THE IDENTIFICATION OF TWO MATURITY LOCI SHEDS LIGHT ON PHOTOPERIODIC FLOWERING IN SORGHUM

A Dissertation

by

REBECCA LEA MURPHY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Biochemistry

The Identification of Two Maturity Loci Sheds Light on

Photoperiodic Flowering in Sorghum

THE IDENTIFICATION OF TWO MATURITY LOCI SHEDS LIGHT ON PHOTOPERIODIC FLOWERING IN SORGHUM

A Dissertation

by

REBECCA LEA MURPHY

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	John E. Mullet
Committee Members,	Scott A. Finlayson
	Gregory Reinhart
	Dorothy E. Shippen
Head of Department,	Gregory Reinhart

August 2012

Major Subject: Biochemistry

ABSTRACT

The Identification of Two Maturity Loci Sheds Light on

Photoperiodic Flowering in Sorghum.

(August 2012)

Rebecca Lea Murphy, B.S, Centenary College of Louisiana

Chair of Advisory Committee: Dr. John Mullet

Harnessing the control of flowering time in *sorghum bicolor* has been essential to programs committed to the development and improvement of this crop. The success of such programs was dependent on the utilization of six Maturity Loci, photoperiod-responsive floral repressors discovered through classic heritability studies. However, the identities of the genes underlying these loci have remained largely unknown. The elucidation of these genes allows for accelerated marker-assisted breeding programs and contributes to the understanding of flowering time in short day plants. Thus, in these studies, two Maturity Loci were identified using a map-based cloning approach, and alleles of each were sequenced in the germplasm. Expression analysis of individual genes by qRT-PCR and the transcriptome by RNAseq was utilized to characterize their response to photoperiod.

Maturity Locus 1 (Ma_1), the most effective of the loci, was identified as *PSEUDORESPONSE REGULATOR 37*, a component of the circadian clock. Sequence analysis revealed an allelic series at this locus, each conferring photoperiod insensitivity

to varying degrees. It was demonstrated that the expression of this gene is regulated by the circadian clock, yet also highly dependent on light. Moreover, PRR37 was found to up-regulate floral repressors while down-regulating activators, providing a mechanism of flowering control consistent with the external coincidence model.

Maturity Locus 6 (*Ma_6*) also generated interest through its genetic interaction with *Ma_1*, and was identified as *Grain Yield, Plant Height and Heading Date 7* (*Ghd7*). Sequence analysis of *Ghd7* revealed several severe mutations and these were traced through several Milo maturity standards, sweet and bioenergy varieties, as well as the pedigree of lines used heavily in the conversion of tropical sorghum to early flowering types. The expression of *Ghd7* mirrors that observed for *PRR37*, and is also regulated by both light and the circadian clock. *PRR37* and *Ghd7* together confer greater repression of floral activators than either alone, but do so independently via pathways that converge on the same downstream "florigen". Thus in sorghum varieties with functional alleles of both, floral initiation is delayed indefinitely in long day photoperiods. The identification of these two genes provides a novel perspective on flowering in short day plants, while also accelerating breeding efforts that ultimately result in improved sorghum varieties for food, forage, and biofuels.

DEDICATION

I would like to dedicate this dissertation to my mother, Sara Murphy, and my grandmother, Shirley Murphy, as their support and love has been an integral part of the success of this endeavor. I also dedicate this to my grandfather, H. C. Murphy, who has yet to miss an opportunity to ask "How 'bout that Milo?"

ACKNOWLEDGEMENTS

I would like to thank my PI and committee chair, Dr. John Mullet, and my committee members, Dr. Scott Finlayson, Dr. Gregory Reinhart, and Dr. Dorothy Shippen for their support and thoughtful perspectives throughout the duration of these projects. Thanks to all of the members of my lab, especially to Dr. Daryl Morishige, who sacrificed his own circadian rhythms to help me collect the ones presented here, and to Susan Hall, without whom the coordination of field and greenhouse studies would have been unmanageable. I would also like to thank the members of the Devarenne and Shippen labs, particularly Dr. Anna Nelson Dittrich and Dr. Andrew Nelson, for their general support and advice concerning certain biochemical aspects of this research. Thanks also go to Drs. Patricia and Robert Klein for their contribution to the data analysis and genetic mapping of the *Ma_1* locus, and to Dr. Bill Rooney for his direction concerning the field studies.

NOMENCLATURE

CAPS	Cleaved Amplified Polymorphic Sequences
DD	Constant Dark
DTF	Days to Flowering
INDEL	Insertion-Deletion
LD	Long Day
LL	Constant Light
PI	Photoperiod Insensitive
PS	Photoperiod Sensitive
PCR	Polymerase Chain Reaction
qRT-PCR	quantitative Real-Time Polymerase Chain Reaction
SD	Short Day
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

SORGHUM BICOLOR: INTRODUCTION AND DEVELOPMENT IN THE U.S.

Sorghum (*Sorghum bicolor* [L.] Moench) is a C4 grass species with origins in Africa. It is grown widely in the United States, where it serves primarily as a source of feed for livestock. In many other countries, sorghum grain and syrup provide a major source of food for approximately 500 million people worldwide (1, 2). The sorghum grain has nutritional properties similar to that of related cereals maize and wheat (3, 4), however; its ability to grow in semi-arid locations with relatively low water and nitrogen requirements confers an advantage over these other crops in marginal lands (3, 5-7). The versatility and heartiness of the sorghum plant have resulted in increasing annual yields of up to approximately 60,000 metric tons per year (8), currently making it the fifth most economically important cereal crop (2). Of this worldwide yield, the U.S. currently produces about one sixth of this, second only to Nigeria (9). However, the journey from the tall, late-flowering varieties originally introduced into American cultivation to the short, high-yield grain sorghum grown today in more than 30 states across the country is a fascinating story involving the foresight and intuition of a few dedicated scientists, several agronomically important genes, and a lot of selective breeding.

This dissertation follows the style of *Proceedings of the National Academy of Science of the United States*.

Today, cultivated sorghum (*S. bicolor* ssp. *bicolor*) is divided into five major races: (1) bicolor; (2) caudatum; (3) durra; (4) guinea; and (5) kafir; each of which is thought to have originated in Africa, or nearby India (Fig. 1) (10). Intermediate races have also formed though interbreeding between the races, and today represent a not insignificant proportion of cultivated sorghum. Bicolor, considered the most primitive of the races, originated in central Africa, where the initial domestication of sorghum was thought to have occurred (11). It is now grown widely, though diffusely, throughout Africa (10). Moreover, the widespread distribution of bicolor is thought to have preceded the evolution of the other races, which arose in spatially distinct regions outside the initial area of domestication. The distribution of bicolor to the west is marked by the region in which guinea race evolved (10, 11). Guinea, even today the predominant cultivated race of west Africa, is more adapted to wet climates, and can tolerate up to 15 times as much rainfall as certain other types (10). It has, however; made its way eastward into the more humid regions of southeastern Africa, primarily as intermediate races through hybridization with other main groups (11). The spread of bicolor eastward resulted in the evolution of the durra race. It is hypothesized that early bicolor gave rise to durras in India, which then made their way back to Africa as a distinct race, though this point is disputed. It may be that durras first evolved in eastern Africa and then spread into Asia, or perhaps evolved concurrently in both locations, with germplasm exchange facilitated by sea travel, though its distribution suggests otherwise (11). Despite its origins, durra is now the primary variety grown in the Ethiopian area (10).

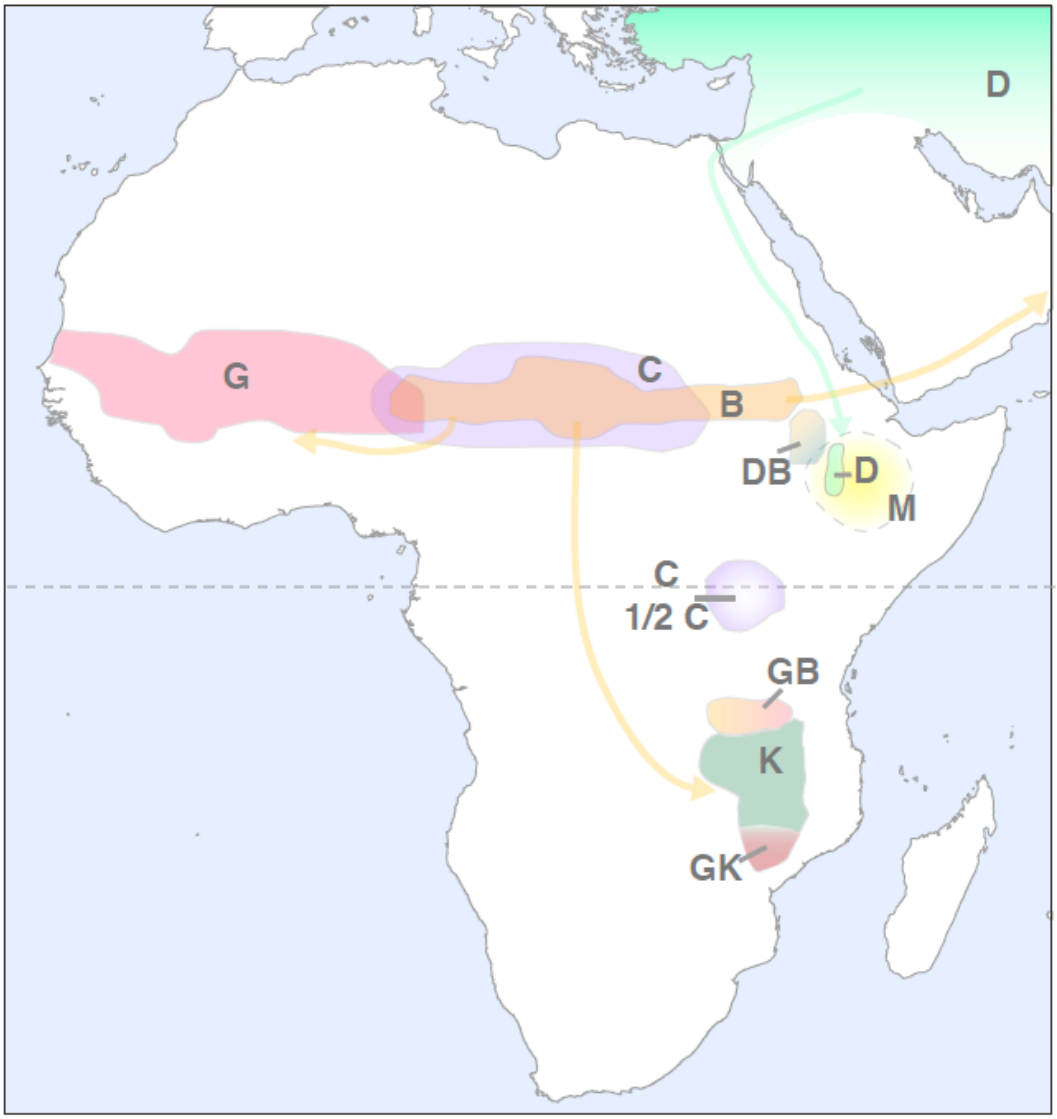

Fig. 1. Origins of the major sorghum races. Central African races include early bicolor (B) orange; guinea (G), pink; caudautum (C), purple. Eastern African durras (D), light green, arose in India/Asia, and kafir (K), dark green, originated in southern Africa. Intermediate races are noted with combinations of these notations, *e.g.* durra-bicolor (DB). The hypothesized region within which the milo group (M) may have originated is marked in yellow, based on information from Quinby and Karper (1946) (1). The dotted line represents the equator. Arrows denote direction of distribution, colored in accordance to the race. Image adapted from (10), p.14. The outline of the map of Africa was downloaded from freeworldmaps.net.

The final two races, the kafirs and caudatums, occupy distinct regions that coincide with their origins. The kafirs arose from the distribution of the early bicolor race to the south. Kafirs are the predominant variety grown in this region of Africa, and generally not found in any other area (10, 11). The younger race, caudatum has origins overlapping that of bicolor, and even extending past its circumference, having segregated out of the territories of early bicolor and guinea (10). Today, caudatum is an important variety to the people of this area, but is generally not cultivated outside this region.

In accordance with the growing sea trade industry, representatives from each of these races eventually migrated to the United States via multiple routes (10). Records indicate that sorghum in the form of Broomcorn, "guinea corn," or "White Milo Maize" was probably grown in the United States by 1853, perhaps even as early as the late 1700s. However, introduction of the founding grain cultivars that gave rise to the sorghum we know today likely began in the mid-1870s with the California-based cultivation of Egyptian White and Brown Durra varieties (Fig. 2). Widespread interest in this durra, especially the white-seeded version, was generated by the late 1800s, and contributed to the growing popularity of the grain crop. In a short time, this type of sorghum could be found growing as far east as Kansas, and its relative success paved the way for the introduction of the kafir and milo cultivars which would ultimately be the source of much of the germplasm present in today's commercial varieties. These kafirs, introduced in 1876 as White and Red varieties, along with Blackhull kafir soon exceeded

the cultivation of their durra counterparts. In fact, by 1936, Blackhull Kafir was being grown more extensively than any other variety in the central US.

The history of the Milo variety, a subset of the durra race, is somewhat less clear. It is thought that Standard Yellow Milo and Giant Milo, a tall, late-lowering variety that was introduced in the 1870s that has since vanished from cultivation, were the progenitors of much of the Milo germplasm utilized in current times, though whether or not these were actually the same genotype is not entirely apparent (1). Nevertheless, a series of mutations in the Milo background eventually gave rise to Dwarf Yellow Milo, Double Dwarf Yellow Milo, Standard White Milo, and Dwarf White Milo, each of which was selected for the varying degrees of short stature that conferred many advantages with respect to grain production. These varieties, in combination with the original kafir introductions, comprised up to 80% of sorghum grown through 1940, and formed much of the original germplasm used to develop the hybrid sorghum utilized today (1). Selection of sorghum with increased yield and improved traits for harvesting, though considered the second phase in the improvement of sorghum cultivars, took place somewhat concurrently with the introduction of new varieties. Optimizing the crops for large-scale, mechanized harvests was the ultimate goal for initial improvement efforts. However, in the beginning, most sorghum was 2 to 3 meters tall and had bent panicles, and as such were primarily harvested by hand. Certain promising cultivars were selected from random Milo by Kafir crosses as early as 1911, as were short plants that arose from spontaneous mutations in taller varieties, particularly in one of the four dwarfing genes dw_1-dw_4 (1). The release of the Wheatland cultivar in 1931, a variety selected for its

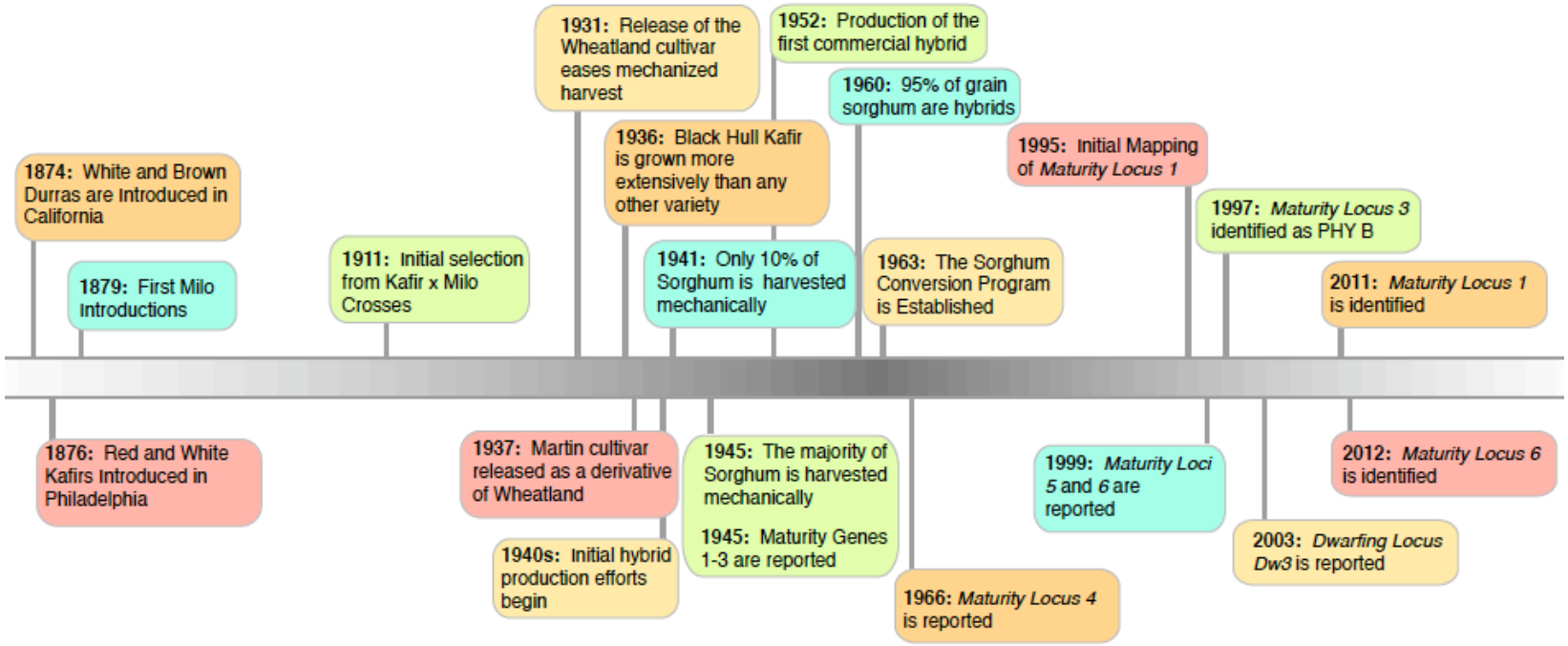

Fig. 2. Timeline of major advances in sorghum development.

early, short phenotype and disease resistance, was developed for use with a wheat combine, along with others, notably Plainsman, Caprock, and Combine 7078. Even so, by 1941, mechanized harvest of this crop was not yet widespread. This changed quickly, however, due not in the least to the serendipitous discovery of the Martin cultivar, the derivative of a single mutant plucked from a field of Wheatland in 1937 (1). Martin, named for the farmer who had the foresight to propagate it, displayed the ideal characteristics for combine harvesting: a short-stemmed plant with a large and upright panicle. The popularity of this cultivar spread quickly, and it was grown widely and used extensively in crosses to more combine-ready types. Thus, via selective utilization of the dwarfing genes, by 1945 the vast majority of sorghum could be harvested mechanically (12, 13). Due to the agronomic importance of these loci, much emphasis has been placed on uncovering the causative mutations. However, aside from the identification of *dw*$_3$ as PGP1, an MDR Transporter that affects polar auxin transport (14), these genes have remained unknown. Nevertheless, the short stature achieved through the use of these loci represents a major landmark in sorghum's success as a commercial crop in the United States.

Another essential milestone in the development of sorghum was the selection and identification of the four classic maturity loci (*Ma*). From crossing different types of Milo, four phenotypes arose, classified as early, comprising cultivars of Sooner Milo (46-60 days), intermediate, representing Yellow and White Milo varieties, (64-74 days), late (76-88 days), and ultra-late, (92-106 days). A 1945 report on the inheritance patterns of these traits (15) revealed that three maturity genes, *Ma*$_1$, *Ma*$_2$, and *Ma*$_3$,

conferred these differing categories of maturity, with *Ma1* being the most effective. Interesting genetic interactions exist between these loci, with *Ma1* being required to observe the effects of either *Ma2* or *Ma3*, suggesting that *Ma1*'s biochemical effects may be downstream of the other loci. Additionally, the late flowering phenotype observed in varieties with a dominant maturity background was eliminated in SD, indicating that the differences in flowering time observed was conferred primarily through photoperiod sensitivity (15). The discovery of these genes had significant implications for commercial sorghum breeders, as the time to maturity affects height, grain and sugar yields. The utilization of these loci was also important for those growing sorghum in northern latitudes where differences in day length are greater than those found at more tropical regions. In 1966, a fourth maturity locus, *Ma4* was reported (16). An initial cross between Milo genotype SM60 and Hegari revealed a subset of F_2 offspring that flowered at 48 days, even earlier than the SM60 parent (16). From this observation, it was clear that a recessive allele was present in Hegari that was distinct from the original three loci. Further crosses between Hegari and established genotypes with known maturity backgrounds revealed that the phenomenon observed in the Hegari x SM60 cross was indeed the result of a novel locus (16). Even at present, Hegari (*Ma1*, *Ma2*, *Ma3*, *ma4*) and its derivative, Early Hegari (*Ma1*, *Ma2*, *ma3*, *ma4*), are the only known varieties with recessive alleles at this locus (17). The identification of *Ma4* in combination with the original three loci comprised the classic set of genes that formed the basis of the sorghum conversion program.

The sorghum conversion program, established in 1963, became necessary because the wide use of only a few cultivars resulted in a bottleneck that greatly reduced the diversity in germplasm (18). Vulnerability to pathogens and environmental stresses drove the need to introduce more exotic cultivar into the established farming system, however, these exotic lines had the tall, late-flowering phenotypes that breeders had spent half a century selecting against in American sorghum. In order to retain the ideal harvesting traits selected for in the first half of the 20th century while introducing beneficial germplasm, the sorghum conversion program used selective breeding practices to obtain the desired combination of loci (19). Exotic varieties were crossed to specific lines developed with a desired set of recessive maturity and dwarfing genes. The resulting progeny were backcrossed to the exotic parent for multiple generations, and selected for the desired flowering time and height at each step. This ultimately resulted in the transfer of recessive maturity and dwarfing genes into an otherwise novel genetic background. The most notable point of conversion with respect to flowering time was the introduction of the recessive *ma$_1$* locus (20), done primarily through breeding with BTx406, a line derived from Martin and various dwarf Milo cultivars (18, 21). Throughout the course of the program, more than 700 lines were converted in this manner (18).

Eventually, it was noticed that certain lines could not be converted effectively from photoperiod sensitive phenotypes to the photoperiod insensitivity, and the four initial loci could not account for this effect. Moreover, it was also observed that certain hybrids actually flowered much later than either of the parents, especially in crosses with

R.07007 (EBA-3) (22). Crosses between R.07007 and several other maturity lines revealed the existence of two new maturity loci, Ma_5 and Ma_6. Most American sorghum is thought to have dominant alleles at Ma_5 and recessive alleles at ma_6, while conversely, R.07007, an Argentinean variety, is ma_5Ma_6 (22). Thus, when a cross is made using R.07007 as a parent, the loci complement, producing an offspring that flowers extremely late, often after 170 days, or even later.

Throughout most of the 1900s, sorghum was selected for properties that enhanced yield and certain harvesting feasibility requirements, primarily by developing early-flowering, short cultivars. Because of these early efforts, sorghum has been well established as food crop for grain and syrup production, as well as a source fodder or forage for livestock. However, recently, in the face of growing energy demands, sorghum has generated interest as a potential bioenergy feedstock (23). Sorghum possesses excellent water and nitrogen use efficiencies that allow it to be grown on marginal land otherwise unsuitable for food production, thus it does not displace other crops grown for this purpose (6, 7). Additionally, total ethanol yields from sweet sorghum are approaching that of sugarcane (23). High biomass lines that can grow to heights of six meters or more are also increasing in popularity, due to the advances in technology for conversion of cellulosic material into ethanol. Production of this type of sorghum requires extremely late flowering times. Sorghum continues to produce vegetative growth up until the point when floral initiation occurs, so extremely photoperiod sensitive varieties accumulate more biomass throughout the season relative to their early-flowering counterparts. Therefore, to generate plants with the desired traits

for this purpose, dominant maturity loci must be introduced into promising bioenergy cultivars.

Whether sorghum is grown for grain, sugar, or biofuels production, achieving optimal yields is strongly dependent on the ability to modulate maturation time. Thus, identifying and characterizing the genes responsible for conferring varying degrees of photoperiod sensitivity becomes essential. These genes have been thoroughly studied with respect to genetics, yet, aside from the identification of *Ma_3* as PHYTOCHROME B in 1997 (24), the molecular basis of photoperiod sensitivity has remained largely uncharacterized for sorghum and many other species. Studying this trait in sorghum directly is both feasible and important, as it will greatly accelerate molecular assisted breeding efforts aimed at crop improvement. Sorghum possesses a relatively small genome, about 730 Mb (25) spread across ten chromosomes, and it did not acquire the whole genome duplication and prolific repetitive elements present in closely related maize (26). Additionally, it retains much colinearity with its more characterized relative, rice, and molecular tools for this plant are being developed, as evidenced by the completion of the genome sequence in 2007 (27). However, the knowledge base required to gain a complete picture of this phenomenon has yet to be fully acquired specifically in this species. Fortunately, the information gleaned from other well-characterized organisms can provide a starting point for understanding flowering time in sorghum.

FLOWERING PATHWAYS IN *ARABIDOPSIS THALIANA*

Flowering has been studied extensively in *Arabidopsis thaliana*, and has formed the foundation of knowledge that can be largely translated to other non-model species. The precise timing of floral initiation is incredibly complex and controlled via several routes, and some, like the autonomous pathway, are intrinsic to the plant, while others, like the vernalization and photoperiod pathways, require coordination with external stimuli (Fig. 3). These pathways are not mutually exclusive, and the combined movement of nutrients, hormones, and molecular signals result in a specific balance in the expression of floral homeodomain genes that are necessary for the transition to reproductive growth, much like a balance scale. However, before a plant can flower it must first pass through the first of three major phases of development: the juvenile phase. The transition from juvenile to adult growth is controlled via age-specific regulation of gene expression, regulated internally.

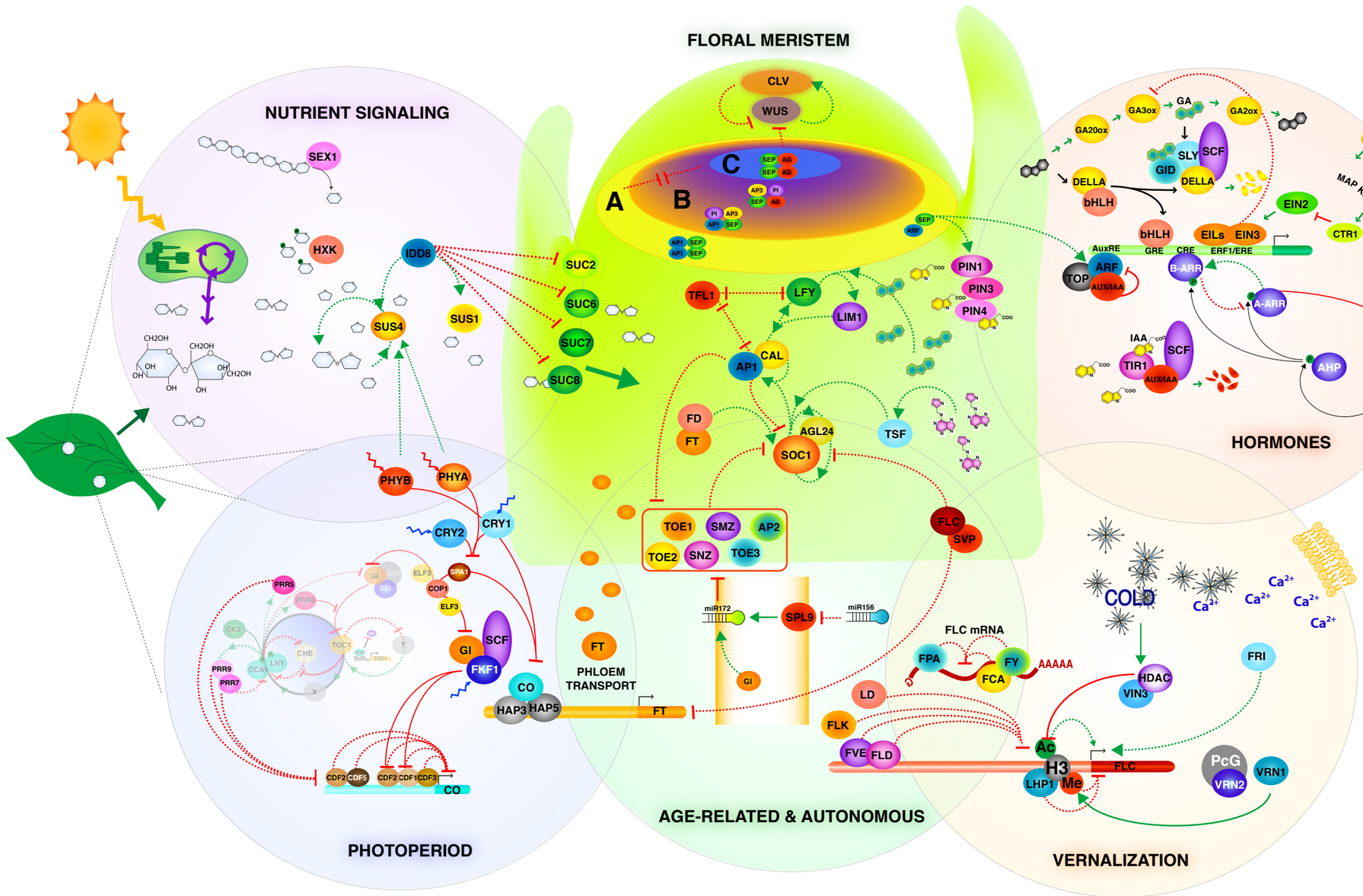

Fig. 3. Integrated floral pathways. Flowering in Arabidopsis is controlled via multiple routes that converge on a few genes, namely FT and SOC1 (middle, orange). These floral integrators subsequently regulate homeodomain genes that specify organ formation. Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions.

THE AGE-RELATED PATHWAY

The juvenile phase is marked physiologically by leaf characteristics that include differences in trichrome formation, leaf ratios, and serration patterns (28). Moreover, during this period, these plants are unresponsive to the external stimuli that would be florally inductive later in their life cycle, including day length and temperature. Molecularly, this phase is marked by the abundance of one small micro RNA transcript, miRNA156 (Fig. 4). The expression of this gene in the apex is high in very young seedlings, and as the plant ages, its expression gradually decreases. While it is being expressed, its targets include a set of related *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) transcripts, which can be divided into four functional groups: *SPL 2/10/11*, *SPL 9/15*, *SPL 3/4/5*, *SPL6/13* (28). The translational repression of these genes promotes juvenile characteristics in various ways, but certain members from two of these groups, *SPL9* and *SPL10*, influence flowering in *Arabidopsis* by activating a second non-coding RNA, miRNA172 (29). Expression of miR172, in contrast to miR156, increases as time goes on. This is due in part to the increased expression of SPL 9/10 protein, which occurs as translation increases in response to diminishing levels of miR156. Once the levels of miR156 are low enough, the plants transition into the adult vegetative phase, marked by adult leaf characteristics and a competency to flower when exposed to inductive conditions.

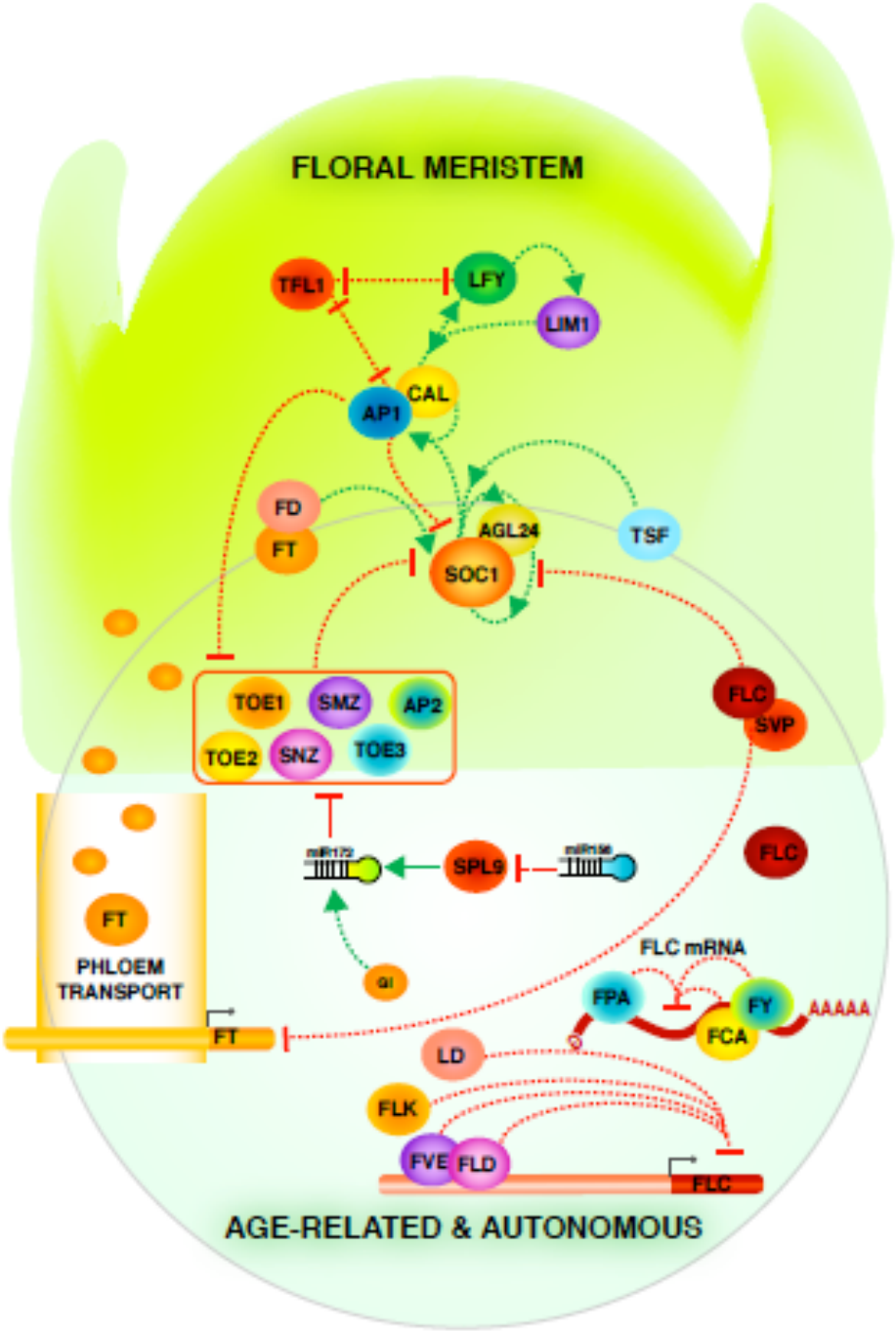

Fig. 4. The age-related and autonomous pathways. The age related pathway involving miRNA156/172, and the autonomous pathway, controlled primarily through FLC, modulate flowering by repressing *SOC1* and *FT*. Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions.

The 21-nucleotide miR172 acts to inhibit translation of *APETALA2* (*AP2*) and a group of *AP2*-like genes comprised of *TARGET OF EAT* (*TOE*) 1, 2, and 3, as well as *SCHLAFMÜTZE* (*SMZ*) and *SCHNARCHZAPFEN* (*SNZ*), which repress the adult phase transition as well as flowering (29, 30). Inhibiting these factors promotes adult growth and releases the repression on floral promotive signals in the apex that, when expressed, result in the transition to reproductive growth. Because this miRNA increases as the plant ages, it can be considered part of the autonomous pathway. However, light, particularly in the blue wavelength, results in increased expression of the *MIRNA172* gene, and miRNA172 transcript is also thought to be positively influenced by photoperiod pathway gene *GIGANTEA* (*GI*) (31). Thus the combined inputs of multiple pathways are likely to influence flowering through miRNA172, as well as the transition from juvenile to adult. Once a plant reaches adulthood, it becomes receptive to the influences of other floral pathways, and sensitivity to these signals is controlled through *FLOWERING LOCUS C* (*FLC*) and the autonomous pathway.

THE AUTONOMOUS PATHWAY

The heart of the autonomous pathway lies in a single MADS-box transcription factor FLC (Fig. 4) (32). FLC binds directly to the promoters of several important floral activators, repressing their transcription and preventing the transition to reproductive growth (32). Included among these is *FT*. *FT* encodes a small RAF-kinase inhibitor-like protein that is part of the PEBP family in plants (33). This family contains six members in *Arabidopsis*, but has expanded to varying degrees in other species (34).

Multiple members of the family can be involved in floral activation, especially in cereals, but in *Arabidopsis* flowering can be activated primarily through FT. Most importantly, this small protein plays the role of "florigen," and is the primary mobile signal that induces the change from vegetative to reproductive growth (35-37). Therefore, in order to relieve the repression that FLC places on *FT* and other inductive molecular signals, most other members of the autonomous pathway work in some way to decrease transcription or translation of the FLC protein. FCA, FY, and FPA bind to the mRNA of FLC (38-41). FCA, an RNA Recognition Motif (RRM) protein, interacts with FY, a 3' RNA processing factor, through its WW domain. This pair, in combination with the activities of a second RRM protein, FPA, works to down regulate FLC by promoting polyadenylation of *FLC* antisense RNA (42). *FLC* chromatin also contains histones that are methylated in a repressive manner, specifically at H3K9 and K3K27, in response to certain environmental stimuli. The repression of FLC through chromatin remodeling may also suggest a parallel role for two histone deacetylases (HDACs), FLD and FVE, which also act to inhibit *FLC*, likely by reinforcing the transcriptionally repressed state of its chromatin (43, 44). Additionally, FLD activity is connected to FLC repression via FCA/FY and FPA, though this mechanism is not well understood. Other factors, *FLOWERING LOCUS K HOMOLOGY [KH] DOMAIN* (*FLK*) and *LUMINIDEPENDENS* (*LD*) have also been shown to decrease FLC gene expression (45-47). Ultimately, the expression of the FLC protein is diminished over time through multiple levels of control, including gene transcription and mRNA processing and silencing. This provides a very precise level of FLC, and because this protein represses

many important floral factors that integrate the outputs of multiple pathways, this results in a tightly regulated sensitivity to outside inductive stimuli (48).

VERNALIZATION

The major components of the autonomous pathway also play a large role in thermoregulation of flowering via vernalization (Fig. 5).

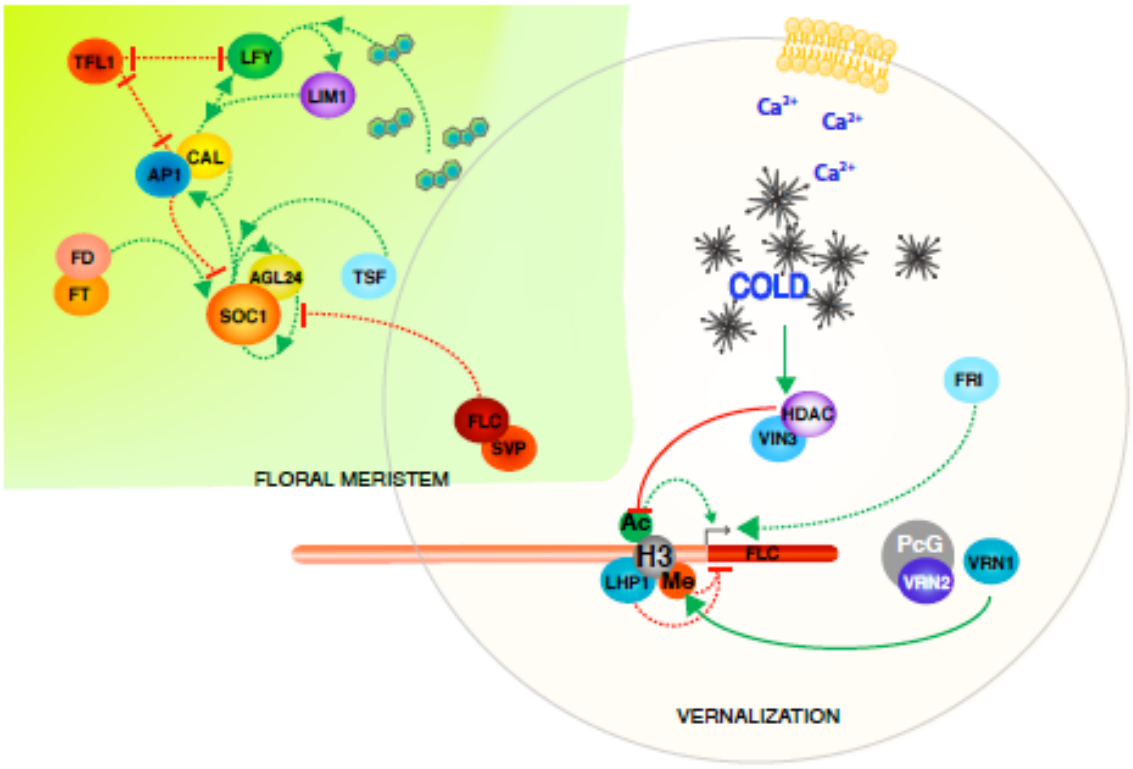

Fig. 5. The vernalization pathway. Exposure to winter temperatures, potentially sensed through changes in membrane fluidity or Ca^{2+} flux, results in the mitotically stable repression of *FLC* through the recruitment of chromatin remodeling factors. Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions.

Vernalization refers to the requirement for prolonged cold that allows certain temperate species to sense the change from winter to spring so that flowering occurs in the appropriate season. As with the autonomous pathway, the central player in the molecular response to cold is FLC. It is thought that plants sense cold through differences in membrane fluidity, or by calcium signaling (49), and prior to sensing winter temperatures, *FLC* is unregulated by FRIGIDA (FRI). However, after a plant has experienced prolonged cold, the expression of VERNALIZATION-INSENSITIVE 3 (VIN3), a Plant Homeodomain protein (PHD) that associates with the Polycomb Repressive Complex 2 (PRC2), is activated in the apex (50, 51). In *Arabidopsis*, the activity of the VIN3-containing PRC2 is restricted to vernalization-specific gene regulation, initiating repressive chromatin structure by recruiting histone deacetylases (HDAC) at FLC (49). This results in FLC down-regulation in response to cold. However, this response is transient, and the FLC chromatin returns to an active state when once again exposed to higher temperatures. This observation suggests that while VIN3 can initiate FLC repression, it is not sufficient to maintain it. In fact, the maintenance of the silenced state requires two additional factors, *VERNALIZATION1* (*VRN1*), a MYB-transcription factor, and *VERNALIZATION 2* (*VRN2*), a component of the PRC2. The VIN3-mediated deacetylation of the FLC chromatin allows the VRN1/VRN2-containing PCR2 to associate with FLC chromatin and facilitates repressive methylation on K9 and K27 of histone 3 (H3). Furthermore, this methylation pattern recruits *LIKE HETERCHROMATIN PROTEIN1* (*LHP1*), which is necessary to maintain this methylation state (50, 52). This cold-induced chromatin remodeling is

mitotically stable, and allows the plant to "remember" that it has experienced winter. Thus, the repression on flowering is lifted by cold such that floral initiation may occur in early spring. Interestingly, orthologs of FLC are absent in the cereals, though certain species, namely wheat, require vernalization. It is thought that the role of FLC is filled via a set of MADS proteins for the autonomous pathway, however; in wheat, the vernalization requirement is mediated through *VERNALIZATION2* (*VRN2*). *VRN2* is a *CONSTANS, CONSTANS-like, TOC1* (CCT) domain protein that represses *FT1* (an FT-like gene) and flowering, probably by competing with CO2, a second CCT protein and floral promoter, for binding in the HAP complex, a modulator of transcription (53). Additionally, in *Arabidopsis*, a small RNA with antisense to FLC called COOLAIR, is up-regulated upon exposure to cold. Though the involvement of this RNA in vernalization is not well understood, it is apparent that many layers of control exist to regulate FLC levels, both internally through autonomous genes, and externally via cold exposure. In addition to FLC, the autonomous pathway controls the activity of a second MADS-box transcription factor, *SHORT VEGATATIVE PHASE* (*SVP*). In leaves, SVP represses FT in response to cool ambient temperatures, and this may more precisely time the initiation of flowering in varieties that coordinate reproduction with the seasonal shift from winter to spring. Moreover, SVP integrates negative stimuli from temperature and age-related responses with those received through hormone signaling (54).

HORMONE SIGNALING

Hormone pathways in plants play many roles in growth and development, and often act to regulate one another (Fig. 6). Though it has been long established that gibberellins, cytokinins, auxin, ethylene and others are involved with floral initiation and patterning, the pathways that connect hormone responses are so complex that the part each plays in flowering is still not entirely understood. However, of these, the flowering response to gibberellins (GA) has been most thoroughly studied. GAs are a class of diterpene hormones that, in a broad sense, accelerate flowering in addition to being and important regulator of general growth. In fact, GA was isolated from the rice fungal pathogen *Gibberella fujikuroi*, which throughout the course of infection caused excessive stem elongation. Additionally, the molecular basis underlying the dwarf phenotype in plants utilized in the 1970s "green revolution" lies in mutations in components of the GA pathway (55). In order for GA to modulate the response to external stimuli and to fulfill growth requirements with proper timing, a major point of control is through regulating biosynthetic enzymes.

GA biosynthesis occurs largely in the plastid and is tightly regulated diurnally and in response to environmental cues (56). Though hundreds of species of GA exist in the plant, only a handful of these are bioactive, namely GA_1, GA_3, and GA_4; the latter is though to be the primary bioactive form in *Arabidopsis* (55, 56). The reactions that convert inactive precursors into these active forms are catalyzed primarily by GA 20- and GA 3-oxidases (GA20ox, GA3ox) and the expression of these enzymes is

responsive to inductive stimuli (56). Additionally, the activity of the active forms can be tempered primarily via the activity of GA 2-oxidase (GA2ox), which catalyzes a

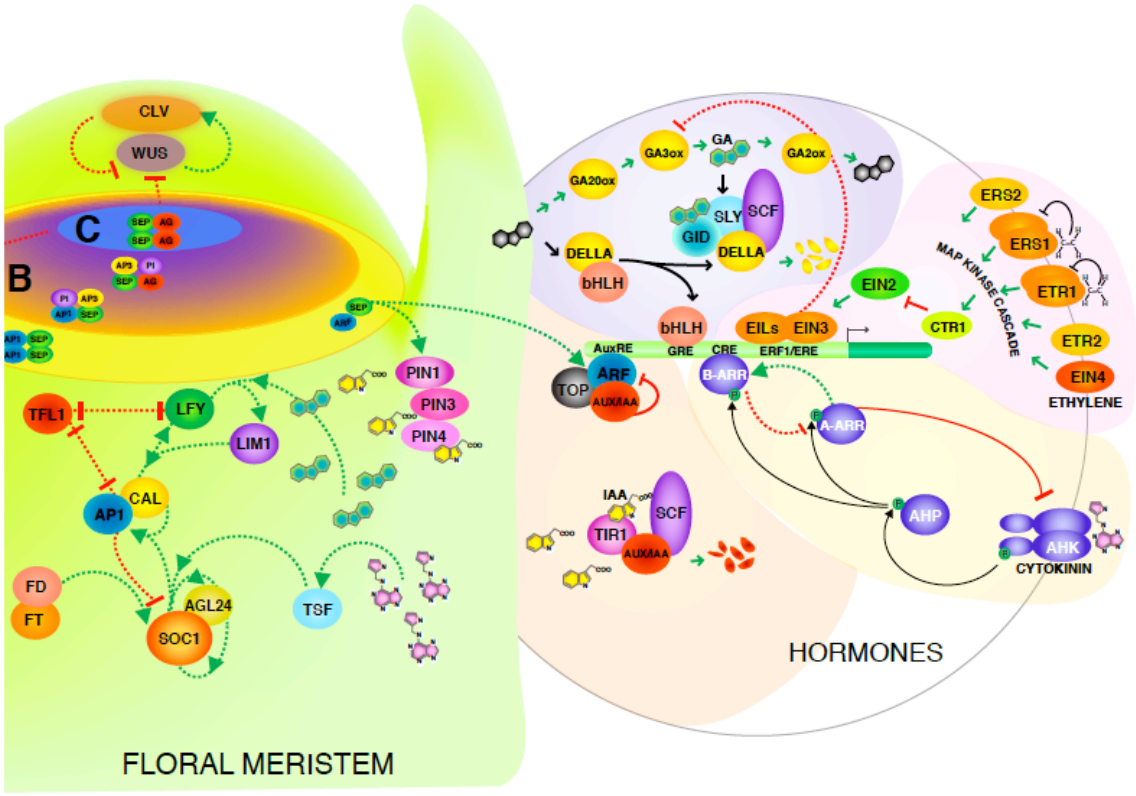

Fig. 6. Integrated hormone pathway. Gibberellins (GA), cytokinins, ethylene, and auxin (IAA) modulate the activity of hormone responsive transcription factors, downstream genes, and the levels of each other. These signals are integrated in the meristem through the flowering genes *LFY*, *SOC1*, *TSF*, and others. Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions.

2β-hydroxylation reaction that results in decreased GA signaling (56). The regulation of relative quantities of specific bioactive species in certain tissues provides the first layer of control in the role of GA responsive floral induction.

A second regulatory point occurs through GA signaling. The GA signaling pathway operates via the "release of restraint" model (57). This refers to the idea that GA-responsive genes are repressed in the default state, due to *Arabidopsis* DELLA proteins, a subgroup of GRAS family transcription factors. In the context of flowering, the five DELLA proteins, GIBEBERELLIC ACID INSENSITIVE (GAI), REPRESSOR OF GA1-3 (RGA), and RGA-like 1, 2, and 3 (RGA1/2/3), act to transcriptionally repress GA-responsive genes by sequestering their activators. This mechanism is illustrated via PHYTOCHROME INTERACTING FACTOR (PIF) and PIF-LIKE (PIL)-mediated light response (58). These PIFs are bound by DELLAs so that they cannot transmit the signals received from light. When GA biosynthesis is up-regulated, the hormone is perceived by GA-INSENSITIVE DWARF1a b, and c (GID1a, b, c) and SLEEPY1 (OsGID2), known GA receptors that bind to DELLAs in a GA-dependent manner and mediate their degradation through the 26S proteasome (58). SLEEPY1 (SLY1), in particular, is an F-box protein that functions as an adapter to regulate GA-specific protein targeting to an SCF E3 ubiquitin ligase. Once the DELLA proteins are degraded, the PIFs they previously bound become free to modulate GA-responsive genes and growth (59).

In the context of flowering, the accumulation of GA in the apex during floral initiation and development is required, and this is thought to be the result of GA

movement in from the leaves in addition to some synthesis directly in the apex (58). GA promotes flowering in short days (SD), and several genes involved in the meristem identity switch from vegetative to reproductive growth in the apex show increased expression in response to GA, including *LEAFY* (*LFY*), which is up-regulated by GA through a GA-responsive MYB factor (GAMYB) (60), and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) (59). Input from multiple external stimuli and internal developmental programs, including not only GA-responsive genes, but from the autonomous and vernalization pathways via FLC and SVP, and photoperiod responsive factors converge on SOC1. The expression level of this gene integrates messages from multiple sources and translates them into one signal that begins the development of floral organs. As these organs begin to emerge, GA continues to play a role throughout the development of the flower. Many of the genes targeted by RGA have catalytic function, and include cell wall modifying enzymes. Nucleic acid binding proteins and transcription factors are also represented to a lesser degree (57). Interestingly, GA plays a large role in later flower development, likely through cell expansion associated with floral organ morphology (57, 61). Moreover, crosstalk between other hormones and GA can influence this process, and the interplay between these pathways control certain aspects of floral initiation, stamen morphology and dehiscence (pollen shed) (62). Many hormones interact synergistically with GA to promote flowering, however some act in opposition, and one of which is ethylene (63).

Ethylene is a small, volatile hydrocarbon (C_2H_4) produced in response to various stimuli, generally including fruit ripening, stress, pathogens, and wounding (64). Its

synthesis, which is regulated by these environmental cues, occurs via the Yang cycle, which utilizes methionine as the initial substrate (55). Ethylene perception occurs in a two-component signaling pathway through a set of five membrane-bound ethylene receptors, ETHYLENE RESPONSE 1 and 2 (ETR1/2), ETHYLENE INSENSITIVE 4 (EIN4), and ETHYLENE RESPONSE SENSOR 1 and 2 (ERS1/2), of which ERS1 and ETR1 function as homodimers (64-68). In the absence of ethylene, this family of receptors send signals to CONSTITUTIVE TRIPLE RESPONSE 1 (CTR1), a Ser/Thr kinase, which is proposed to repress positive ethylene signaling by ETHYLENE INSENSITIVE 2 (EIN2) via a MAP-kinase cascade (64, 69). When ethylene is bound, the negative signal from CTR1 is repressed, allowing EIN2 to promote the activity of a set of transcription factors, EIN3, and EIL1 and 2, which bind and regulate the ethylene response element binding protein (EREBP) ERF1, which then binds to the ERE in the promoters of target genes, eliciting a response (64).

Ethylene is often produced in times of stress, and in such cases the plant may benefit from delayed flowering. Though the effects of ethylene are pleiotropic, it has been demonstrated that there are specific repressive actions in the context of floral initiation. As ethylene is produced, signaling via the CTR1-EIN3 pathway results in transcriptional repression of certain GA20- and GA3-oxidases (63). The inhibition of GA biosynthesis leads to more DELLA accumulation, and specific repression of *SOC1* and *LFY* transcripts, which encode MADS-box proteins essential for floral development. The antagonistic action between GA and ethylene represents flowering time control that is regulated in response to promoting and inhibitory growth conditions. GA promotes

flowering in already inductive photoperiods, and because the GA effects modulated by ethylene are downstream of FT, a primary systemic signal, this mechanism allows the plant a "last chance" to delay flowering, even when receiving positive signals from other pathways. However, even though GA and ethylene regulate the transition to flowering, the initiation of a floral meristem will not occur without a third hormone: auxin.

In the meristem, local biosynthesis and specific spatial accumulation of auxin are necessary for the positioning and patterning of floral meristems. Indole-3-acetic acid (IAA), the most prevalent auxin in plants, is synthesized in the leaves, roots, and shoot apical regions from tryptophan and from tryptophan precursors, and is then sent to exact locations via a highly-specific transport system (55, 70). As the inflorescence meristem elongates, floral meristems that give rise to a flower develop at regular intervals along its length. The precise patterning of these flower primordia is regulated by specific auxin transport (71). IAA exists in the cell wall in its IAAH form, and its entrance into the cell is mediated through diffusion and AUX1 (70, 72). In the cytoplasm, the pH exceeds the pKa, and IAAH is converted to its acidic IAA form, which can no longer pass freely through the cell membrane (72). In order to leave the cell, IAA must be transported out, and herein lies the key to auxin-dependent patterning. Auxin export is mediated by the PINFORMED (PIN) auxin transport proteins, which are polarized in response to light, gravity and other stimuli through PINOID (PID), a protein kinase, and protein phosphatase 2A (PP2A) (73, 74). This directed auxin transport results in the localization of PIN proteins along one side of the membrane, creating a flow of auxin that occurs in

the desired direction. This directional transport of IAA is used to create auxin maxima that determine where floral primordia will arise (71, 75).

In the appropriate cells, auxin induces the transcription of several gene families, including *AUXIN/INDOLE-3-ACETIC ACID* (*Aux/IAA*) genes. Aux/IAAs are transcriptional repressors that bind to AUXIN RESPONSE FACTOR (ARFs) proteins, preventing them from inducing *AuxRE* genes in the absence of auxin (55). When a sufficient level of IAA accumulates, it is perceived by TIR1, an auxin receptor and F-box component of the SCF E3 ubiquitin ligase. TIR1 binds to Aux/IAAs in an auxin-dependent manner, and marks them for degradation by the 26S proteasome, relieving the repression on ARF-mediated auxin response genes (70, 74). In the floral meristem, precise patterns of gene expression are required for the correct positioning and morphology of floral organs (71). It has been shown that certain homoeodomain proteins that regulate organ formation bind *AuxRE* genes and regulate the transcription of auxin transport genes and ARFs, linking auxin to floral development in a molecular way (76). Additionally, crosstalk between IAA and other hormone signaling pathways, including cytokinin, can modulate the auxin response (77, 78).

Cytokinins are an adenine-derived class of hormones that were identified as cell-proliferation factors that influence many areas of plant growth and development (55, 79, 80). Like GA, cytokinin levels are maintained through balancing the expression of biosynthetic and metabolic genes. The cytokinin signal is perceived by membrane bound *ARABIDOPSIS* HIS KINASE (AHK) receptors, which autophosphorylate in response to hormone binding (55). Subsequently the phosphate is transferred to an

ARABIDOPSIS HIS PHOSPHOTRANSFER protein (AHP), which shuttles the signal from the cytosol to the nucleus through the phosphorylation of A-type and B-type *ARABIDOPSIS* RESPONSE REGULATOR (ARR) proteins (79). Once phosphorylated, B-type ARRs induce the transcription of cytokinin responsive genes (CRF), including A-type ARRs, which then provide negative feedback in the signaling system by repressing the AHK receptors.

In addition to mediating cell division and growth, cytokinin signaling also plays a direct role in the regulation of flowering-related genes. Cytokinin movement has been correlated with photoperiod; cytokinin export from leaves was increased concurrently with accumulation of the hormone occurring in the apex in response to an inductive LD treatment (81, 82). Moreover, treatment with N^6-benzylaminopurine (BAP), a synthetic cytokinin, induced flowering and activated the expression of *TWIN SISTER OF FT* (*TSF*), a floral activator in non-inductive conditions (83). *TSF* expression is regulated in the leaves by the photoperiod pathway in a manner parallel to FT, which is the primary floral activator in LD. This response is integrated through SOC1 (81), tying the production of cytokinin and other hormones into the overall context of flowering.

NUTRIENT SIGNALING

It had long been hypothesized that "florigen," a mobile signal that originates in leaves and is transmitted through the phloem to the shoot apex, exists in plants (84). However, the exact composition of this signal was not known. Eventually, through an elegant series of grafting experiments, it was demonstrated that the FT protein, produced

only in the phloem companion cells in leaves, does exactly that (35, 36). However, FT is not the only signal present in the vasculature of plants in the reproductive phase. Carbohydrates, produced in the leaves as products of photosynthesis, are transported to other parts of the plant, including the shoot apex (Fig. 7) (85, 86). In fact, it has been shown that levels of carbohydrates spike in leaves and at the SAM just prior to flowering, and that these increases accelerate flowering, indicating a correlation between sugar signaling and floral induction (85). The transmission of starch as a signal requires

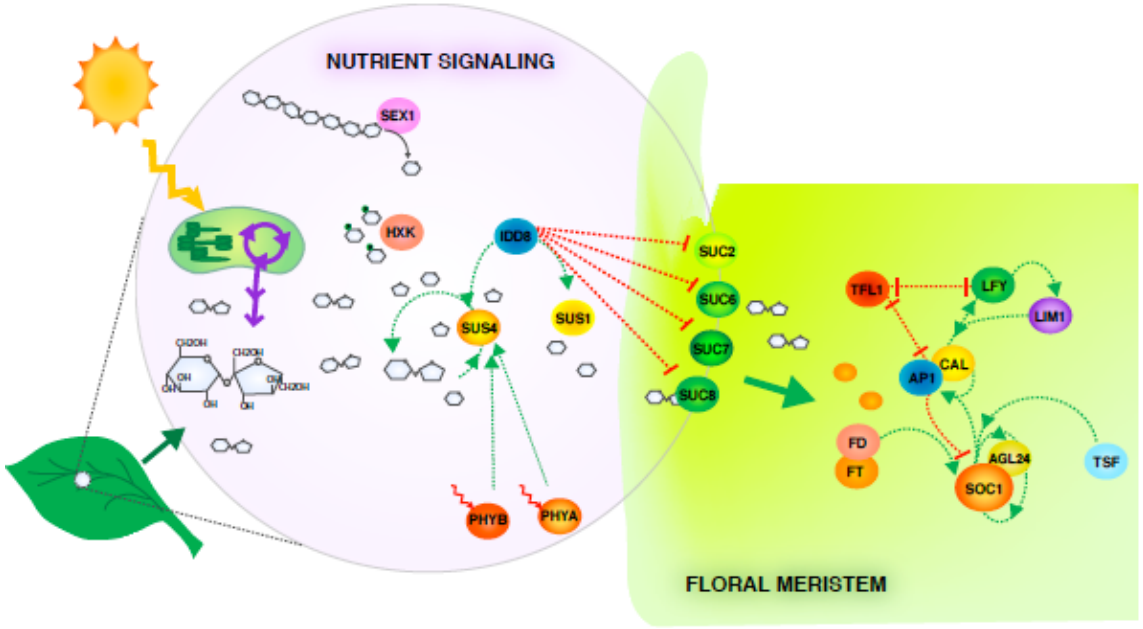

Fig. 7. Nutrient signaling. Starch metabolic enzymes and sucrose transporters mediate mobilization of sugar reserves from the leaves to the apex. Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions. Jagged yellow, red, and blue arrows represent visible sunlight, red/far red, and blue wavelengths, respectively.

multiple components, including starch synthesis and mobilization enzymes, like *STARCH EXCESS1* (*SEX1*), as well as hexokinases (HXK), which are implicated as potential sugar-sensing enzymes (87, 88).

Though the carbohydrate composition in the vasculature is variable and contains multiple types of starch and sugar, the major product downstream from the light-independent reactions of photosynthesis is sucrose (86). In *Arabidopsis*, a family of six sucrose synthase (SUS) enzymes utilizes a nucleoside diphosphate to catalyze the reversible conversion of sucrose into ADP and UDP glucose and fructose (89). Interestingly, the transcription of some SUS members is induced in response to certain light conditions. SUS1 is induced in high light, while SUS4 is regulated in a photoperiod-dependent manner, and increases in expression in response to LD conditions, likely through the effects of PHYTOCROME A and B (PHYA/B) (90). This increase in expression is also due to direct transcriptional regulation by INDETERMINATE DOMAIN 8 (IDD8) (90). IDD8 is a four zinc finger domain transcription factor that is part of a sixteen-member family and regulates sucrose transport and metabolism. Its transcription is not responsive to photoperiod itself, but is involved in photoperiodic regulation of flowering through the genes in modulates. Along with the up regulation of SUS1 and SUS4, IDD8 represses transcription of four sucrose-proton symporters, SUC 2,6,7, and 8. This transcriptional modulation by IDD8 promotes flowering, and *idd8* mutants show later flowering phenotypes (90). Moreover, IDD8 is not the only one in this family that regulates flowering. *INDETERMINATE 1* (*ID1*) was identified in maize because of its strong promoting effect on floral initiation

(91). However, it has been demonstrated that expression of this gene is correlated with increases in photosynthesis and carbon fixation (90) suggesting another link between carbon assimilation and flowering. The consequences of sugar signaling with respect to flowering are still not well understood. However, it may be hypothesized that sugar levels in the vascular tissues may have several roles in floral transition. First, the transmission of sucrose from source to sink may establish a gradient in the plant for other mobile signals to follow. In fact, sucrose may also be a signal that elicits cytokinin transport from the roots in response to floral inductive signals (83). Additionally, though it has been well established that FT moves through the phloem to the apex, it is not known how this occurs. Phloem loading and unloading are processes that require energy from the plant (92). It is likely that sucrose acts not only as a signal, but also provides nutrients to the regions of the plant where these actions occur. Flower development is an energy expensive endeavor, and this increase in nutrients may help nourish the apex while it is going through morphological changes. In fact, it has been shown that sugar signaling is especially important in SD photoperiods, when photosynthesis may be limited (85). However, regardless of the specifics of this mechanism, sugar metabolism and transport is connected to one of the most important pathways in day length-sensitive flowering plants: the photoperiod pathway.

PHOTOPERIODIC FLOWERING AND THE *ARABIDOPSIS* CIRCADIAN CLOCK

Flowering in response to photoperiod has been studied extensively in the long day plant *Arabidopsis*, and requires precise coordination between external stimuli and

internal gene expression regulated through the circadian clock (Fig. 8). The circadian, or ~24-hour rhythm present in plants is a phenomenon that has captured the attention and intrigue of scientists for centuries. Though these rhythms had been observed through external physiological processes like daily flower openings, plant growth, and leaf

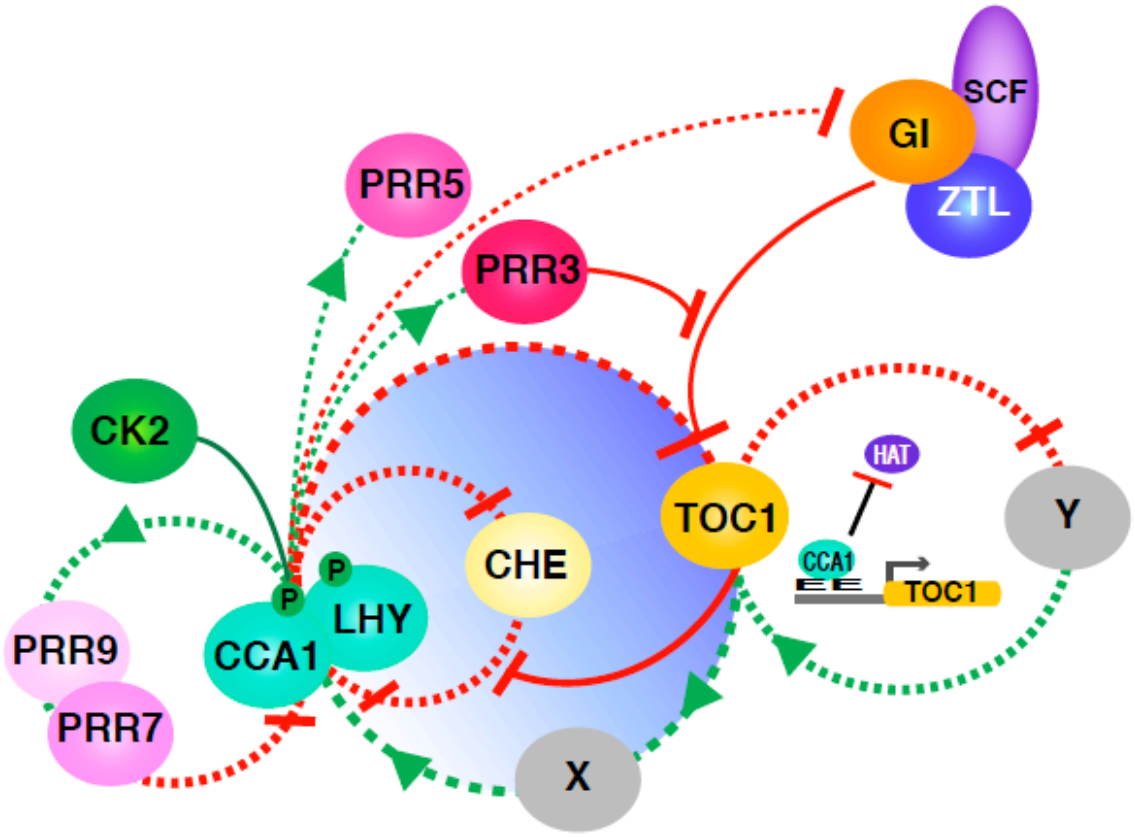

Fig. 8. The circadian clock. The feedback loop between morning-expressed CCA1/LHY, and evening-expressed TOC1 makes up the core oscillator. The relative levels of each are regulated by secondary oscillators, post-translational modifications, 26S-proteasome-mediated degradation, and chromatin remodeling. Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions.

movement was not obtained until 1729. French astronomer Jean Jacques d'Ortous de Mairan found movement, evidence that these occurrences were controlled by an internal mechanism that the cyclical leaf openings and closings of *Mimosa pudica* (heliotrope plant) were not only observed in the presence of cues from sunlight, but that the rhythms persisted even in plants placed in the constant darkness of his basement (93-95). Though the endogenous nature of the cyclical leaf movements was not fully recognized at that time, it marked the beginning of what would become the now widely studied field of plant circadian biology. The rhythmic nature of plant processes would continue to attract others to the study, including Linnaeus, Charles Darwin (96) and Erwin Bunning, and many physiological manifestations of the clock combined with genetic approaches in several plant varieties would be used to study clock characteristics such as period length, heritability, and temperature dependence throughout nineteenth century. However, it was not until advent of clock related forward genetics in the 1970s and 80s that several molecular tools would be developed that would markedly change the way circadian clocks could be studied (94).

One of the most pivotal discoveries in the modern study of plant rhythms came after the observation that certain transcripts accumulate in a circadian fashion. Specifically, it was demonstrated that the mRNA abundance of certain light-harvesting chlorophyll a/b binding proteins (CAB) is regulated by the circadian clock (94, 97, 98). Moreover, it was determined that a small region of the *Arabidopsis CAB2* promoter was adequate to confer this rhythmicity (99, 100). This information was harnessed to develop one of the most powerful tools available at that time for studying circadian

rhythms in plants: a reporter encoding for firefly *luciferase* expressed under the control of the clock responsive region of the *CAB2* promoter (101). Peaks and troughs in luciferase luminescence allowed the oscillations of the clock to be tracked in a manner that paralleled the endogenous accumulation of *CAB2* transcript. Utilization of this new reporter in a population of mutagenized *Arabidopsis* plants allowed the Kay group and others to screen for defects in the circadian clock, and from these initial screens arose the first series of clock mutants at the *timing of cab expression 1* (*toc1*) locus (102). Loss of function (LOF) mutations at this locus result in plants with a significantly shortened period (2-4 hours), while over expression (OX) mutants yield global arrhythmicity. From the severity of these phenotypes it was inferred that *toc1* might play a role close to the core oscillator. The eventual cloning of this gene identified TOC1 as a pseudo response regulator (PRR) protein, characterized by an N-terminal psuedoresponse (PR) domain and a C-terminal *CONSTANS, CO-like, and TOC1* (CCT) motif (103). Additional factors, including the blue light photoreceptors *ZEITLUPE* (*ZTL*) and *LUX ARRYTHMO* (*LUX*), were also identified by the *cab2*::*luc* screen from their aberrant circadian phenotypes (100).

Not surprisingly, however, other clock-associated factors were identified based not on their clock rhythms, but on their flowering phenotypes. Among the first of these were *EARLY FLOWERING 3* and *4* (*ELF3/4*) and *LATE ELONGATED HYPOCOTYL* (*LHY*), initially reported as early- and late-flowering mutants with altered response to photoperiod, respectively (100, 104, 105). Upon closer inspection, *LHY* was found to encode a Myb-transcription factor that, when over expressed due to the specific allele in

*lh*y plants, results in arrhythmic expression of multiple clock-regulated transcripts (104). The clock and flowering phenotypes of this mutant place *LHY* near the heart of the core oscillator alongside *TOC1*, while also reiterating the importance of proper clock function in translating exogenous signals into the appropriate endogenous response.

A third component of the core oscillator, *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), was discovered through a reverse genetics approach, the purpose of which was to identify proteins binding to a certain DNA element present in the promoter of the light-responsive Lhcb1*3 gene (100, 106). The CCA1 protein, a second Myb-transcription factor, was bound to this element, though the role CCA1 plays in the clock was not readily apparent (107). It was eventually determined that the over expression of this gene resulted in clock arrhythmia, and effect was apparent in the expression patterns of all clock-regulated genes assayed.

TOC1, *CCA1*, and *LHY* were placed into the clock model as part of a feedback loop that makes up the core oscillator (100). The global arrhythmicity observed in OX lines for all three genes suggested that they must be required for proper function of the clock at a very upstream level. Further experiments had determined that some level of functional redundancy exists between the parologous Myb-transcription factors LHY and CCA1, and they demonstrated similar morning peaks in their expression patterns (94). It was also found that in CCA1 and LHY OX lines, *TOC1* transcript levels were low, suggesting repression of this gene by CCA1/LHY. The opposite phasing of *TOC1* expression in combination with the finding that it is required for *CCA1*/*LHY* expression allowed it to be placed in a complementary position on the evening half of the loop.

Moreover, these and other experiments revealed an element (the evening element (EE)) in the *TOC1* promoter that is also present in many evening-expressed genes, bound by both CCA1 and LHY, suggesting direct repression of *TOC1* by these Myb-factors and solidifying their position in the basic clock feedback loop. Thus, by the end of the 1990s, the first models of the plant circadian clock began to emerge (94).

Armed with a growing suite of clock mutants and working models of the clock, rapid progress in the development of molecular techniques for *Arabidopsis* greatly accelerated the identification of many more clock components and allowed the functions of these components to be characterized on a biochemical level. From these recent studies, we can begin to grasp the complexity required to form a clock that is both robust against environmental changes but retains some plasticity in order to respond to particular developmental cues. The plant circadian clock as it is understood today is comprised of not one, but a series of interconnected feedback loops that, once entrained to environmental cues, retains self-sustaining properties. The most central of these is the original CCA1/LHY and TOC1 loop. The role of the negative arm is played by the partially redundant functions of both LHY and CCA1, while the role of the positive arm is fulfilled by evening expressed TOC1 (94). As each day begins, expression of both *LHY* and *CCA1* reaches a peak. Once activated through various post-translational modifications, including phosphorylation by CASEIN KINASE 2 (CK2), these proteins bind directly to the EE present in the promoter of *TOC1* (108, 109). This binding inhibits the activity of certain histone acetyltransferases (HATs), leaving the chromatin

of the *TOC1* promoter in a hypoacetylated, presumably condensed state that is not conducive for transcriptional activation (110).

As the day continues, CCA1/LHY activates the expression of *PRRs 9/7/5/3* in descending order so that their expression peaks in a sequential fashion (111, 112). These PRRs fall into the same family as TOC1 (PRR1), containing similar PR and CCT motifs, and are related to the *Arabidopsis* response regulators (ARR) (109). Though they lack the necessary aspartate residue for participation in the classical ARR phosphorelay, these proteins serve both to maintain the rhythm of the clock as well as acting as molecular liaisons between the core oscillator and multiple downstream pathways. Moreover, PRRs 9 and 7 transcriptionally repress *CCA1*/*LHY*, forming a second feedback loop within the context of the original clock (Fig. 8) (94, 113). CCA1/LHY also participate in a third loop, negatively repressing the expression of *CCA1 HIKING EXPEDITION* (*CHE*), which itself acts to repress transcription of *CCA1* (114). The combinatorial negative effects of CHE and the PRRs on the expression of these genes cause a decrease in their transcription throughout the day.

As expression of LHY and CCA1 begins to diminish, the repression of *TOC1* is alleviated and expression of this gene begins to increase, peaking in the evening. TOC1 fulfills the role of the clock's positive arm, and as such is required to begin the reactivation of *LHY/CCA1* transcription that closes the loop and marks the start of a new cycle. The mechanism by which this occurs is not entirely understood; its action is currently modeled through an unknown factor "X," which may mediate TOC1's activity on *CCA1/LHY* (115). However, one way in which TOC1 may also act is by binding to

the CHE protein, preventing this factor from repressing *CCA1/LHY* transcription. TOC1 also participates in a fourth feedback loop, regulating its own expression through "Y," another unknown component or set of components required for accurate modeling of the clock *in silico* (115). Additionally, the stability of the TOC1 protein is modulated via an interaction between GIGANTEA (GI) and ZTL, which functions as a blue light-dependent F-box component of SKP CULLIN F-BOX (SCF) E3 ubiquitin ligase (116, 117). When GI and ZTL interact, the SCF complex is recruited to and marks TOC1 for degradation. However, turnover of TOC1 via this mechanism is inhibited through PRR3, which binds to and protects TOC1 from degradation (118). This interaction is dependent on the night-specific phosphorylation of both TOC1 and PRR3, adding another layer of complexity to the regulation of this clock protein. Once an appropriate level of TOC1 is achieved during the night, it allows transcription of *CCA1/LHY* to begin again, closing the loop and beginning a new cycle.

In accordance with de Mairon's original experiments in mimosa, the circadian nature of the clock allows it to continue cycling even when deprived of external stimuli. However, external cues play important roles in entraining the clock to a cycle that is suitable for the environment and modulating the responsiveness of the clock to stimuli during particular times of day, a phenomenon referred to as "gating (94, 119)." Temperature stimuli are involved in this process, though the effects of light exposure have been characterized somewhat more thoroughly. Though the effects of circadian gating can be observed readily, the molecular basis for this phenomenon is still not well known. Of course, the environmental light cues are received by photoreceptors, which

transduce the signal to clock components. Multiple photoreceptors are involved in the daily resetting and maintenance of the clock (120). Red and far red sensing phytochromes (PHYA – PHYE in *At*) and blue light responsive cryptochromes (CRY1,2), and phototropins (PHOT1,2) receive these signals and modulate the expression and activity of specific downstream genes, particularly those involved in photomorphogenesis, phototropism, photosynthesis, and, most importantly for flowering: the photoperiod response.

FLORAL INITIATION IN *ARABIDOPSIS*

Flowering in response to photoperiod in *Arabidopsis*, a long day plant, occurs via what is referred to as the external coincidence model (Fig. 9*A*) (121). As the internal rhythms of a plant's molecular components cycle, the coincidence of light with various phases of this clock allows the plant to "tell time" and respond to seasonal cues (Fig. 9*B*). Photoperiodic flowering via this model is controlled largely through regulation of CONSTANS (CO), a CCT domain protein (122), in the phloem companion cells of the leaf, which then directly up regulates a mobile "florigen" signal, which travels to the shoot apical meristem (SAM) and induces reproductive growth (123). CO expression is regulated in a circadian fashion so that mRNA accumulates in both long and short days. However, in long days, this expression is coincident with light (124). The CO protein is only stable during the day, making this concurrent expression required for floral initiation. The modulation of CO at several levels confers the differential response to flowering under inductive and non-inductive periods. As the day begins, CO

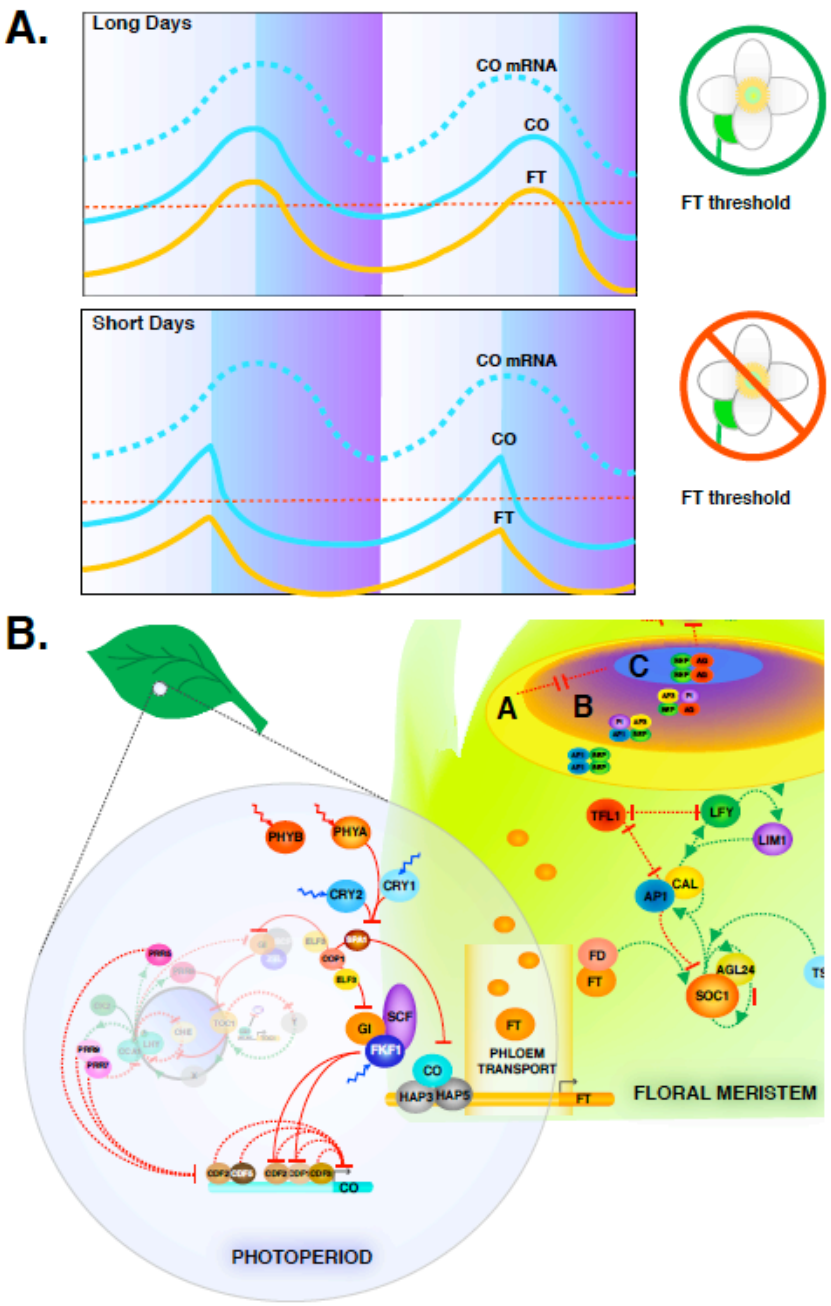

Fig. 9. The photoperiod pathway. Flowering in response to photoperiod occurs via the external coincidence model (*A*). In this model, the expression of *CO*, the cornerstone of photoperiod sensitivity, is tightly regulated by multiple mechanisms so that it can only up-regulate *FT* to inductive levels in LD conditions (*B*). Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions.

transcription is repressed by the binding of at least four zinc finger transcription factors, specifically CYCLING DOF FACTORS 1-3, and 5. (CDF1-3, 5) (125, 126). This repression is lifted in the afternoon of long days via two distinct mechanisms. The first of these is the inhibition of expression of CDFs 2, 3, and 5 by PRRs 9,7, and 5 (127). The expression of these *PRR*s is under circadian control so that high levels of these proteins only occur in the afternoon and evening phases of the clock (112). The transcriptional repression of the *CDF*s therefore occurs only during that time. A second mechanism relies on specific interactions, which ultimately control CDF protein stability. Clock-regulated GI becomes more highly expressed in the afternoon and interacts directly with the FLAVIN-BINDING KELCH REPEAT, F-BOX1 (FKF1) protein, a photoreceptor with ubiquitin ligase activity, in a blue light-dependent manner (128). This complex then binds to CDF1/2, targeting them for degradation by the 26S proteasome (125). The blue light dependency of the GI-FKF1 interaction means that it is stable only in the afternoons of long days, further relieving the repression of CO in that photoperiod. Coupled with afternoon PRR mediated transcriptional repression of *CDF*s, a mechanism emerges by which differential expression of *CO* occurs in long days. Thus, under long day, but not short day photoperiods, *CO* mRNA is allowed to accumulate during a time coincident with light exposure (129). In this point lies the key to the external coincidence model. Though *CO* mRNA is produced even in SD, in the absence of certain light signals the protein is bound and marked for degradation by a complex formed between kinase-like coiled-coil/WD-repeat protein, SUPRESSOR OF PHYA-105 1 (SPA1), and CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1), an

E3 ubiquitin ligase (130). In the afternoons in LD, the blue light dependent action of CRY2 allows it to bind to SPA1–COP1, preventing CO from being targeted by this complex (130). CRY1 also participates in the stabilization of CO in LD, along with far red-dependent PHYA, though PHYB antagonizes this action (131). However, the PHYB-mediated destabilization of CO protein occurs primarily in the morning, while CRY1, CRY2, and PHYA act only under late afternoon light conditions (131). Therefore, if the expression of CO is not concurrent with afternoon light, it cannot transmit a floral induction signal.

The biochemical mechanism by which CO functions has only recently begun to be explored. The CO protein contains several B-box zinc fingers and is part of a large family that possesses CCT domains, and this group is enriched for genes whose products are involved in the clock and developmental processes (132). This domain is required for downstream function of CO, and it had been proposed that it facilitates interactions among other members of the family. Among these members are the Heme Activating Proteins (HAP, also NF-Y), which are further categorized in to HAPs 2, 3, or 5 (NF-Y A, B, or C) (133). These interact to form a HAP2/3/5 heterotrimeric complex, which then acts to regulate transcription of multiple CCAAT box-containing genes. In the *Arabidopsis* genome, there are 36 HAP-encoding genes, and the variation and interchange between members may allow for finer transcription regulation (134). Moreover, the CCT domain of CO has the most similarity to that found in HAP2, and it has been shown that CO most likely competes with this HAP for binding to the HAP3/5 heterodimer in both *Arabidopsis* and wheat (53, 132). Ultimately, the differential

function and stoichiometry of the individual HAP2/3/5 and CO/HAP3/HAP5 complexes may modulate the function of CO, regulating its binding to its direct downstream target, *FT*. In addition to the competition for HAP complex binding, competition may exist for the binding sites at the *FT* promoter. *TEMPRANILLO 1* and *2* (*TEM1/2*), members of a RAV transcription factor family, bind directly to the promoter in sites that are predicted to be CO sites as well (135). Thus, the relative levels of the TEM1/2 and CO can determine whether sufficient FT is produced to initiate reproductive growth, and inductive photoperiods ensure that CO levels are high enough to promote this. In this mechanism, it has also been proposed that GI may have yet another role in floral activation: this protein binds to repressors TEM1/2 and SVP, likely inhibiting their activity (136).

The shoot apex, in the absence of floral inductive signals maintains indeterminate growth through a set of spatially regulated homeodomain genes. Just below the topmost layers of the meristem, the expression of *SHOOT MERISTEMLESS* (*STM*) and *WUSCHEL* (*WUS*) promotes indeterminacy in vegetative, inflorescence, and floral meristems (137, 138). *WUS* expression is localized to a very specific space through a CLAVATA (CLV) / WUS feedback loop. WUS enhances the expression of *CLV3* in the topmost cell layers. This signal acts as a ligand for the CLV1/2 heterodimer, which functions as a repressor of *WUS* (139). This feedback loop contains the indeterminacy signal to several cells, maintaining the meristem while protecting against its over-proliferation. However, upon the initiation of flowering, the meristem must undergo significant changes in morphology that require a complete overhaul of transcriptional

programming.

Upon transitioning to reproductive growth, input from multiple external stimuli and internal developmental programs, including *FLC* from the autonomous and vernalization pathways and GA-responsive genes, converge on *FT* and *SOC1*, and their expression integrates messages from multiple sources into one flowering signal. FT, transported through the vasculature from the leaf to the shoot apex, interacts with FLOWERING LOCUS D (FD) in order to up-regulate several genes in the floral pathway, including *SOC1* (140). Activation of *SOC1* via CO-FT in leaves represents input from the photoperiod pathway, however, age, vernalization, and hormones also play important roles in its regulation in the apex. In the earliest stages of floral development, SOC1 accumulates in the apex and forms a positive transcriptional feedback loop with AGAMOUS-LIKE 24 (AGL24). SOC1 and AGL24 also interact physically, together up regulating another MADS-box protein, LFY, which transduces the floral initiation signals received from SOC1/AGL24 into physiological changes in meristem determinacy (140, 141). LFY then forms a positive feedback loop with APETALA1 (AP1) and partially redundant CAULIFLOWER (CAL), MADS proteins positively regulated through FT and SOC1 (140, 142). However, the expression of *AP1/CAL* and *LFY* must be limited to the floral meristems that arise on the flanks of the inflorescence meristem. If *AP1/CAL/LFY* are expressed inappropriately, only a single flower will be produced. Therefore, the expression is antagonized by *TERMINAL FLOWER 1* (*TFL1*), an FT-like gene which promotes meristem indeterminacy and inhibits the commitment to floral programming (143). AP1 and LFY also act to repress

TFL1 expression (143, 144). Thus, as more and more *LFY* expression occurs via SOC1/AGL24, and more *AP1* is expressed through LFY and FT, the balance shifts from a TFL1-dominated inflorescence meristem to one where *AP1/LFY* are predominately expressed. When a sufficient amount of LFY and AP1 have accumulated, the floral meristem makes a commitment to terminal organogenesis and the expression of a suite of homeodomain genes are regulated in a temporospatial manner resulting in floral organ initiation (145). The development of reproductive structures requires a substantial amount of organization, and in the first phases of this switch, AP1, along with partially redundant CAULIFLOWER (CAL), orchestrates this process (142). AP1 ultimately specifies the identity of certain organs, but initially, this protein plays an essential role in switching off meristem indeterminacy by acting as master transcriptional repressor on a genome-wide scale. AP1 has nearly 2000 binding sites throughout *Arabidopsis* genome (146), and surprisingly, AP1 initially represses its own activators, including FD, SPL9, and SOC1/AGL24. However, it also turns off genes that repress flowering and promote indeterminacy, including *AP2* and *AP2*-like genes *TOE1-3*, *SNZ*, and *SMZ*, along with *TEM1/2* (146). Once the commitment to flowering has been made, AP1 does not necessarily act to repress indeterminacy, but instead begins to up regulate organ identity genes. The complexity of the AP1-regulated transcriptional program suggests that AP1 plays many roles in the switch to reproductive growth. However, ultimately this factor is necessary to set both the timing and patterning of the flower.

The development of reproductive structures takes place according to the ABC model of floral organogenesis elucidated largely through the works of the Meyerowitz

group (Fig. 10) (145, 147). In this model, the expression of A-, B-, and C-identity MADS homoeodomain genes is confined within each of a series of four concentric circles, or whorls, and the organs are produced from the outside in. In the outermost whorl, the expression of A-identity genes *AP1* and *AP2* result in the formation of sepals. The innermost whorl is marked by the expression of the C-identity gene *AGAMOUS* (*AG*), activated by LFY and WUS, which gives rise to a carpel, the female reproductive

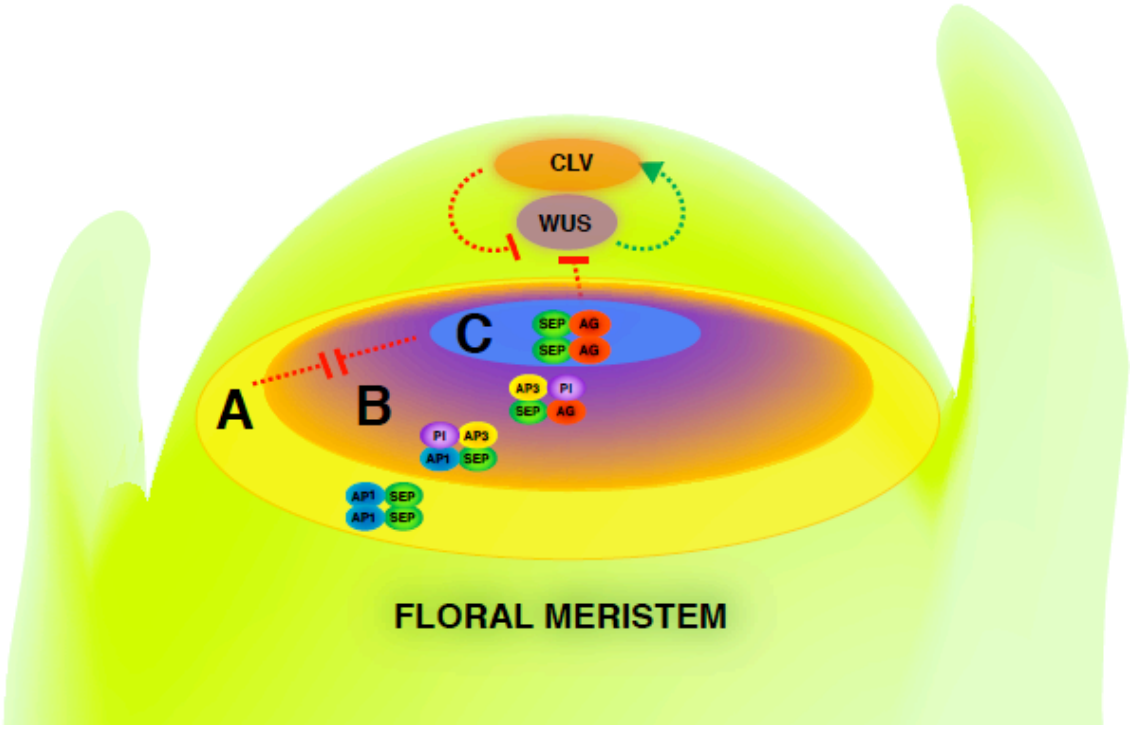

Fig. 10. The ABC model of floral organogenesis. A-type and C-type genes are expressed in the two outermost and innermost whorls, respectively. B-type genes overlap A-type and C-type in the central two whorls, creating four distinct zones of gene expression. These give rise to sepals, petals, stamen, and carpels, from the outside in. Dotted lines denote transcriptional activation (green) or repression (red).

organ. Once AG is expressed, it begins to inhibit *WUS* and thus terminate indeterminate meristem activity (144, 145). Moreover, AG has been shown to bind to GA biosynthesis promoter elements where GA is required for proper organ formation, again integrating hormones into the floral development pathway (148). These A and C genes mutually antagonize the expression of one another, defining the region where these organs can develop. Between the sepals and the carpel in the two central whorls, LFY, AP1, and UNUSUAL FLORAL ORGANS (UFO) activate the expression of B-identity genes *AP3* and *PISTILLATA* (*PI*), allowing for the specification of the remaining structures (144). In the outermost of these two, petals are produced where A and B gene expression overlaps, while the second-most central whorl gives rise to stamens through the interaction of B and C genes. The function of each class of genes in the original model was determined elegantly through the observation of phenotypes found in mutants of each kind (147). When A-identity genes are eliminated, C gene expression expands to cover the entirety of the floral primordia and will produce carpels in the outermost and innermost whorls, with stamen in the central two as a result of B-gene expression (149). When the opposite is true and C genes are lacking, sepals and petals form not only in the outer whorls, but also in the place of stamen and carpels. Though the roles of each class of genes has been well defined, the identification of D and E-identity genes adds another layer of complexity to this model (149). D-class genes, including *SEEDSTICK* (*STK*), and *SHATTERPROOF1* and *2* (*SHP1* and *2*) primarily confer ovule identity in combination with the carpel identity gene AG (150). The E class, comprised of *SEPALLATA 1-4* (*SEP1-4*), act in conjunction with other classes to confer organ

specificity (151, 152). *SEP1* and *2* are expressed in all regions of the primordia, while *SEP3* and *SEP4* are expressed in the more inner whorls (144). The mechanism by which the E class *SEP* genes are proposed to work in combination with the remaining classes is referred to as the Quartet Model (144, 153). In this model, the SEP proteins bind the members of A, B, and C class proteins in various combinations, and subsequently function as a complex that can bind to the promoters of target genes. In the A domain, a pair of SEP proteins is predicted to bind an AP1 homodimer; likewise, in the C domain, AG homodimers can bind two SEP proteins. In the two central whorls, SEP proteins are thought to bind to the B class proteins AP3 and PI, but the fourth member of the quartet is dependent on whether an A (AP1) or a C (AG) class gene is expressed. The varying combination of MADS transcription factors in each whorl confers DNA binding specificity such that only the target genes required for formation of the respective organ are activated.

In addition to modulation of floral homeodomain genes via protein-protein interactions, some of these players are regulated through post-translational mechanisms. Notably, the processing of *AG* mRNA is controlled through HUA1 and 2, as well as HUA ENHANCERs 2 and 4 (HEN2 and HEN4) (38, 154). It is proposed that the activity of HUA1 and 2 results in proper splicing and polyadenylation of the *AG* pre-mRNA, though the details of this mechanism are not known. A second set of genes, *HEN1* and *PAUSED* (*PSD*) represent an miRNA methyltransferase and an exportin-like protein, respectively, are also involved in *AG* mRNA production, though *AG* is not the only target (155). However, the many layers of regulation involved ultimately

demonstrate that flower formation is an incredibly complex phenomenon. Input from multiple signaling pathways is integrated to initiate reproductive growth at the proper time, and subsequent temporospatial regulation of very specific sets of homoeodomain genes that will eventually give rise to floral organs is required. This process has been studied extensively in *Arabidopsis*, however, questions about these molecular details are still, to varying degrees, unanswered in many agronomically important species. Of these, rice (*Oryza sativa*) has been studied most extensively, and though many similarities exist between rice and *Arabidopsis*, differences exist from the protein function level to the level of whole pathways.

PHOTOPERIODIC FLOWERING TIME CONTROL IN *ARABIDOPSIS* AND THE CEREALS

Photoperiodic flowering control in rice and several other cereals involves conserved genes that have a similar role to those in *Arabidopsis*, as well as a subset of novel genes and orthologous genes with novel functions (Fig. 11), and these small differences ultimately add up to the difference between a long day and a short day plant. The most central component of *Arabidopsis* photoperiod signaling, the *GIGANTEA-CONSTANS-FT* module, is well conserved in rice, however, with one important difference; the rice *CO* ortholog, *HEADING DATE 1* (*Hd1*), can act as a repressor (156). In both LD and SD, OsGI induces the expression of *Hd1*, and in SD, *Hd1* up-regulates *HEADING DATE 3A* (*Hd3a*), the rice *FT* ortholog (157). Conversely, in LD,

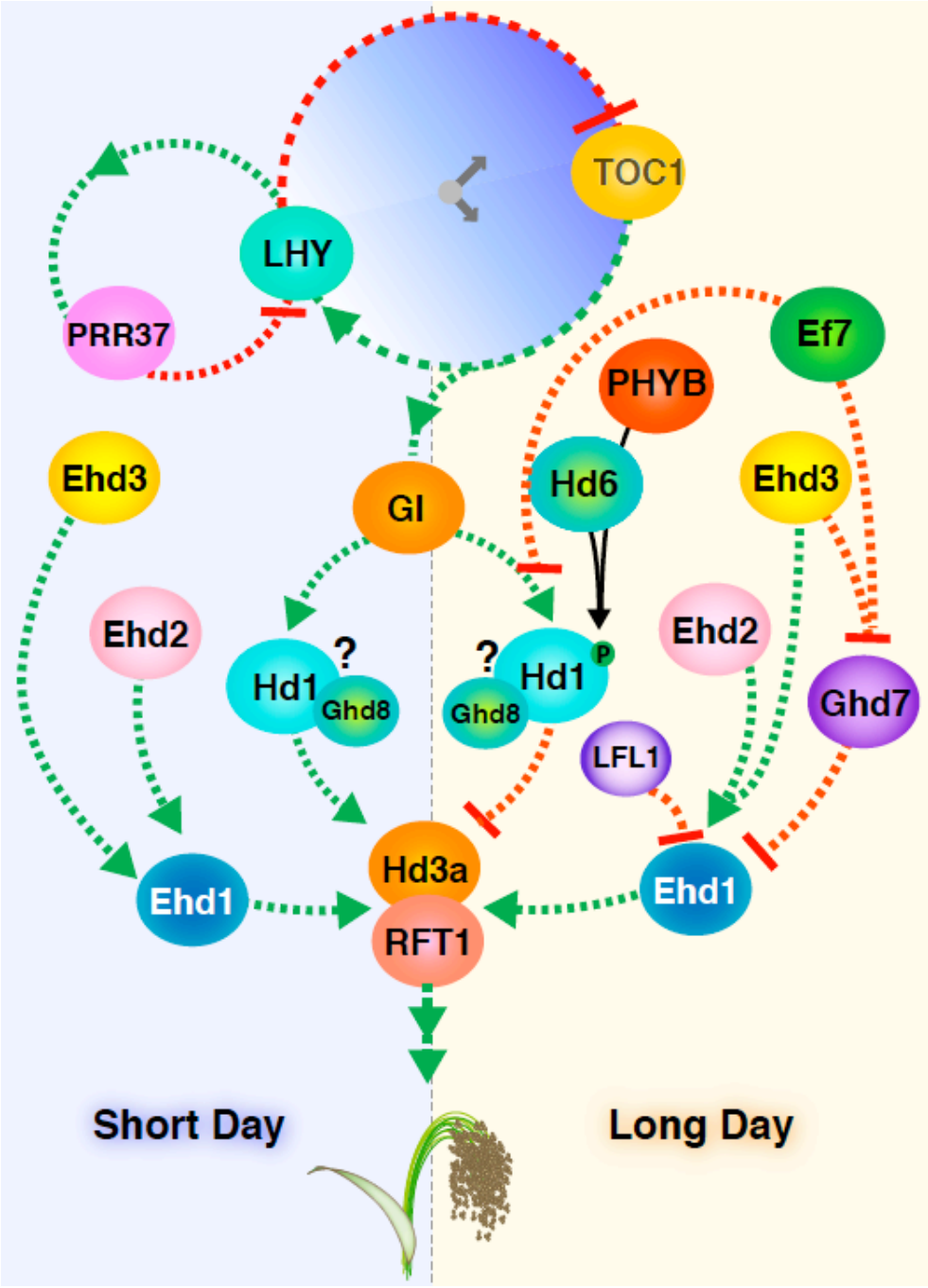

Fig. 11. Photoperiodic flowering in rice. In short days, florigens FT and RFT1 are up-regulated by both the Hd1 and Ehd1 pathways. In LD, these inductive mechanisms are reversed via post-translational modifications (Hd1) and transcriptional repression (Ehd1). Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme denote physical interactions.

phosphorylation of Hd1 by Hd6 (CK2α) results in a polarity switch that turns Hd1 into a repressor (158, 159). This phosphorylation is phytochrome-dependent, as illustrated by the switch in function of Hd1 from a repressor to an activator in LD in *photoperiod sensitivity 5* (*se5)* plants (159), which are phytochrome-deficient mutants. The activity of CO in LD and SD may also occur in conjunction with *Ghd8*, which encodes a protein similar to *At*HAP3 (160). These types of proteins have been shown to bind CO in *Arabidopsis* and wheat (53), and though it is not known if Hd1 binds Ghd8 in rice, it can be hypothesized that this may occur. In fact, the *ghd8* plants display late flowering in SD and early flowering in LD, which is consistent with the phenotypes observed in *hd1* plants, further suggesting an interaction (160). Moreover, this switch is reinforced by the expression patterns of *Hd1*. In SD, *Hd1* peaks at midnight, when phytochrome signaling cannot induce the change in Hd1 polarity, in contrast to LD, where the transcription of *Hd1* occurs in during daylight (161). However, rice is a facultative SD plant, and so despite the repressive function of Hd1 in LD, rice plants can eventually flower in under these conditions. This is due primarily to an *Hd3a* paralog, *RICE FLOWERING LOCUS T 1* (*RFT1*). RFT is an activator of flowering in LD, and its expression is regulated through a pathway unique to rice and other tropical cereals. Moreover, it functions similarly to Hd3a in that it is a mobile signal expressed in the leaves and transported through the vasculature to the apex (162). In the RFT1-dependent flowering model, *OsMADS50*, a homolog of *AtSOC1* with different function, and *EARLY HEADING DATE 1* (*Ehd1*), a B-type response regulator with no orthologous

gene in *Arabidopsis*, induce the expression of *RFT1 (91, 163)*. This *Hd3a*-independent pathway allows rice to flower, if delayed, in non-inductive conditions.

In addition to the up-regulating *RFT1* in LD, Ehd1 also up-regulates *Hd3a* in SD via a distinct pathway that does not exist in *Arabidopsis*. The expression of *Ehd1* is strongly repressed in LD by Grain number, plant height, and heading date 7 (Ghd7) and LEC2 and FUSCA3 Like 1 (LFL1), inhibiting the downstream expression of *Hd3a* (164, 165). However, LD-specific expression of *Ghd7* requires the plant to anticipate day length. *Ghd7* expression is induced via phytochrome signaling, but is subject to gating by the circadian clock so that it can only be activated by light during specific times of the day. It is thought that this gate is set up by *Os*GI, though the exact mechanism by which this occurs is not fully understood (166). In photoperiods perceived as LD, *Ehd1* is repressed by Hd1, in addition to the repression already conferred by Ghd7 and LFL1. This multi-layer repression results in very low levels of expression of *Ehd1*, and thus *Hd3a*, in LD. In SD, repression of *Ehd1* by Ghd7 is attenuated, and the gene is further activated by OsMADS51, Early heading date 2 (Ehd2), a zinc finger protein; and Early heading date 3 (Ehd3), a PHD finger protein; none of which has an ortholog in *Arabidopsis* (91, 167). This allows Hd3a to accumulate to critical levels to initiate flowering. Though this pathway exists in sorghum, it is not common to all cereals, and is, in fact, absent in certain temperate grasses (168). The *Ghd7-Ehd1-Ehd2* pathway confers extreme photoperiod sensitivity, allowing the plant to sense small changes in photoperiod. It can be hypothesized that these additional components arose so that seasonal flowering can be controlled in tropical regions where day length differences are

minimal. As crops dispersed from the equatorial region, they had to adapt to greater differences in day length and temperature, and many varieties grown in higher latitudes have mutations in some component of this pathway (168, 169).

A second source of variation in floral pathways between species lies in the nature of their "florigen." In *Arabidopsis*, the FT-like family consists of a set of six posphatidylethanolamine-binding proteins (PEBP) including FT itself, MOTHER OF FT AND TFL1 (MFT), BROTHER OF FT AND TFL1 (BFT), TSF, ARABIDOPSIS THALIANA CENTRORADIALIS (ATC), and TFL1. Of these, FT acts as the primary inductive signal, acting somewhat redundantly with TSF and MFT, while TFL1 represses floral meristem formation (170, 171). The functions of BFT and ATC have yet to be established. In monocots, this family has generally expanded, and in rice it includes about 19 members, though Hd3a and RFT1 are the major promoters of the floral transition (172). Though sorghum is closely related, an RFT1 ortholog has not been found in its genome.

Perhaps one of the most striking PEBP families is the one present in maize. This group has been expanded to include 25 *Zea Mays CENTRORADIALIS* (*ZCN*) genes that are further categorized into *FT*-like, *MFT*-like, and *TFL1*-like (34). However, only a few of these are thought to play a role in the floral transition. By analyzing expression patterns and binding partners of a subset of these proteins, it was determined that ZCN8 is the most likely candidate for maize florigen. Its expression is photoperiod responsive, and it is localized to the leaves during the proper time for floral initiation (34, 173). Moreover, it was shown to bind FDL1 (the ortholog of FD), consistent with FT signaling

in *Arabidopsis* (34). *ZCN8*, as well as some others from this family are represented in the sorghum genome, though these have not been studied as extensively as in maize. In fact, though day length-dependent flowering in sorghum is the primary factor in determining its maturation time, much is left to be elucidated in this species and other cereals.

DISSERTATION OVERVIEW

Many of the flowering pathways present in *Arabidopsis* are conserved in the temperate grasses as well as in sorghum, a tropical species, including age-dependent flowering, and flowering in response to hormone signaling. By contrast, in sorghum, other major pathways that play a large role in the initiation in *Arabidopsis*, such as vernalization, do not influence flowering. In fact, floral initiation in sorghum is controlled primarily though photoperiod. Photoperiodic flowering in sorghum has long been studied from an agronomic perspective at the physiological and genetic level, however; molecular data is not as readily available. Sorghum has orthologs of major components of the *GI-Hd1-Hd3a* pathway in rice, though it lacks RFT1. The *Ghd7-Ehd2-Ehd1* module is also present in sorghum, as is at least some portion of the *ZCN*/*FT*-like family found in maize, though sorghum-specific details surrounding these mechanisms are not known. Studies in *Arabidopsis*, rice, maize, and other cereals have provided a framework by which to compare sorghum flowering genes and the potential role of the six maturity loci. Aside from Ma_3, identified as phytochrome B (24), it is not known how the maturity loci may fit into this framework. All of the loci are flowering

repressors under LD conditions, and may be either the equivalent of a rice or maize gene, or something novel. It is important to determine how sorghum commits itself to the reproductive phase in order to optimize breeding efforts and yields, as well as providing more information about flowering in short day plants. Therefore, the central theme of my research was to identify certain maturity loci, and place them into a sorghum-specific photoperiodic flowering network.

In Chapter II, Ma_1, the most influential of the maturity loci in the repression of sorghum flowering, was elucidated as the *PSEUDO-RESPONSE REGULATOR 37* gene using a map-based cloning approach. Sequencing of this gene in multiple accessions revealed various mutations at this locus, suggesting the presence of an allelic series within sorghum germplasm. The expression of *PRR37* is controlled by not only the internal circadian clock, but also by external light cues from the environment, providing a mechanism by which photoperiod can control flowering that is consistent with the external coincidence model. Furthermore, the expression of multiple downstream flowering genes, including *CO*, *Ehd1*, *FT* (*OsHd3a*), and *ZCN8* are modulated by PRR37. The results of this study provide a novel perspective on flowering in SD grass species within the context of the models proposed for LD species, as well as a means by which breeders may more efficiently modulate flowering time to achieve higher yields from all varieties of sorghum.

In Chapter III, Ma_6, a recently identified maturity locus that interacts genetically with the Ma_1 system to greatly delay flowering, was elucidated as the ortholog of the rice *GRAIN NUMBER, PLANT HEIGHT AND HEADING DATE* 7 (*Ghd7*) gene using a

map-based cloning approach. Sequence analysis from diverse germplasm including grain, sweet, and bioenergy varieties revealed that the functional allele was present in all genotypes showing extremely late flowering. Several recessive alleles were associated with a decrease in flowering time in multiple varieties. Moreover, tracing the alleles of this gene through certain founder genotypes of the sorghum conversion program suggests that ma_6, though not identified until 1999, was co-selected with recessive alleles of ma_1 by grain sorghum breeders who favored earlier flowering lines. The expression of *Ghd7*, mirroring that of *PRR37*, is controlled by both the circadian clock and light, providing a mechanism by which these two genes act in conjunction to delay flowering to varying degrees depending on day length. The expression of *CO* is affected primarily by PRR37, however *Ehd1*, *ZCN8*, and *ZCN12* are repressed to a greater extent in LD in varieties with dominant alleles of both genes than varieties with *PRR37* alone, indicating that Ghd7 down regulates *FT-like* genes synergistically with PRR37 through the *Ehd1* pathway. The results of this study provide a model by which the two maturity genes act synergistically to repress *FT* and flowering, while also providing novel insight into the selection of early flowering sorghum in the years prior to the initiation of the sorghum conversion program.

CHAPTER II

IDENTIFICATION AND CHARACTERIZATION OF *MATURITY LOCUS 1* (*Ma_1*)*

INTRODUCTION

Sorghum bicolor is an important food source for millions of people worldwide and an excellent energy crop for sustainable production of biofuels (174). Energy sorghum is selected for delayed flowering, which results in increased biomass yields, while grain sorghum is selected for earlier flowering to increase grain yields and facilitate mechanical harvest. The precise control of flowering time is essential for optimal production of either of these sorghum crops, and flowering in this species is primarily regulated by photoperiod. Therefore, because of its critical importance to crop yield, varying degrees of photoperiod sensitivity were selected for by sorghum improvement programs as long ago as the early 1900s (175).

Flowering in *Arabidopsis*, a long day-induced species, occurs via the external coincidence model (121, 161, 176). This model refers to the coincidence between the internal rhythms of the circadian clock and external light exposure. The core clock oscillator, formed by a feedback loop comprised of morning-expressed Myb

---

*Portions of this chapter are reprinted with permission from "Coincident light and clock regulation of *psuedoresponse regulator protein 37* (*PRR37*) controls photoperiodic flowering in sorghum" by Rebecca L. Murphy, Robert R. Klein, Daryl T. Morishige, Jeff A. Brady, William L. Rooney, Frederick R. Miller, Diana V. Dugas, Patricia E. Klein, and John E. Mullet, 2011. *Proceedings of the National Academy of Science of the United States of America*, 108, 16469-16474, Copyright ©2012 by the National Academy of Sciences.

transcription factors CCA1 and LHY and evening expressed PRR protein TOC1, results in clock-dependent expression of downstream genes, including GI, CO, and FT, the primary floral inducer in *Arabidopsis*. Light-dependent protein interactions further modulate the functionality of these proteins such that FT can only accumulate to critical levels under LD conditions (161).

In rice, a SD species, orthologs of GI, CO (Hd1), and FT (Hd3a) act to modulate flowering time, though some, like Hd1, affect flowering in a repressive rather than activating manner (156, 177). Moreover, in certain grasses, additional regulators have been identified that are unique to these species, including *Ehd1* and *Ehd2*, floral activators that have no orthologs in *Arabidopsis* (163, 178). The identification of these genes and others in SD tropical grasses is indicative of the diversification that occurred as plants adapted to varying latitudes and photoperiods.

Within the species, sorghum exhibits quite variable photoperiod sensitivity (175, 179). This variation is largely controlled by a set of genes known as maturity loci, termed Ma_1 through Ma_4, which were identified in through genetic studies beginning in the 1940s (15, 16, 180). The utilization of these loci was essential for the improvement of sorghum in U.S., and recessive alleles of each were used in various combinations in sorghum breeding programs.

More recently, two additional loci, Ma_5 and Ma_6, were identified as late flowering off-types from such programs, which conferred the extremely late flowering desired in bioenergy and forage lines (22). Dominant alleles at any of these maturity loci represses flowering. However, Ma_1 is the most effective of the six, and has the

largest impact on flowering. In fact, mutations in *Ma$_1$* were critical for the early domestication and dispersal of sorghum in Africa (180), and sorghum breeders across the U.S. selected for recessive alleles at this locus. Because of the importance of Ma1 in sorghum improvement, uncovering the identity of the underlying gene has been emphasized. In 1995, the first quantitative trait locus (QTL) for *Ma$_1$* localized this gene to a position on chromosome 6 (20). Since that time, haplotype analysis has been used to understand how recessive alleles of this locus are distributed among agronomically important germplasm (21), but until 2011, the identity of this gene had not been reported. In this chapter, we report the identification of *Ma$_1$* as *PSEUDORESPONSE REGULATOR 37* (*PRR37*), and analyze the relationship of *Ma$_1$* with the transcription of downstream flowering factors, such as *GI*, *CO*, and *FT*. Moreover, the expression of *PRR37* itself is regulated in a manner consistent with the external coincidence model, providing a link between the flowering mechanism in the LD plant *Arabidopsis* with that found in the SD grass *Sorghum bicolor*.

RESULTS AND DISCUSSION

Map-Based Cloning of *Ma_1*

The gene corresponding to *Ma_1* was cloned using three mapping populations derived from genotypes that vary in flowering time due to differences in alleles at this locus (Fig. 12). A previous study localized *Ma_1* to the long arm of chromosome SBI-06 (21), therefore, two mapping populations were created to refine the *Ma_1* region, and a third

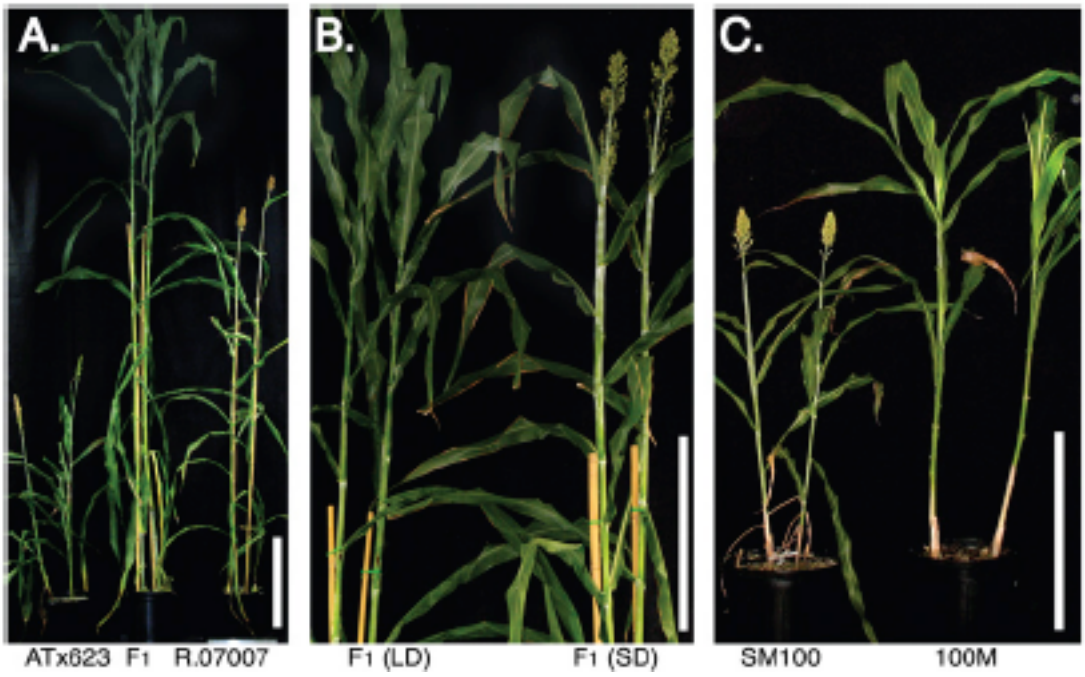

Fig. 12. Phenotypic analysis of *Ma_1*. (*A*) LD-entrained ATx623 and R.07007 flower by 54 and 68 d, respectively; ATx623 x R.07007 F_1 plants remain vegetative for >150 d. (*B*) Flowering is induced in LD-entrained ATx623 x R.07007 F1 hybrids exposed to SD; continued exposure to LD represses flowering. (*C*) In LD, SM100 flowers in 54 d; 100M in 120 d. (Scale bar, 0.5 m.)

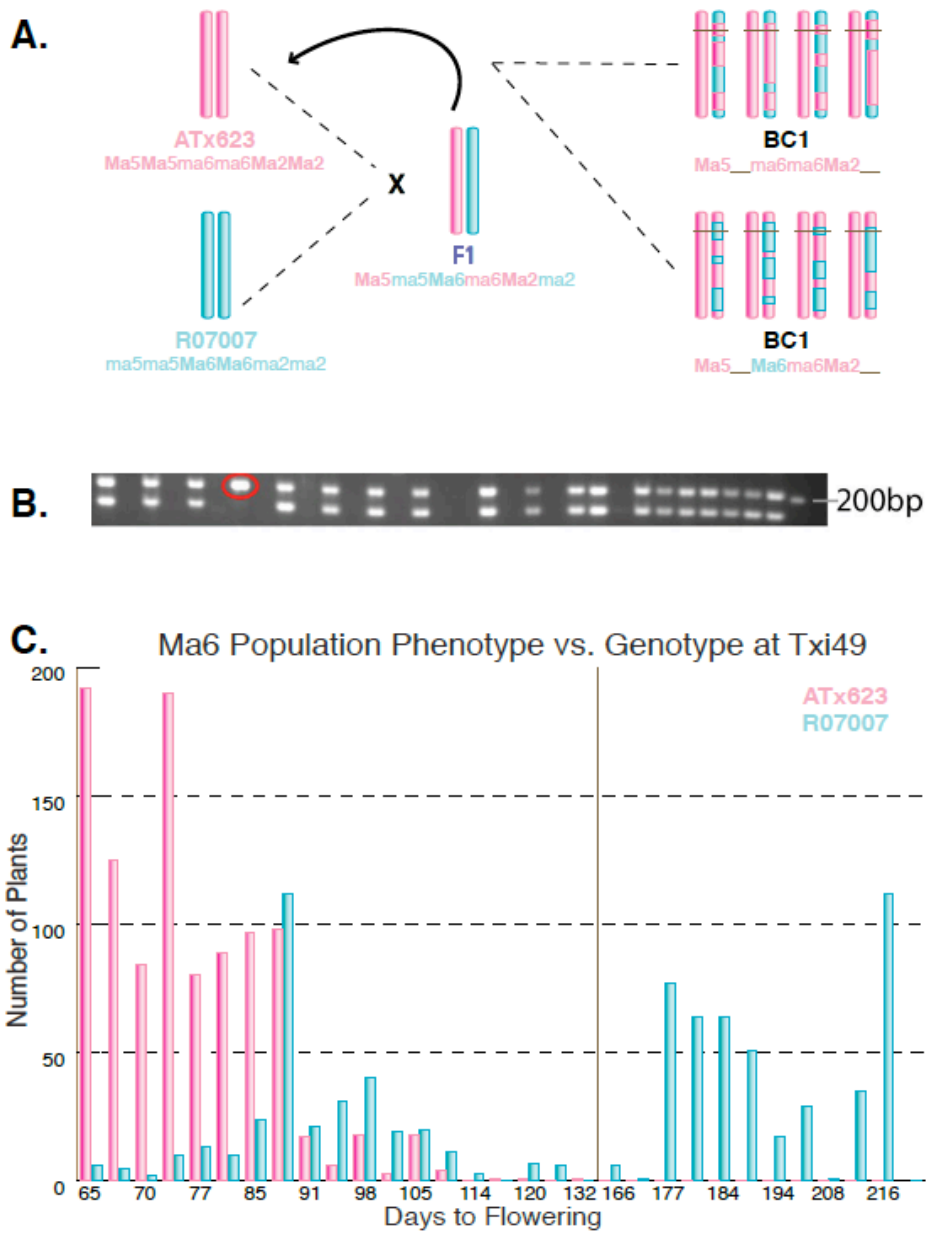

Fig. 13. *Ma$_1$* fine mapping population structure. (*A*) The population structure of the ATx623 x R.07007 BC_1F_1. (*B*) A representative example of the gel-based *Xtxi48* marker used to screen this population (n=1821). A single recombinant was found using this marker, and is circled in red. (*C*) All of the plants that were late flowering (>160 days) were also dominant (R.07007; blue) at *Xtxi49* (at PRR37). Earlier flowering lines that were dominant (<130 days, blue) were used for mapping *Ma$_6$*, as discussed in Chapter III.

population was subsequently created for high-resolution mapping. *Ma_1* was initially fine mapped using a population created by crossing two photoperiod insensitive, early flowering genotypes, ATx623 (*ma_1*, *Ma_5*; 54 Days to Flowering (DTF)) and R.07007 *Ma_1*, *ma_5*; 65 DTF), to generate photoperiod sensitive, late flowering F_1 hybrids that are useful for biomass production (Fig. 12*A*). Exposure of 65-day-old vegetative F_1 plants to SD resulted in flowering 36 days later, whereas F_1 hybrids kept in LD remained vegetative indefinitely (Fig. 12*B*). The sorghum inbred ATx623 was crossed to R.07007, and F_1 plants were subsequently used as the pollinator in a backcross to ATx623 to eliminate allelic effects of *Ma_5* and *Ma_2*, additional flowering loci known to be recessive in R.07007 (Fig. 13*A*) (22).

The resulting population of 1821 plants was screened with a series of PCR-based markers (Fig. 13*B* and *C*), resulting in the identification of a statistically significant region on SBI-06 that mapped coincidently with the reported location of *Ma_1* (20, 21). More specifically, variation in flowering time due to the *Ma_1* locus was linked to marker *Xtxp434* on SBI06 (Fig. 14*A*). Additional markers in this region were used to further define the *Ma_1* region.

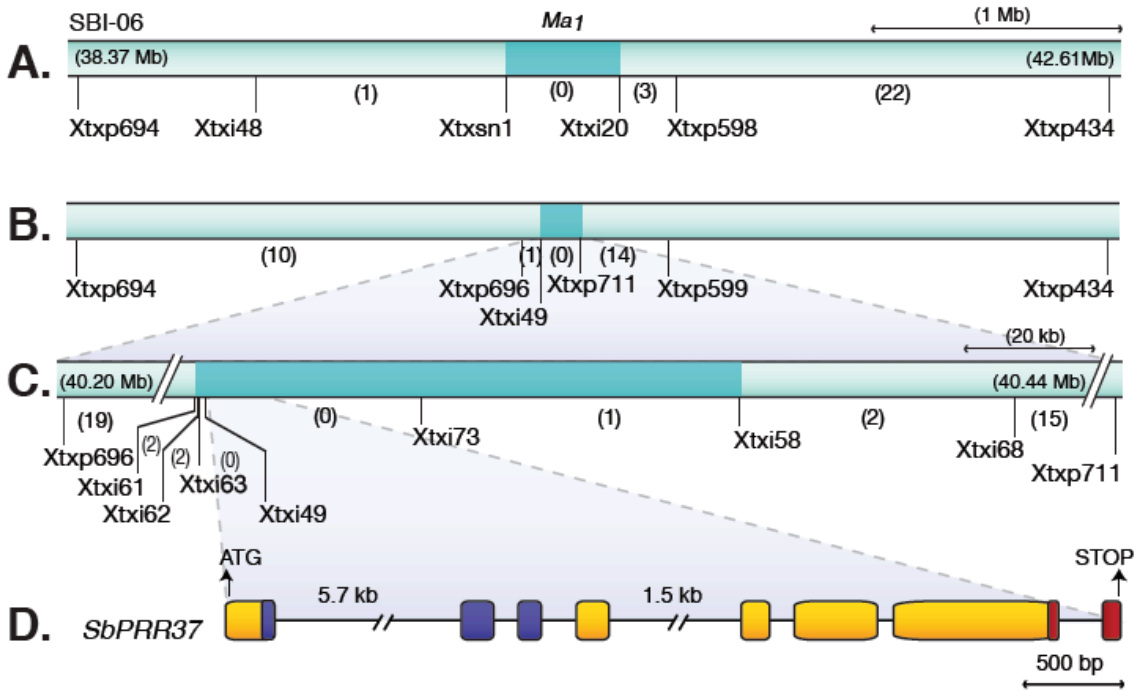

Fig. 14. Fine mapping the *Ma*$_1$ locus. (*A*) *Ma*$_1$ locus delimited to an ~700-kb region between *Xtxsn1* and *Xtxi20* in a BC_1F_1 population (n = 1,821) derived from ATx623 and R.07007 (*B*) *Ma*$_1$ mapped to an ~240-kb region delimited by *Xtxp696* and *Xtxp711* using a population of F_2 plants (n = 122) derived from 100M and BTx406. (*C*) The *Ma*$_1$ locus was refined between markers *Xtxi62* and *Xtxi58*. Recombination events are shown in parentheses, physical coordinates are at the end of each chromosome segment, and the *Ma*$_1$ locus is shaded in blue. (*D*) *SbPRR37* is the only gene present in this interval.

Table 1. Genes present in the ~700-kb interval mapped in the ATx623 by R.07007 BC_1F_1 population.

Gene no.	Start (bp)	End (bp)	Annotation (Phytozome v5.0)
Sb06g014410	39,903,279	39,906,744	Similar to HCF106 family protein
Sb06g014420	39,918,495	39,921,352	Leucine rich repeat protein/F-box
Sb06g014430	39,948,142	39,953,971	no functional annotations
Sb06g014440	39,962,396	39,969,146	Transport protein particle (TRAPP) component
Sb06g014450	39,970,615	39,972,162	Berberine and berberine like, D-lactate dehydrogenase, FAD binding, oxidoreductase
Sb06g014460	39,994,622	39,997,657	Divergent CCT motif
Sb06g014470	40,037,697	40,043,627	Ribosomal protein L4/L1 family
Sb06g014480	40,047,428	40,054,072	Poly A polymerase head domain, tRNA nucleotidyltransferase activity
Sb06g014490	40,079,857	40,082,075	no functional annotations
Sb06g014500	40,082,952	40,085,195	Putative helicase related
Sb06g014504	40,102,573	40,105,282	Ulp1 protease family, C-terminal catalytic domain, Sentrin/SUMO-specific protease
Sb06g014508 (**Xtxsn1**)	40,133,864	40,693,944	General transcription factor 2-related zinc finger
Sb06g014510	40,175,447	40,175,656	no functional annotations
Sb06g014520	40,193,736	40,195,230	no functional annotations
Sb06g014530	40,210,136	40,211,600	Ribosomal protein L11; RNA-binding domain
Sb06g014550	40,216,040	40,217,587	Iron/ascorbate family oxidoreductase
Sb06g014560	40,243,699	40,244,912	Sulfotransferase
Xtxi63	40,279,273		
Sb06g014570 (**txi49**)	40,280,414	40,290,602	PRR37-like
Xtxi58	40,364,576		
Sb06g014580	40,400,791	40,403,993	Serine protease family; S10 serine carboxypeptidase

Table 1. Continued

Gene no.	Start (b)	End (b)	Annotation (Phytozome v5.0)
Sb06g014590	40,406,258	40,401,377	ATP-dependent CLP protease
Sb06g014630	40,434,828	40,436,984	no functional annotations
Sb06g014640	40,437,379	40,440,378	no functional annotations
Sb06g014650	40,460,341	40,462,677	no functional annotations
Sb06g014670	40,493,258	40,498,431	RNA-binding; similar to Mei2
Sb06g014676	40,594,546	40,608,430	no functional annotations
Xtxi20	40,595,306		
Sb06g014680	40,688,839	40,689,723	no functional annotations
Sb06g014710	40,761,789	40,771,377	Glycyl-tRNA synthetase and related class II tRNA synthetase
Sb06g014720	40,771,644	40,776,006	ubiquinol-cytochrome c reductase iron-sulfur subunit-related
Sb06g014730	40,781,189	40,783,917	Extosin family
Sb06g014740	40,784,078	40,784,927	Pollen proteins Ole e I family
Sb06g014743	40,789,072	40,795,991	no annotated domains
Sb06g014746	40,799,353	40,800,508	no functional annotations
Sb06g014750	40,827,904	40,836,238	no functional annotations
Sb06g014760	40,827,904	40,836,238	no functional annotations
Xtxp598	40,828,046		

Markers and genes used as markers are highlighted in bold type. The physical coordinates and functional annotation of each gene are given as based on Phytozome v5.0.

Twenty-two recombination events were detected between *Xtxp598* and *Xtxp434*. An additional three recombinants were found between *Xtxi20* and *Xtxp58*, delimiting the downstream border of the locus. In the upstream region, one recombination event was detected at *Xtxi48*, the breakpoint of which was found in the promoter of putative gene Sb06g014508, defining the *Ma_1* interval to ~700-kb between *Xtxsn1* and *Xtxi20* (Fig. 14*A*). This genomic region is one of low gene density, encoding 34 putative genes Phytozome v5.0 (Table 1). One gene, *PSEUDO-RESPONSE REGULATOR 37* (*PRR37*; Sb06g014570) was identified as a likely gene candidate for *Ma_1* based on the known roles of PRR proteins in flowering in Arabidopsis. Despite the relatively large number of offspring screened, further refinement of the locus was not possible in this population due to the lack of recombination within the genetic region. Therefore, two additional populations were created by Dr. Robert Klein's group to map *Ma_1*, using several historically important grain-producing cultivars that possess different *Ma_1* alleles. 100M and SM100 are nearly-isogenic lines (NILs) that contain dominant and recessive *Ma_1* alleles, respectively (180), and differ in flowering time by ~60 days when grown in LD (Fig. 12*C*). A mapping population was created by crossing 100M (*Ma_1*) to the elite inbred BTx406, which derives its *ma_1* allele from the same source as SM100, but provides a level of polymorphism more suitable for mapping (180). BTx406 is also of historical importance as the genetic donor of the *ma_1* allele used to convert tropical late flowering sorghum to photoperiod-insensitive cultivars useful for grain sorghum breeding (181). Genetic analysis of this F_2 population (n=122) was performed by screening it with PCR-based markers, refining the *Ma_1* locus to a 240-kb region

delimited by markers *Xtxp696* and *Xtxp711* (Fig. 14*B*). An additional 255 F_2 plants from this population were genotyped and phenotyped, and a total of 16 plants were identified with crossovers in the interval from which F_3 progeny could be derived. This effort resulted in a set of recombinant 100M x BTx406 F_3 progeny that were utilized to further refine the *Ma1* locus. These 100M x BTx406 F_3 progeny were utilized in conjunction with a 100M x Blackhull Kafir F_2 population (n=1925) to define the position of the *Ma1* locus on the high-resolution map (Fig. 14*C*) Blackhull Kafir (*ma1*) is a founder genotype from an ancestral lineage different from 100M (15, 180). Analysis of these F_2 plants in combination with derived F_3 families, allowed the *Ma1* locus to be reduced to an 86-kb interval delimited by markers *Xtxi62* and *Xtxi58* (Fig. 14*C*). The best candidate for *Ma1*, *SbPRR37*, was the sole gene present among the stretches of repetitive DNA found in this region (Fig. 14*D*) (Phytozome v5.0).

The *SbPRR37* Allelic Series

To confirm the identity of the *Ma1* gene as *SbPPR37* and to characterize alleles of this locus that modify photoperiod-sensitivity, full-length cDNAs were sequenced for a select set of founder and elite sorghum cultivars. The structure of *SbPRR37* alleles was examined by aligning full-length cDNA sequences from photoperiod-sensitive (*Ma1*) and -insensitive (*ma1*) parental genotypes to genomic DNA sequences. The 3165 nucleotide *SbPRR37* mRNA from R.07007 and 100M contained three untranslated and eight protein-coding exons (Fig. 14*D*; Fig. 15*A*; Fig. 16). This transcript encodes a 739 amino

acid, ~93kDa protein, that contains a predicted N-terminal pseudo-receiver domain (residues 99-207) and a C-terminal CCT domain (residues 682-727), present in all

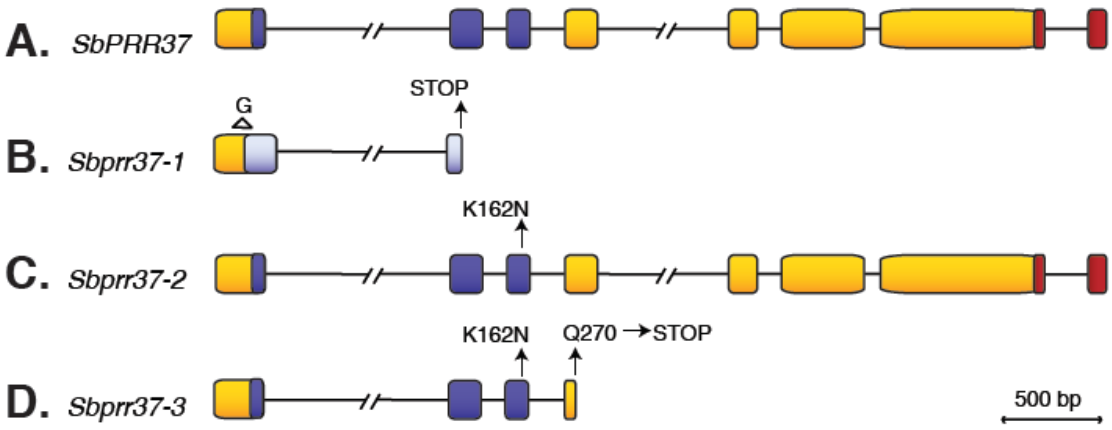

Fig. 15. *PRR37* allelic series. (*A*) Functional SbPRR37 allele in 100M and R.07007. (*B*) Recessive Sbprr37-1 allele from SM100 and BTx406 with a single nucleotide deletion and frameshift upstream of the PRR domain. (*C*) Sbprr37-2 allele from Blackhull Kafir with a missense mutation in the PRR domain at conserved Lys162 residue. (*D*) Sbprr37-3 allele from ATx623 containing both the Lys162Asn substitution and a nonsense mutation at Gln270 resulting in premature termination. Exons are shown as boxes, and introns as solid lines. Yellow boxes, protein coding sequence; blue boxes, pseudoreceiver domain; red boxes, CCT domain; light blue boxes, missense coding post frameshift.

known plant pseudoresponse regulator proteins. Sorghum *PRR37* was compared to other plant PRR proteins by Dr. Patricia Klein using the method described by Turner *et al*. (Fig. 17) (182). This analysis showed that sorghum *PRR37* is most closely related to Arabidopsis *PRR7*, two maize PRR37-like proteins (encoded by GRMZM2G033962 and GRMZM2G005732), rice *PRR37* (LOC_Os07g49460) and PRR proteins encoded by barley *Ppd-H1* and wheat *Ppd-D1a* (182, 183).

Fig. 16. Annotation of the full-length *SbPRR37* cDNA. Exon–intron splice junctions and key amino acid residues are shown. The position of the amino acid substitution is shown above the respective codon. Downward-facing triangles indicate insertion positions of nucleotides or introns, and upward-facing triangles indicate nucleotide deletions. Nucleotides encoding pseudoreceiver domain and CCT motif residues are highlighted in yellow and blue, respectively. Alt. Splice, alternatively spliced. This annotation is based on the sequence of 100M transcript, splice variant a (GenBank accession no. JF801188). Figure prepared by Dr. Robert Klein (184).

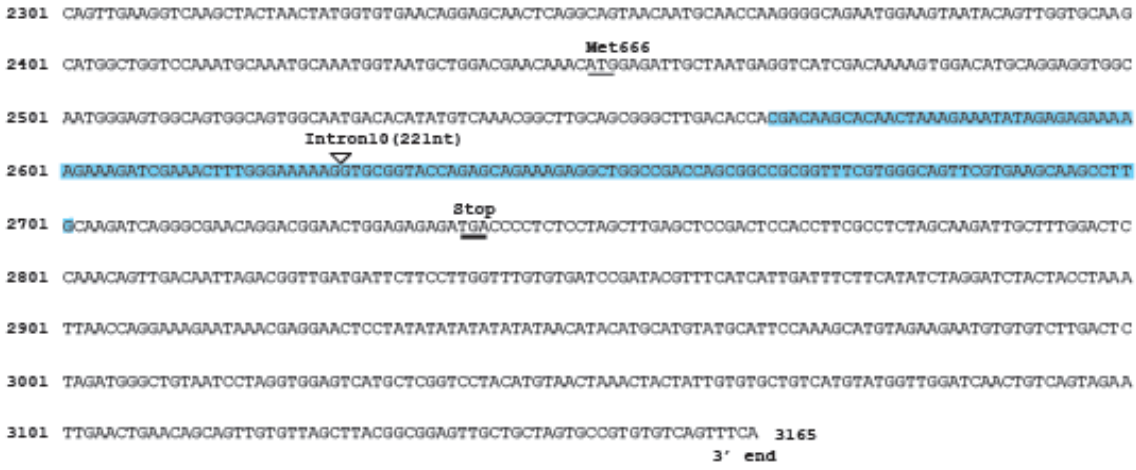

Fig. 16. Continued.

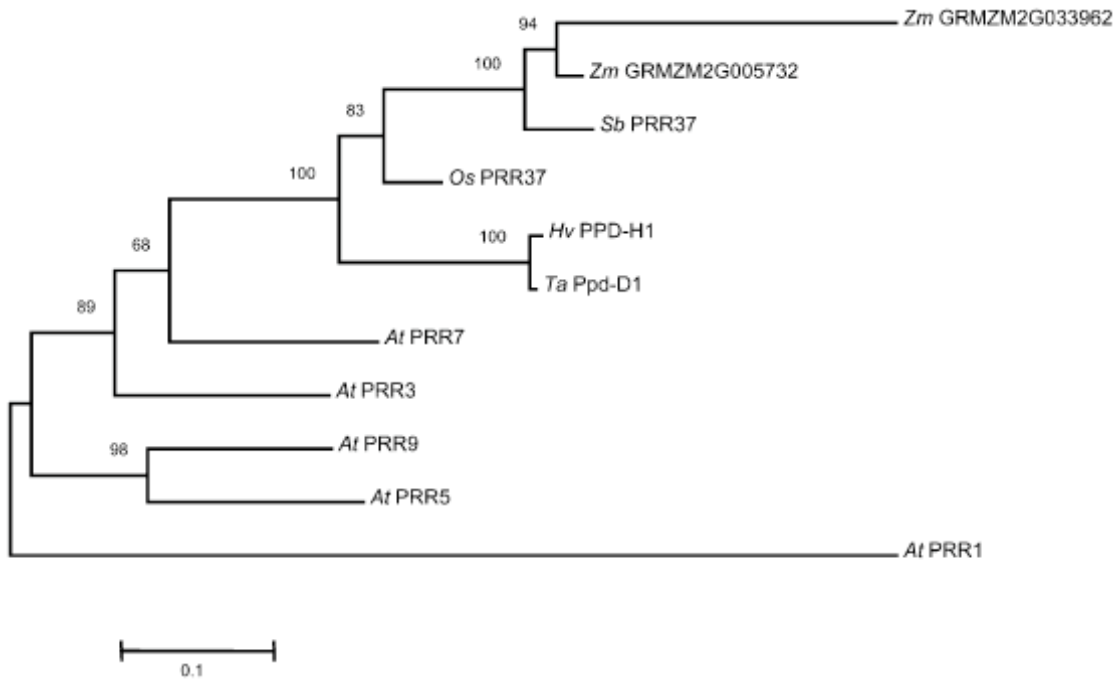

Fig. 17. Similarity of sorghum PRR37 protein to PRR proteins from maize, rice, barley, wheat, and *Arabidopsis*. The neighbor joining tree (185) is based on concatenated pseudoreceiver domain/CCT domain peptide sequences (182). Maize PRR protein sequences GRMZM2G033962 and GRMZM2G005732 were downloaded from Gramene (http://www.maizesequence.org/Zea_mays/Transcript/). Barley Ppd-H1 (AAY42109.1), wheat Ppd-D1 (ABL09464.1), rice PRR37 (NP_001060743), sorghum PRR37 (Sb06g014570.1) and Arabidopsis PRR1 (BAA94547.1), PRR3 (BAB13744.1), PRR5 (BAB13743.1), PRR7 (BAB13742.1), and PRR9 (BAB13741.1) protein sequences were downloaded from the National Center for Biotechnology Information. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches (186). Evolutionary analyses were conducted in MEGA5 (187). Analysis carried out by Dr. Patricia Klein, Department of Horticulture, TAMU (184).

Table 2. Sequence variation within the *SbPRR37* coding region.

Allele[*]	Coding Region Haplotype	Nucleotide Position[†]	Protein Modification[‡]	PRR Domain[¥]	CCT Motif[¥]
SbPRR37[**]	Alternative 3' splice junction - Intron7	1231	QA-insert	No	No
SbPRR37[**]	Alternative 5' splice junction - Intron 8	1376	GTSNRNCMKQKYTN-insert	No	No
Sbprr37-1	G-deletion	608	***Frameshift, Premature termination***	**Yes**	**Yes**
Sbprr37-2	G>T	1006	***K162N***	**Yes**	No
	G>T	1721	D401Y	No	No
	C>A	2270	Q584E	No	No
Sbprr37-3	G>T	1006	***K162N***	**Yes**	No
	C>T	1329	***Q270Stop***	No	**Yes**
	G>T	1721	na[!!]	-	-
	C>G	2270	na[!!]	-	-

[*] Total of 8 founder sorghum genotypes were examined for *Ma_1* coding region haplotypes. Genotypes examined are listed in Supplemental Table 2.
[†] Position in *Ma_1* mRNA based on full-length sequence of sorghum genotype 100M as shown in Fig. 16.
[‡] Protein modifications in ***bold-italics*** type denote mutations that alter the pseudo-receiver domain and/or CCT motif.
[¥] Protein modification that truncate, eliminate, or alter key amino acid residues in the pseudo-receiver domain or CCT motif are denoted as **Yes**.
[**] Alternative splicing variants were observed in wild type and mutant alleles.
[!!] Mutation occurs after premature Stop codon in *ma_1* allele.

The coding sequences from photoperiod-insensitive, *ma_1* genotypes revealed mutations in the PRR37 protein that are predicted to disrupt function (Fig. 15*B-D*) as well as other background nucleotide polymorphisms (Fig. 16; Table 2). The nucleotide

sequence of the coding region from the *Sbprr37-1* allele (genotypes BTx406 and SM100) was identical to *SbPRR37* except for a single nucleotide deletion upstream of the pseudo-receiver domain, resulting in the premature termination of the Sbprr37-1 protein (Fig. 15*B*). Allele *Sbprr37-2* (from cultivar Blackhull Kafir) differs from SbPRR37 by three amino acid substitutions; two substitutions are present in regions of low conservation among PRR37 proteins (Fig. 16; Table 2), but the third substitution occurs in the pseudo-receiver domain at Lys162, a highly conserved amino acid in all pseudo-receiver and receiver-domain proteins (Fig. 15*C*). The substitution of an uncharged Asn for a positively charged Lys at this position could alter the functionality of the pseudo-receiver domain. Recessive allele *Sbprr37-3* from ATx623 harbored both the Lys162Asn substitution found in Sbprr37-2 and an additional nucleotide substitution resulting in an in-frame stop codon (Q270→Stop) prior to the CCT motif (Fig. 15*D*). It had been proposed by Quinby (175) more than 50 years prior that the wide range of flowering times observed among sorghum varieties may be due to only a few large-effect genes that exist as series of unique recessive alleles. The findings from this study support this hypothesis. The *Sbprr37-1* functional mutation occurred in tropical Standard Milo introduced into the U.S. from Columbia in the mid 1800s (1, 180). By comparison, the *Sbprr37-2* allele can be traced to Kafirs from South Africa that were introduced in 1876 (175, 180). The *Sbprr37-3* represents a second Kafir allele originally present in the progenitor cultivar Combine Kafir-60 (1). The full extent of the *Sbprr37* allelic series remains to be determined, but these results suggest multiple independent

mutation events in *SbPRR37* have occurred during sorghum's adaptation to temperate climates worldwide.

Coincident Clock- and Light-Regulated *SbPRR37* Expression

In wheat, misexpression of *Ppd-D1a* is correlated with reduced photoperiod sensitivity, indicating the importance of *PRR37* expression in photoperiodic regulation of flowering in this LD-grass (183). Therefore, in order to understand the significance of its expression pattern in sorghum, *SbPRR37* mRNA levels were quantified in

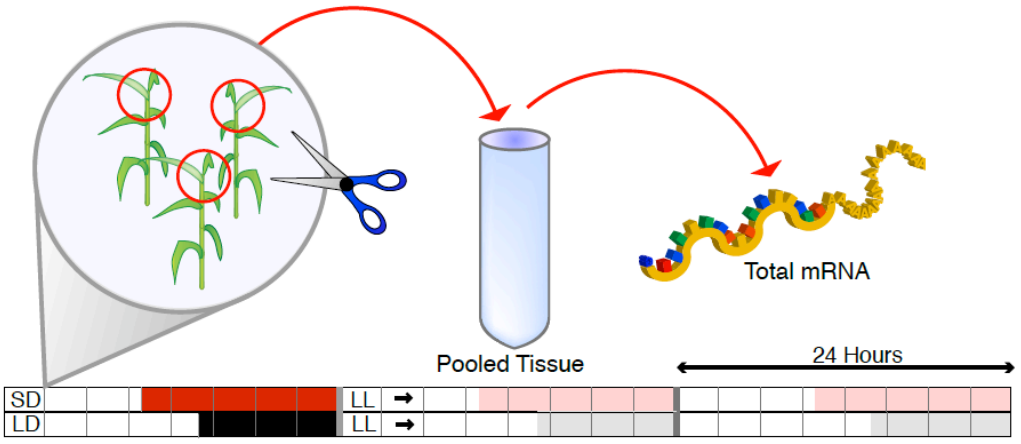

Fig. 18. Gene expression analysis of *PRR37* and flowering genes. Plants were grown under LD for 32 days, and subsequently treated with LD (14h, black shading) or SD (10h, red shading). Leaf tissue was collected and pooled from three plants in either group at three hour intervals (vertical gray lines) for one 24-hour day/night cycle, followed by 42 hours of constant light (Light red and gray shading). Total mRNA was extracted and used in subsequent qRT-PCR analysis.

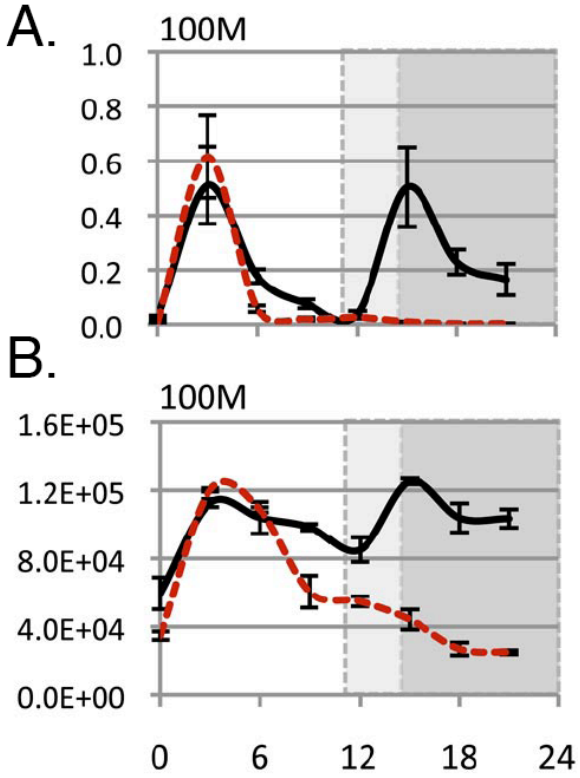

Fig. 19. Expression analysis of the alternatively spliced cDNA. Intron 8 (*A*) and intron 7 (*B*) Fig. 16; Table 2). The expression patterns of these cDNA variants parallel that seen in overall transcript abundance in long and short days. Additionally, these splice sites are found in equal proportion in *PRR37* and *prr37* genotypes, suggesting that these differences do not necessarily contribute to overall phenotypic differences. Expression of the variant observed in intron 7 was extrapolated from the area under curve as obtained from the ABI 3130xl instrument and Peak Scanner software (188). The ordinate represents the average area under the curve ± SEM and is based on three biological replicates. For the variant observed in intron 8, expression was detected using qRT-PCR as described in the Materials and Methods Section. The ordinate represents normalized expression relative to a calibrator sample (189) and is based on three biological replicates ± SEM.

photoperiod sensitive 100M and F_1 plants under LD, SD and circadian cycling conditions (Fig. 18). Analysis of cDNA revealed several PRR37 splice variants (Table 2; Fig. 19*A* and *B*); however, the abundance of splice variants and full-length transcripts was regulated in a similar manner, therefore overall *PRR37* transcript abundance was quantified by qRT-PCR.

In LD, 100M and F_1 plants show peaks of *SbPRR37* mRNA abundance in the morning and in the evening approximately 3 hrs and 15 hours after lights were turned on, respectively (Fig. 20*A* and *B*). The daily bimodal cycling pattern of *SbPRR37* mRNA abundance persisted in continuous light and temperature (LL) (Fig. 20*A* and *B*, days 2/3, star), indicating that the circadian clock modulates *SbPRR37* expression under these free-running conditions. By contrast, in SDs, 100M and F_1 plants showed only the morning-phase peak of *SbPRR37* mRNA abundance (Fig. 20*A* and *B*, day 1). However, when plants grown in SD were transferred to LL, both the morning and evening-phase peaks of *SbPRR37* mRNA abundance were observed (Fig. 20*A* and *B*, days 2/3, star). This suggests that *SbPRR37* expression is light-dependent and that the disappearance of the evening peak of *SbPRR37* expression in SD is caused by the lack of light during the evening. The light dependence of *SbPRR37* expression was further analyzed by transferring F_1 plants grown in LD to continuous dark (DD) (Fig. 20*C*). In DD, neither peak was observed, consistent with a requirement for light for *SbPRR37* expression. The results described above indicate that *SbPRR37* expression is dependent on illumination

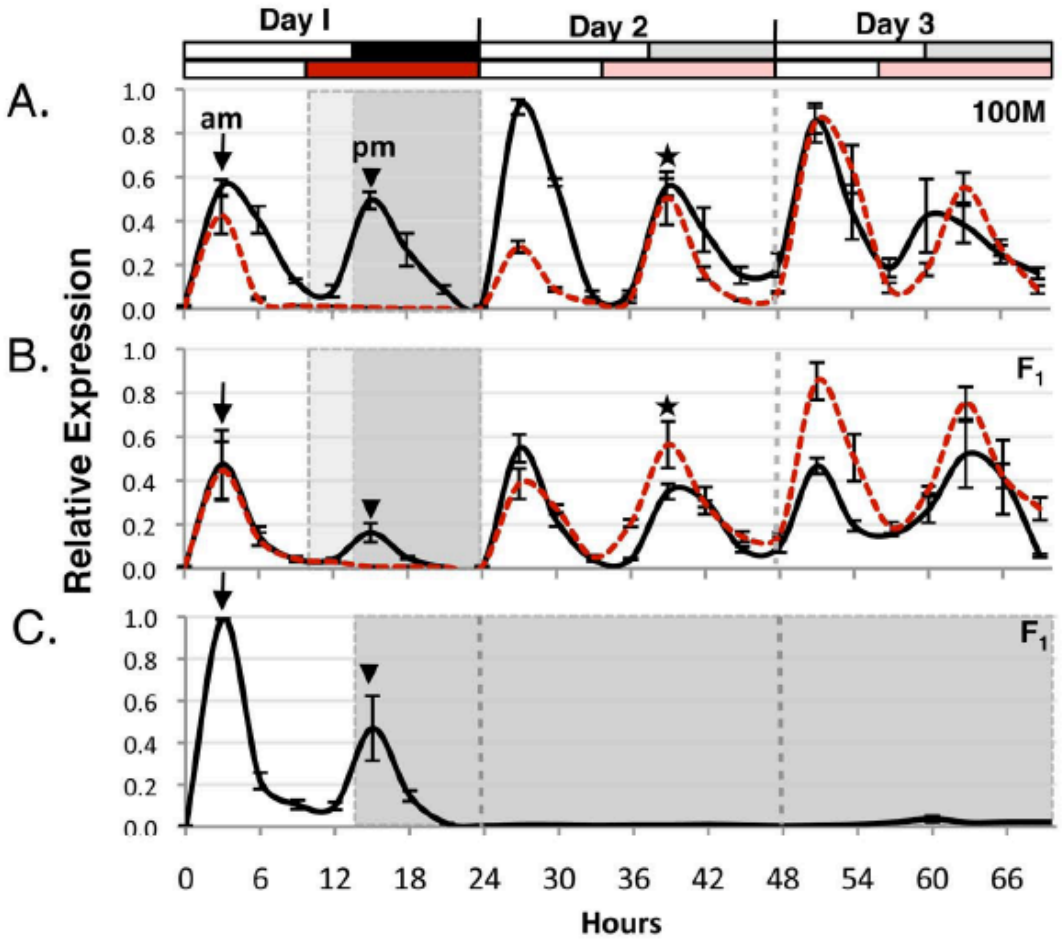

Fig. 20. Clock-regulated SbPRR37 expression is light dependent. Plants were grown in 14-h light:10-h dark LD (solid line) or 10-h light:14-h dark SD (red dashed line) and then released into LL at time 24 h. Relative expression of *SbPRR37* was analyzed at 3-h intervals by quantitative RT-PCR. In 100M (*A*) and ATx623 x R.07007 F_1 plants (*B*), *SbPRR37* expression increased in the morning (arrow) and evening (arrowhead) of long days. (*C*) ATx623 x R.07007 F_1 plants grown in 14-h:10-h LD and then released into DD at time 24 h. The ordinate represents normalized expression relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar at the top of the figure indicates the dark period for LD-treated plants, and the gray bars indicate subjective dark during LL conditions. The red bar indicates darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. The light gray shading within the plot area indicates darkness for SD-treated plants only, and the dark gray shading indicates darkness for both LD- and SD-treated plants.

of plants during times of the day when output from the circadian clock has the potential to activate it via circadian gating. This mode of regulation is consistent with the external coincidence model of flowering time regulation (121). In LL or LD, output from the circadian clock activates *SbPRR37* transcription in the morning and evening and the continuous production of PRR37 in LD is proposed to repress flowering. In SD, output from the clock increases *SbPRR37* expression during the morning but not in the evening because the evening phase of potential clock activation of *SbPRR37* expression occurs in darkness. In SD, lack of increased *SbPRR37* expression during the evening phase is proposed to reduce the level of the repressor PRR37, allowing floral initiation.

The important contribution of the evening peak of *SbPRR37* expression to floral repression in LD was supported by analysis of the genotype R.07007 (Fig. 21*A*). This genotype is photoperiod insensitive and flowers early in LD due to recessive ma_5 and ma_2 (22), despite possessing a functional *SbPRR37* allele. In LD, the morning phase increase in SbPRR37 expression was observed in R.07007 and 100M (Fig. 20*A*; Fig. 21*A*, arrow). However, the increase of SbPRR37 expression in the evening that occurs in 100M was not observed in R.07007 (Fig. 20*A*; Fig. 21*A*, arrow, arrowhead). The evening peak is restored under LL conditions, although shifted three hours later than peaks of *SbPRR37* mRNA abundance observed in 100M or the F_1 (Fig. 20; Fig. 21*A*, star). The molecular basis for altered *SbPRR37* expression during the evening and under LL conditions in R.07007 is not known, but may be associated with recessive ma_5 or ma_2 alleles in the R.07007 background, which have been shown to be genetically

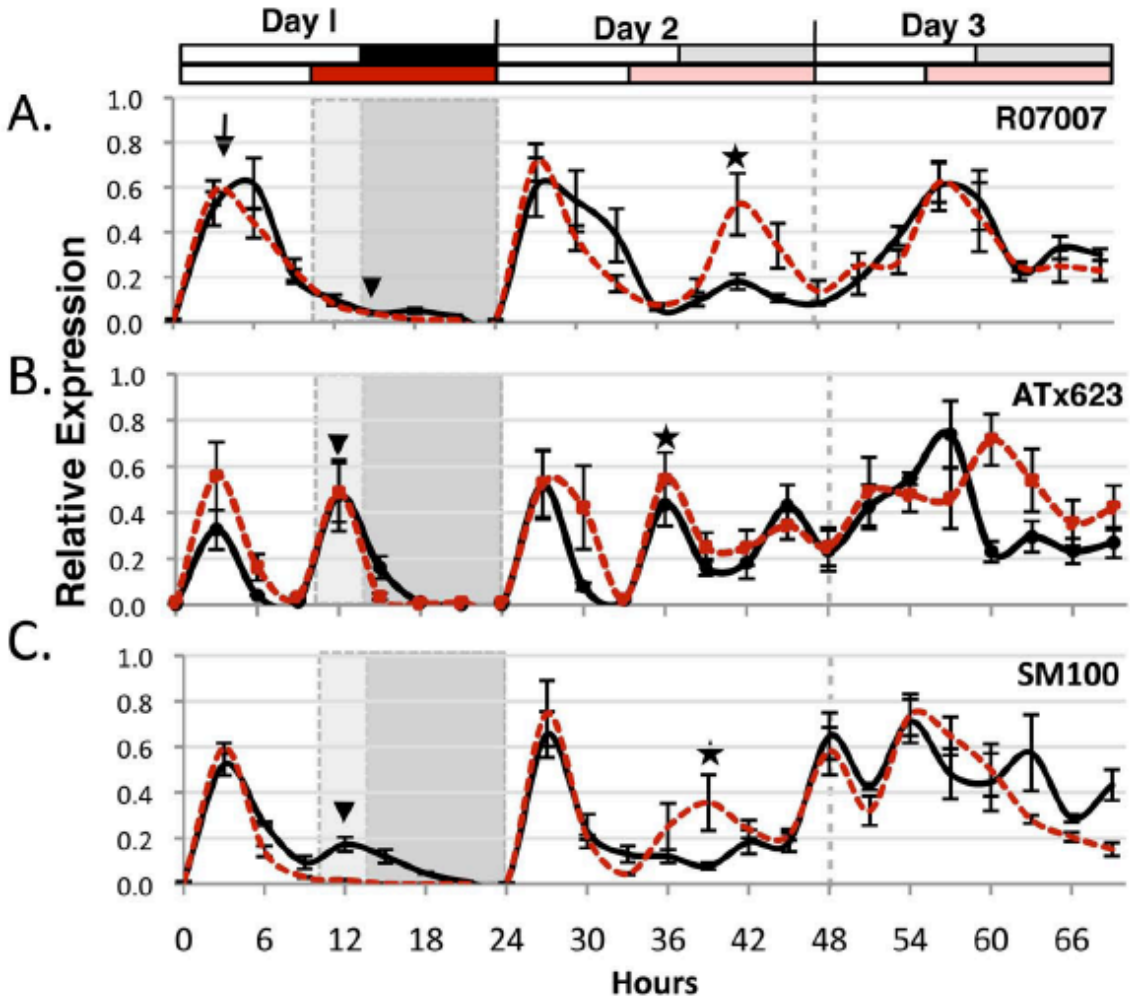

Fig. 21. *SbPRR37* expression in PI genotypes. (*A*) *SbPRR37* expression in R.07007 plants grown in LD and then transferred to LL at time 24 h. (*B*) The *Sbprr37-3* mutation in ATx623 results in a nonfunctional protein. The expression patterns are similar to 100M and the F_1, with the exception of a 3-hour early shift. (*C*) The *Sbprr37-1* mutation in SM100 also results in a nonfunctional protein. The expression patterns are similar to ATx623, also showing an early shift. The ordinate represents normalized expression relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar at the top of the figure indicates the dark period for LD-treated plants, and the gray bars indicate subjective dark during LL conditions. The red bar indicates darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. The light gray shading within the plot area indicates darkness for SD-treated plants only, and the dark gray shading indicates darkness for both LD- and SD-treated plants.

interdependent with Ma_1 (22). Regardless, these results show that evening-phased expression of SbPRR37 is correlated with repression of flowering in LD-photoperiods in sorghum. By contrast, expression analysis of the recessive *Sbprr37-3* and *Sbprr37-1* alleles in the early flowering parents ATx623 and SM100, respectively (Fig. 21, *B* and *C*), revealed that the dual peaks observed in LD in 100M and the F_1 are present, though shifted three hours early relative to those genotypes. As with the altered expression patterns observed for R.07007, the basis for this phenomenon is not known. However, it has been proposed that *PRR7* expression may be self-regulating through a negative feedback loop (190), and the lack of functional protein to participate in self-repression activity in these sorghum genotypes may result in an expression cycle that is faster than that observed for their dominant *SbPRR37*-possesing counterparts. Additionally, though *PRR7* expression in *Arabidopsis* is connected to other PRRs in the circadian clock, obvious expression differences as a result of the *PRR37* lesion were not detected in sorghum (Fig. 22*A-I*).

Regulation of Clock and Flowering Time Genes by *SbPRR37*

We next investigated the mechanism by which PRR37 represses flowering under LD conditions in sorghum. Genes in the canonical *Arabidopsis* flowering pathway also contribute to the control of flowering time in rice and other grasses, and the most noted of these, CO, is a repressor of flowering in rice in LDs (156). Therefore, in order to gain further understanding of how PRR37 modulates the floral induction pathway, we characterized the expression of the sorghum ortholog of *CO* over a 40-hour time course

in 100M and SM100 in LD and SD to determine if this gene was regulated by PRR37 as previously shown in barley (182). 100M grown in LD (solid line) showed two increases of *CO* mRNA abundance in leaves each day, similar to the daily bimodal *CO* expression pattern observed in maize in LD (Fig. 23*A*) (191). The first peak of *CO* mRNA occurred in the evening approximately 15 hours after lights on, and the second increase occurred during the last several hours of the night, peaking at dawn (24 hours) (Fig. 23*A*, upper, arrowhead). By contrast, in SD (red dashed line), the peak of *CO* expression at dawn was greatly reduced. In addition, the second peak of *CO* expression observed at dawn in 100M grown in LD is absent in SM100 (*prr37-1*) (Fig. 23*A*, lower, arrowhead). These results indicate that the reduction of *CO* mRNA abundance at dawn in SM100 plants grown in LD is due to *prr37-1*, and that PRR37 is required for differential expression of CO in response to photoperiod in sorghum. *CO* expression is regulated by the circadian clock through the action of GI in *Arabidopsis* and rice (75-77). Therefore, it is possible that PRR37 alters *CO* expression through an indirect effect on clock gene expression. Small differences in the patterns of *TOC1* (Fig. 24*A-E*), *LHY* (Fig. 24*F-J*), and *GI* (Fig. 25*A-E*) expression were noted but could not be directly connected to the altered expression of *CO* in SM100 compared to 100M. We interpret these results to indicate that, although PRR37 may have an effect on clock gene expression, this protein also directly regulates *CO*. Regulation of *CO* expression by PRR7, an ortholog of SbPRR37, independent of the clock-*GI* pathway was also proposed in Arabidopsis (192).

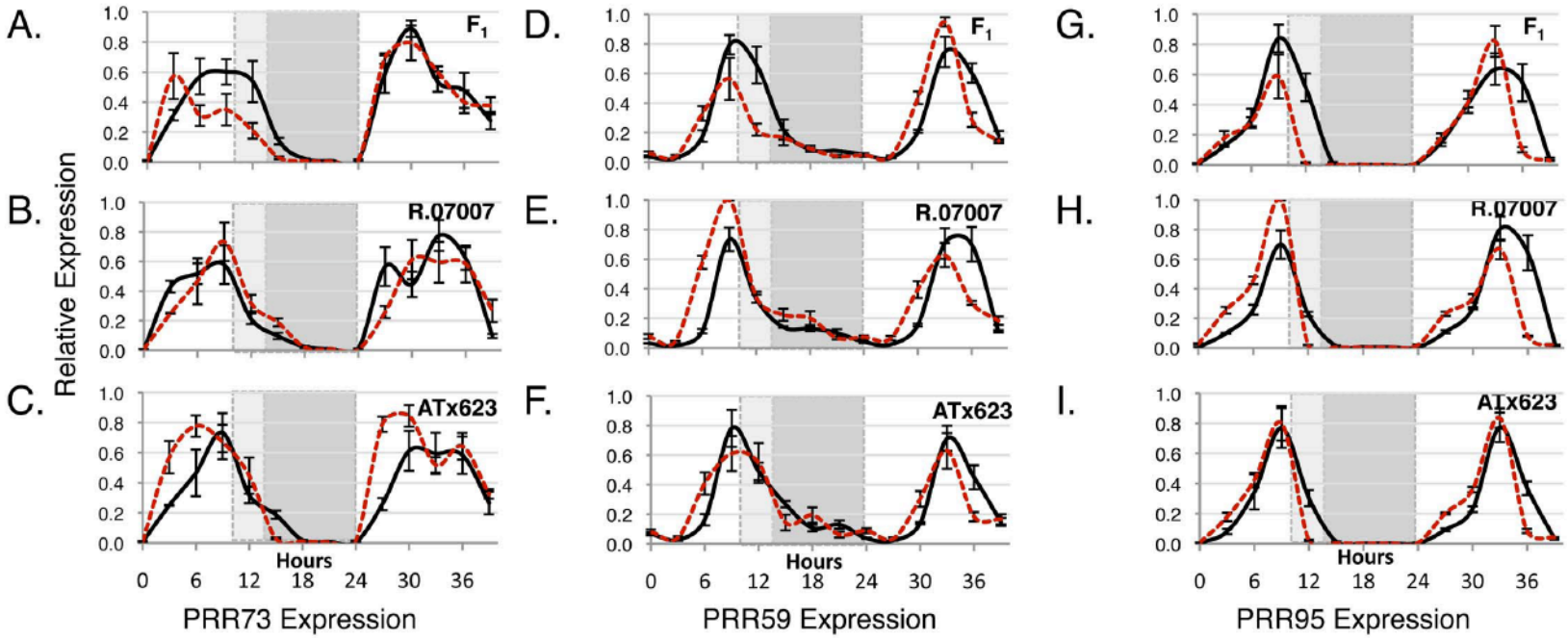

Fig. 22. General *SbPRR* expression is not affected by *prr37*. Plants were grown in LD and then transferred to LL at time 24 h. *SbPRR73* (*AtPRR3*) expression in (*A*) ATx623 x R.07007 F_1 (*B*) R.07007 (*C*) ATx623. *PRR59* (like *AtPRR5* and *9*) expression in (*D*) ATx623 x R.07007 F_1 (*E*) R.07007 (*F*) ATx623. *PRR95* (like *AtPRR9* and *5*) expression in (*G*) ATx623 x R.07007 F_1 (*H*) R.07007 (*I*) ATx623. The ordinate represents normalized expression relative to a calibrator sample and is based on three biological replicates ± SEM (189).

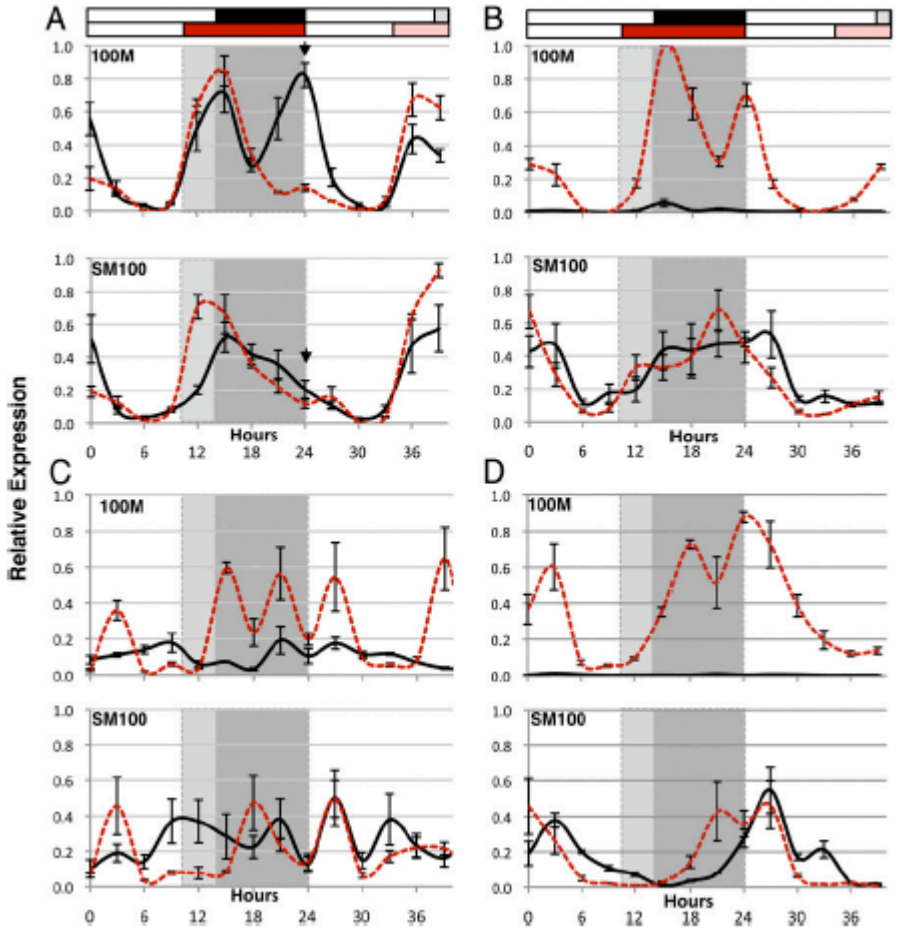

Fig. 23. *SbPRR37* modulates expression of downstream flowering genes. Plants were treated under14-h light:10-h dark (LD, solid line) or 10-hlight:14-h dark (SD, red dashed line) conditions. (*A*) Relative *CO* expression in 100M peaks at dawn (arrowhead) in plants treated in LD, but not in SD. This peak is absent in SM100 under either condition. (*B*) Relative *Ehd1* expression is repressed under LD in 100M, but is activated under both LD and SD in SM100. (*C*) Expression of *FT* is repressed in LD in 100M, but SM100 expression levels are equivalent in LD and SD. (D) Expression of *ZCN8* is elevated in SD-treated 100M plants but is repressed to near undetectable levels in LD. In SM100, expression is de-repressed in LD. The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar above the plot indicates the dark period for LD-treated plants; gray bars indicate subjective dark during LL conditions. Red bars indicate darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. Light-gray shading within the plot area indicates darkness for SD treated plants only; dark-gray shading indicates darkness for both LD- and SD-treated plants.

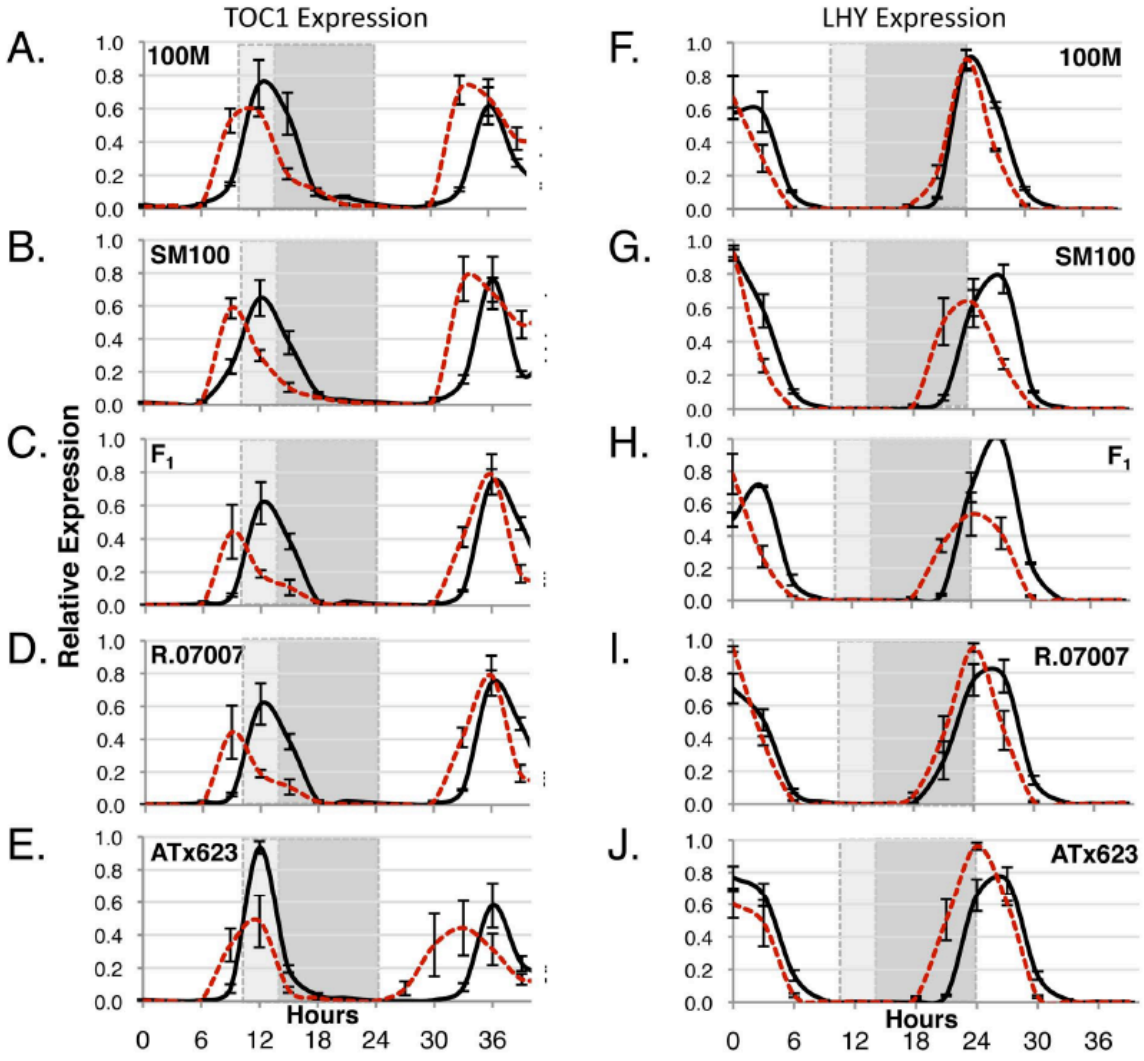

Fig. 24. Expression of core clock genes in *SbPRR37*, *Sbprr37-1*, and *Sbprr37-3* genotypes. (*A*-*E*) No expression differences were observed in the core clock gene *TOC1* between *PRR37* genotypes (100M (*A*) and ATx623 x 007 F_1 (*C*)), the *Sbprr37-1* mutant (SM100), or in *Sbprr37-3* types (ATx623) in either LD (solid black line) or SD (red dashed line). (*F*-*J*) *LHY* also is expressed in a similar manner in all genotypes. The ordinate represents normalized expression relative to a calibrator sample (189) and is based on three biological replicates ± SEM. The light-gray shading within the plot area indicates darkness for SD-treated plants only, and dark-gray shading indicates darkness for both LD- and SD-treated plants.

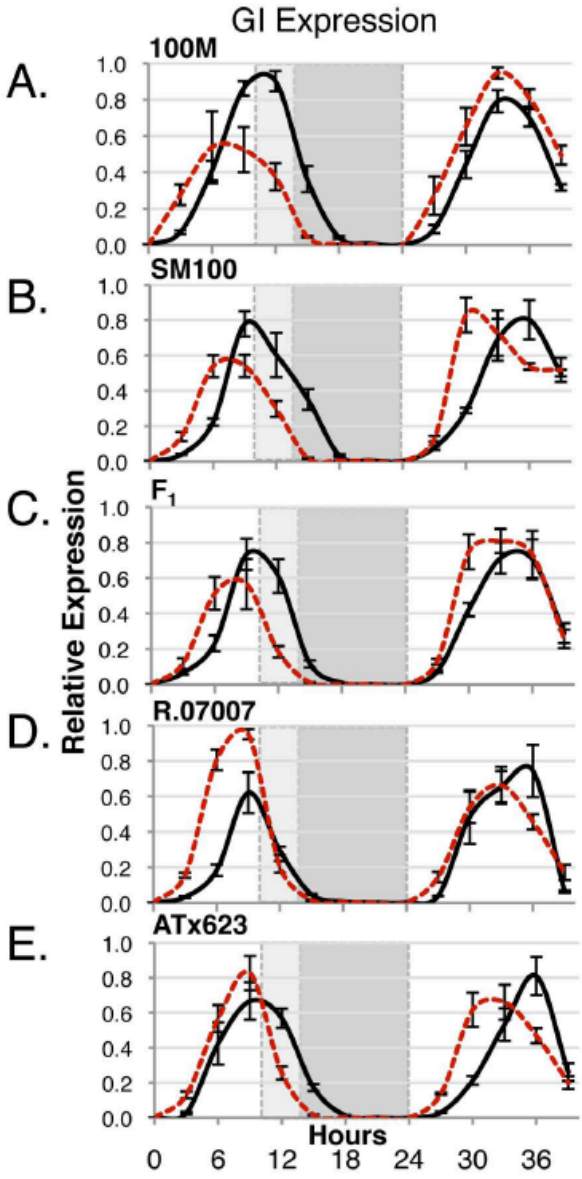

Fig. 25. Flowering gene GI follows the expression pattern of LHY in *PRR37* genotypes. (*A*, *C* and *D*) 100M, the ATx623 x 007 F_1, and R.07007 (*B*) the *Sbprr37-1* mutant (SM100), or in (*E*) *Sbprr37-3* types (ATx623) in either LD (solid black line) or SD (red dashed line). The ordinate represents normalized expression relative to a calibrator sample (189) and is based on three biological replicates ± SEM. The light-gray shading within the plot area indicates darkness for SD-treated plants only, and dark-gray shading indicates darkness for both LD- and SD-treated plants.

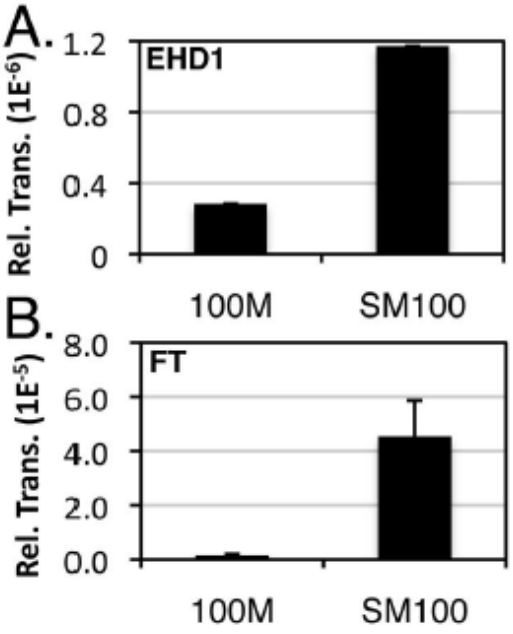

Fig. 26. Direct comparison of *Ehd1* and *FT* expression. (*A*) *Ehd1* (*B*) *FT* (*Hd3a*) transcript levels in 100M and SM100. Relative transcript abundance was calculated at 15 h into the LD, 1 h after the beginning of the dark period. This time point was selected because it corresponds to the evening peak of *SbPRR37* expression in LD (A and B, arrowhead). The analysis showed that *Ehd1* (*A*) mRNA levels were significantly lower in the leaves of LD-treated 100M compared with SM100 (P value < 0.001). *FT* (*B*) expression levels were similarly decreased in LD-treated 100M plants compared with SM100 (P value < 0.01). When 100M plants were transferred to SDs, expression of *Ehd1* and *FT* increased by 17- and 7.06-fold, respectively. By contrast, upon SD transfer, the increases in *Ehd1* and *FT* levels in SM100 were only 2.35 and 0.34, respectively, consistent with the de-repression of these floral activators in the *prr37-1* background. The ordinate represents the absolute transcript abundance normalized relative to 18S ± SD (189). Statistical significance was calculated using a two-tailed Student's t test; actual P values were 0.005 and 0.0002 for *Ehd1* and *FT*, respectively.

In rice, *Ehd1* encodes a BRR-type transcription factor unique to grasses that has been shown to promote flowering (163). Because *Ehd1* has a role in floral activation via a pathway separate from *CO*, we identified the sorghum ortholog of *Ehd1* and found that

in 100M expression of this gene was strongly repressed in LD (Fig. 23*B*, upper), while in SM100, LD and SD levels were similar (Fig. 23*B*, lower). Moreover, when 100M plants grown in LD were transferred to SD, expression of *Ehd1* at 15 hours after lights on increased ~17-fold (Fig A). In contrast, transfer of SM100 plants to SD increased *Ehd1* mRNA levels only 2.4-fold (Fig. 26*A*). These results are consistent with the de-repression of *Ehd1* in the *prr37-1* background (SM100). *FT* is part of a 6-member PEBP-domain gene family in *Arabidopsis* and >20-member gene family in maize (173). Several members of the PEBP-domain gene family encode florigens that modulate flowering in rice and maize (162, 173). In rice, Hd3a and RFT1 act synergistically to promote the transition from vegetative to reproductive growth (162). No ortholog of *RFT1* was found in sorghum (162) (Phytozome v5.0). However, a collinear sorghum ortholog of rice *FT* (*OsHd3a*) was present in sorghum and this gene was regulated by photoperiod and PRR37 (Fig. 23*C*). The sorghum ortholog of *OsHd3a* (*SbFT*) was expressed in 100M leaves at lower levels in LD compared to SD (Fig. 23*C*, upper). In contrast, *SbFT* showed elevated expression in SM100 plants in LD and SD (Fig. 23*C*, lower). When 100M plants were transferred from LD to SD for one week, SbFT mRNA levels increased 7.1-fold during the evening phase (15 hours after lights on). By contrast, transfer of SM100 plants from LD to SD for one week increased *SbFT* levels in SM100 only 0.3-fold indicating the absence of repression of *SbFT* expression in LD in the *prr37-1* background (Fig. 26*B*).

Sorghum *ZCN8*, the collinear ortholog of the maize florigen *Zea mays CENTRORADIALIS 8* (173), was expressed at low levels in LD in 100M and at elevated

levels in SD (Fig. 23*D*, upper). Similar to *Ehd1* and *SbFT*, *SbZCN8* expression in SM100 plants was de-repressed regardless of photoperiod (Fig. 23*D*, lower). Additionally, expression analysis of *ZCN12*, a second florigen candidate gene in maize that responds to photoperiod (173), showed a pattern of expression in 100M and SM100 similar to SbZCN8 (Fig. 27, *A* and *B*). In summary, expression of sorghum orthologs of

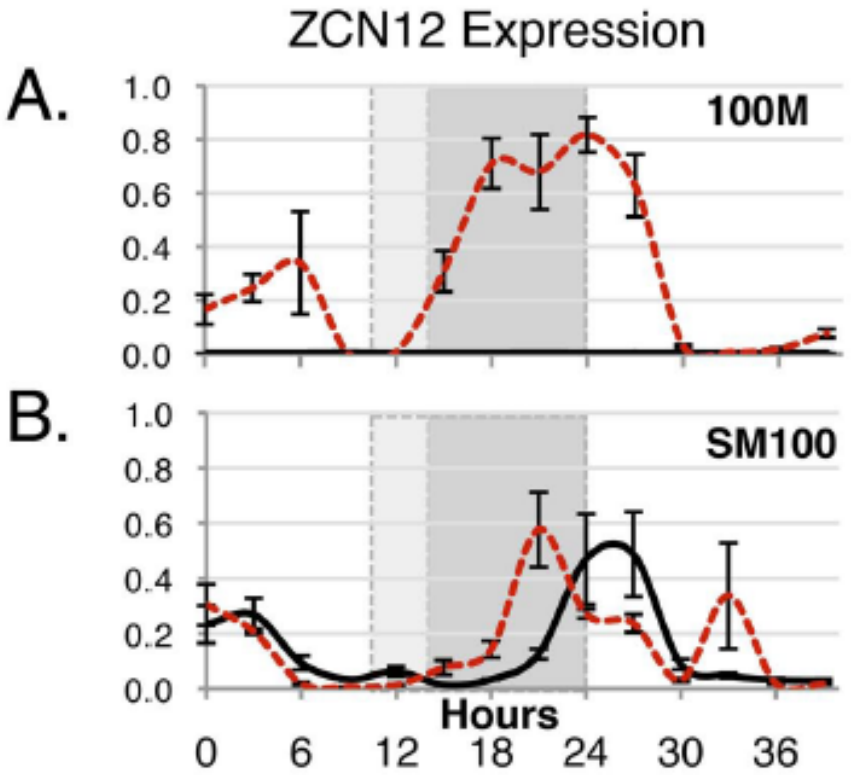

Fig. 27. *ZCN12* is regulated in response to photoperiod. *ZCN12*, a second candidate for florigen in maize, is activated in response to the SD photoperiod while remaining repressed to near-undetectable levels in LD-treated 100M plants. SD- and LD-treated SM100 shows no real differences in expression levels. The ordinate represents normalized expression relative to a calibrator sample (189) and is based on three biological replicates ± SEM. The light-gray shading within the plot area indicates darkness for SD-treated plants only, and dark-gray shading indicates darkness for both LD- and SD-treated plants

genes that are involved in floral induction in other grasses including *Ehd1*, *FT*, *ZCN8* and *ZCN12* is regulated by photoperiod in 100M (*PRR37*), but not in SM100 (*prr37-1*), a genotype that lacks a functional PRR37.

Discussion

This study demonstrates that SbPRR37 is a central repressor in a regulatory pathway that controls sorghum flowering in response to photoperiod. A working model for this regulatory network is shown in Fig. 28. In LD, light-dependent circadian-regulated increases in *SbPRR37* expression in the morning and evening are proposed to result in a sufficient level of PRR37 throughout the day to repress *FT*, other genes encoding florigens, and floral initiation. In SD, the evening peak of *SbPRR37* expression is reduced or eliminated leading to floral induction consistent with the external coincidence model of flowering time regulation (121, 176). By contrast, *Arabidopsis PRR7*, the ortholog of *SbPRR37*, shows only a single morning-phase peak of expression (193) indicating that evening phase expression of this gene may be a special feature of grass species. The light-dependent induction of *SbPRR37* expression and the clock-mediated evening phase peak of expression enable *SbPRR37* to regulate flowering time in response to photoperiod; PRR37 mRNA levels in the evening phase decrease as day length is reduced. The photoreceptor(s) that mediate light-induced *SbPRR37* expression are currently under investigation, however, phytochrome B is likely involved because

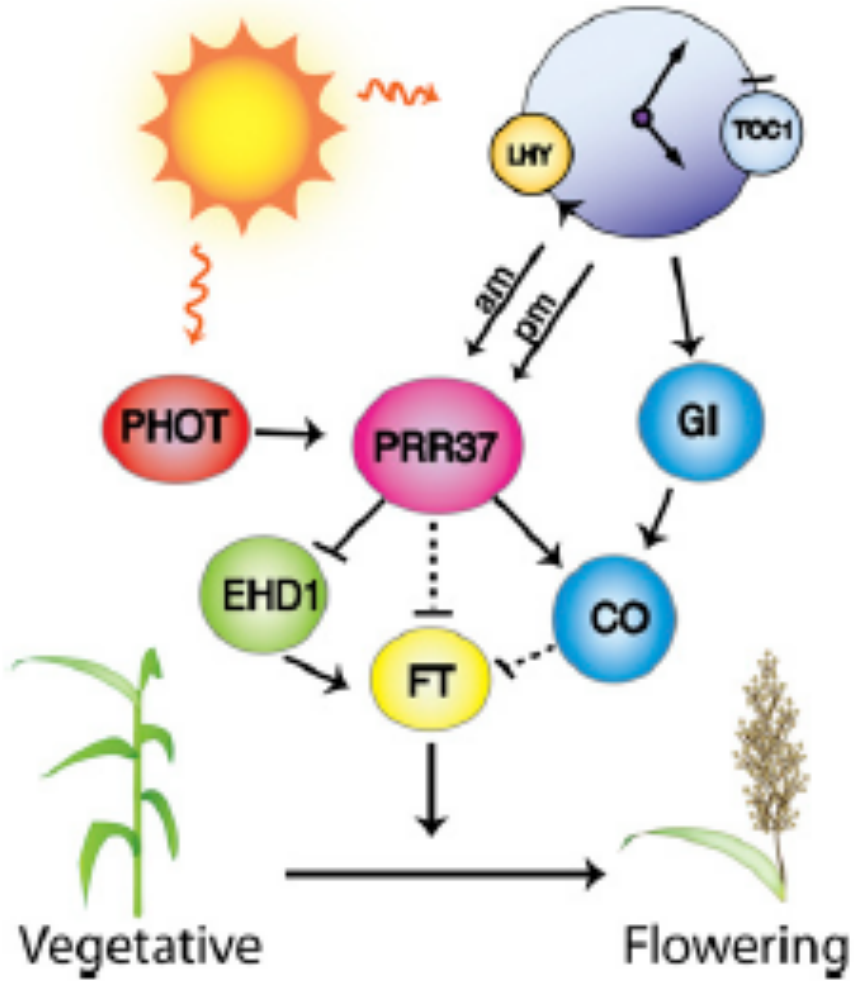

Fig. 28. Model of photoperiodic flowering time regulation in sorghum. PRR37 is a central floral repressor that blocks transition from the vegetative phase to flowering in LD. PRR37 represses *FT*, *ZCN8*, and flowering by activating expression of *CO*, a repressor of *FT* in rice, and by inhibiting *Ehd1*, a grass-specific inducer of *FT*. *SbPRR37* expression is regulated by the circadian clock and light in a manner consistent with the external coincidence model. It is proposed that photoreceptors (PHOT) such as phytochromes mediate light activation of *SbPRR37* expression coincident with output from the circadian clock, resulting in increased *SbPRR37* expression in the morning and evening in LD. In SD, *SbPRR37* expression is not activated in the evening, leading to floral induction.

recessive alleles of this gene cause early flowering in LD in sorghum (*ma*$_3$, *ma*$_3^R$) (24), barley (194) and rice (195). SbPRR37 is proposed to repress *FT*, *SbZCN8* and *SbZCN12* and flowering in LD in part by inhibiting expression of *Ehd1*, an activator of *FT* and flowering in rice (163), by increasing expression of *CO*, a repressor of flowering in rice in LD (196), and possibly by other mechanisms that modulate *SbFT* and *SbZCN8* expression.

This study provides insight into the mechanism of photoperiodic regulation of flowering time in the SD-grass sorghum. The importance of PRR37 in photoperiod regulation was first documented in the LD-grasses barley (182) and wheat (183). In these grasses, PRR37 activates *FT* and flowering in LD whereas in sorghum PRR37 represses *FT*, *ZCN8* and flowering in LD. The molecular basis of this difference in PRR37 activity in sorghum, a short day plant, and the long day grasses barley/wheat may relate to differences in *CO* activity on the formation of CCAAT-box binding complexes involved in floral gene expression (132, 196). In addition to documenting how sorghum regulates flowering time in response to photoperiod, this study identified important alleles of *SbPRR37* that were critical for the domestication and utilization of this tropical grass for grain production in temperate regions worldwide (1, 179). This information will allow plant breeders to more precisely control flowering time in grass species thus increasing yield and sustainable production.

MATERIALS AND METHODS

Genotyping and Phenotyping Mapping Populations

All sorghum accessions used in this study, as well as relevant descriptors, are listed in Table 3. For flowering date determinations, plants were grown in different LD environments in the greenhouse or in the field in the summer seasons of 2003-2010. Days to mid-anthesis were evaluated in the greenhouse under 14-h light/10 h dark photoperiods (30-34/20-25°C), and in field locations in College Station, Vega, and Plainview, TX (USA). For genotyping, plant DNA was extracted with either the FastDNA Spin Kit (MP Biomedicals, Solon, OH, USA) or by disruption of a leaf punch using a GenoGrinder (BT&C/OPS Diagnostics, Bridgewater, NJ). Plants from all three mapping populations were subjected to marker analysis as previously described (197).

To identify polymorphisms between parental genotypes, SSR markers were developed utilizing SSRIT (198). *De novo* sequencing and analysis of promoter and intron regions flanking predicted open reading frames identified additional PCR-based SNP and INDEL markers used for high-resolution linkage analysis. PCR primers were designed with Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) or PrimerQuestSM (Integrated DNA Technologies, Inc) to amplify selected regions of the sorghum genome (http://www.phytozome.net/sorghum).

Table 3. Pedigree and *Maturity Locus 1* classification for sorghum genotypes utilized in linkage analyses and gene expression studies of Ma_1.

Sorghum genotype	Classification at *Maturity Locus 1*	Pedigree, year/decade of germplasm development
100M Milo	Ma_1	Derivative of Early White Milo-Double (circa 1911) X Dwarf Yellow Milo (circa 1936). Developed by Quinby as maturity loci genetic stock for flowering time studies.
SM100 Milo	ma_1, *Sbprr37-1*	Derivative of Early White Milo-Double (circa 1911) X Dwarf Yellow Milo (circa 1936). Developed by Quinby as maturity loci genetic stock for flowering time studies
Blackhull Kafir	ma_1, *Sbprr37-2*	Pure line selection from Standard Kafir, circa 1919
ATx623	ma_1, *Sbprr37-3*	Selected from the cross of elite lines BTx3197 X SC170-6-4 (Kafir x Zera-zera), circa 1964
R07007	Ma_1	Derivative of EBA-3 (Dual purpose sorghum from Argentina), Texas AgriLife release 2007.
BTx406	ma_1, *Sbprr37-1*	Martin derivative. Martin originated from undetermined outcross of Wheatland (Kafir by Milo)

Pedigree and classification at *Maturity Locus 1* are based on germplasm release notices on historical records maintained at Texas A&M University. Information from Dr. Robert Klein (184).

Purified PCR products were used in sequencing reactions with Big Dye Terminator v3.1 (Applied Biosystems) and run on an ABI 3130xl Genetic Analyzer. Sequence assembly and analysis was carried out using Sequencher® v4.8 (Gene Codes Corporation). Physical locations of genetic markers and associated high-resolution map units were based on the whole genome sequence of Sorghum bicolor (http://www.phytozome.net/sorghum; v5.0). All markers used in this study are listed with their physical coordinates in Table 4.

Table 4. Markers used for fine mapping the *Ma_1* locus.

Marker	Forward	Reverse primer	Marker type	SBI-06 marker location (b)
Xtxp694	GGGCCCTGTTACATCCTTAAT	TGCCTGATCTTAGAGAAACACC	SSR	38,368,626
Xtxi48	TCAAGGCAAGATTGACGAAGCCAC	GGCTTGTAGCAGTAGCACTTGTGT	INDEL	39,098,709
Xtxsn1	ACCTTTCAGTGTGGTGCAAATGGG	TTGGCCCGATAGCAGTCCGATAAA	SNP	39,901,458
Xtxp696	TCGATCGATTCTCCTGCTTT	GTAGGTGCACCCAGTGCTTC	SSR	40,200,439
Xtxi61	GCCTTTGCAAGCAAAAATCT	TCTCGAGCCTAATCCCAATC	INDEL	40,278,618
Xtxi62	CGGTGCGTAGCAAATGTAAA	GGTCCAATGCAGAAGACGAT	INDEL	40,278,905
Xtxi63	CTCCTTTTGCTCCACGTCAT	GCATGCAGATGGCTGAGTTA	INDEL	40,279,273
Xtxi49	CGAGCCAATTTCACCTCCTA	GCCAATGCATGTTTCATAGC	INDEL	40,279,611
Xtxi55	TTTATGCCCGGTGTGTCTG	CATCACTGCACATGAACCAC	INDEL	40,286,452
Xtxi73	AATTTTCTATGCAATTAAGAAGAG	GCCTTAAGAGCCGGGAATA	INDEL	40,313,421
Xtxi58	GGAGCTATTGCTATGCTGCTT	CTCAGAGTACAGCAGCTCCAAC	INDEL	40,364,576
Xtxi68	TAGAGCCTTTGTGCAGCATTC	CAACCAATTGCCCTTGTTTAC	INDEL	40,408,262
Xtxp711	CACCTAGCAGAGGGCAAGAG	CACACTCATTGCTTGCCTGT	SSR	40,440,214
Xtxi66	GGGAGCGTTGAAACTTGATG	GCAAGCACAGACGAACTCAC	INDEL	40,491,401
Xtxi20	GCCTCCACTTGCGAATGAT	ATACATAACTTGTGGGTCGAAAG	INDEL	40,595,306
Xtxp599	TGAAAACGAACCAAACACACTC	TTTAAATACTTCCTCCATTCCAAA	SSR	40,801,774
Xtxp598	GTGGCGCACAGCTAAAAGT	TTTGGTCCGATCTTTTGGAG	SSR	40,828,046
Xtxp434	CGAGGTCCAGGAGTACACG	CGGCCTCCAGGAGGAGTAAT	SSR	42,610,344

Sequences of forward and reverse primers are given, as well as polymorphism classification. Physical coordinates are given as obtained from Phytozome v7.0.

Sequencing of *SbPRR37* Alleles

To examine the *SbPRR37* gene for functional mutations that contribute to the temperate-zone adaptation of sorghum, either the full 10kb *SbPRR37* genomic region or expressed cDNA was sequenced from six genotypes including historically prominent cultivars. PCR-amplified products from genomic DNA (Phusion® High-Fidelity DNA Polymerase, New England BioLabs, Inc) were isolated using the QIAquick PCR Purification and Gel Extraction Kits (QIAGEN). Sequencing of the purified PCR products was carried out in a reaction using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Capillary sequencing was performed on the Applied Biosystems 3130xl Genetic Analyzer. The results were assembled and analyzed using Sequencher® v4.8 (Gene Codes Corporation). For cDNA sequence determination of *SbPRR37* alleles from genotypes in Table 3, RNA was extracted from leaf tissue of 2-week old plants using the miRNeasy Mini Kit (QIAGEN). After extraction, RNA concentration was determined spectophotometrically, and RNA integrity was visually assessed by denaturing agarose gel electrophoresis. Ten µg of total RNA was DNase-treated with Turbo DNA-Free Kit (Applied Biosystems/Ambion, Austin, TX, USA), and this RNA (1µg) was used for first-strand cDNA synthesis (SuperScript™ III First-Strand Synthesis System; Invitrogen, Carlsbad, Ca, USA) primed with either Oligo(dT)20 or random hexamers. After inactivation of reverse transcriptase, the cDNA was diluted to a final concentration of 5 ng/µl with dH2O. Based on the full-length sequence of SbPRR37 ESTs from 100M, SM100, and Blackhull Kafir, PCR primers (5'-

TCCTCCAGAAGAAGACCATCA-3'; and 5'-CAACAGCACCTTTCGACAAA-3') were designed adjacent to the transcription start and stop sites of SbPRR37, and cDNA from the remaining sorghum genotypes list in Table 3 were PCR-amplified from first-strand cDNA template. PCR products were cloned into the pCRII-TOPO™ vector and One Shot TOPO10 Competent™ cells (Invitrogen).

Forty-eight independent clones were selected from each genotype for sequence determination of the SbPRR37 allele. Phred and Phrap were utilized for sequence base calling and sequence assembly, respectively, and Consed Graphical Tool for sequence finishing (http://www.phrap.org/phredphrapconsed.html). Intron-exon borders were annotated by aligning the genomic and cDNA sequences.

SbPRR37 alleles were analyzed for conserved protein domains using NCBI BLAST against the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). The entire coding sequence of different *SbPRR37* alleles was compared to the GenBank nr database using BLASTP to identify homologues (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Nucleotide and protein sequence alignments were assessed with ClustalW2 (http://workbench.sdsc.edu/) and shading style in the background was applied with Boxshade 3.2.1 (http://www.ch.embnet.org/software/BOX_form.html).

Gene Expression Studies

Sorghum genotypes 100M, SM100, ATx623, R07007, and ATx623 x R07007 F_1 plants were grown in a greenhouse in Metro-Mix 200 (Sunshine MVP; Sun Gro

Horticulture, Canada CM, Ltd.) under long day conditions (14hr days) and were fertilized once after two weeks using Peters Professional Allrounder fertilizer (The Scotts Company LLC). After 32 days, the plants were transferred to a growth chamber for one week under either long (14hr days) or short (10hr days), at a light intensity of ~300µmol s-1 m-2 at ~50% humidity with 30° C day temperatures and 23° C night temperatures. At day 39, one week after SD or LD treatment, the three topmost leaves from three different plants (pooled) were harvested from each genotype every three hours for one 24 hour light-dark cycle and two additional 48-hour constant light (constant 30° C) or constant dark cycles (constant 23° C), as indicated.

Leaf tissue was homogenized in liquid nitrogen using a mortar and pestle, and total RNA was extracted from each sample using the TRI REAGENT® Protocol for samples high in polysaccharides (Molecular Research Center, Inc.). The resulting RNA was further purified using the RNeasy Mini Kit with on-column DNase digestion (QIAgen). RNA samples were quantified twice each using a NanoDrop 1000 Instrument (Thermo Fisher Scientific, Inc) and the average value for each sample was used. Five micrograms of each RNA sample was vacuum dried and resuspended in denaturing buffer and RNA integrity was visualized on a 1% MOPS gel using the Molecular Imager® Gel Doc™ XR running Quantity One®v4.6.8 software (Bio-Rad Laboratories, Inc.) (Fig. 29). First-strand cDNA synthesis was performed using the SuperScript™ III First-Strand Synthesis System (Invitrogen, Carlsbad, Ca, USA), primed with a 9:1 ratio of random hexamer/oligo dT mix using 4 µg of total RNA. The reactions were diluted

to 10ng/µl cDNA in 1x TE buffer for subsequent use in qRT-PCR expression analysis. This process was repeated twice more for a total of three biological replicates.

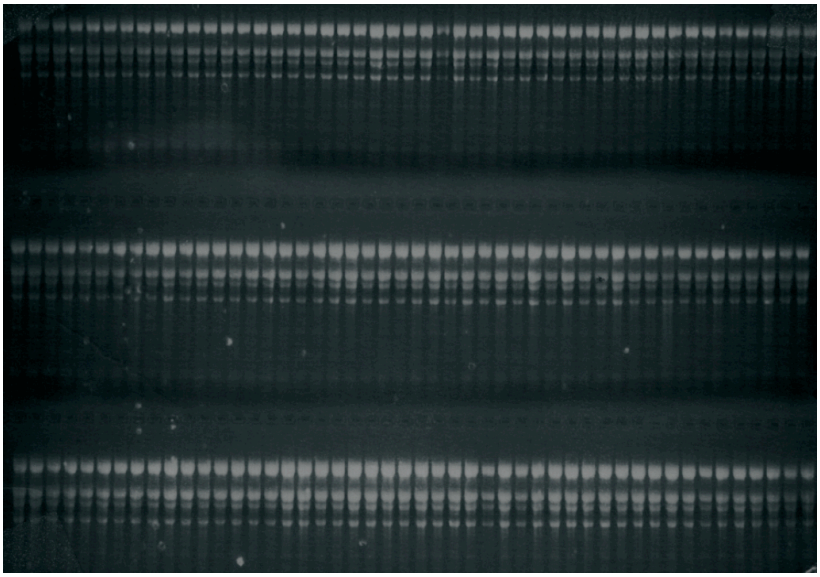

Fig. 29. Representative RNA gel. Samples (5µg) are from one biological replicate for ATx623, R.07007, and the ATx623 x R.07007 F_1 in both LD and SD.

Primers for qRT-PCR-based expression analysis of select clock and flowering time related genes were designed using PrimerQuestSM (Integrated DNA Technologies, Inc). Orthologs of rice and maize genes were identified based on protein and nucleotide sequence similarities, as well as position within collinear regions (Gramene Genome Browser release #33: http://www.gramene.org/). Sorghum gene and cDNA sequences were obtained through Phytozome v5.0, and primers were engineered to span an intron or 3'UTR sequence of each of the following genes: *PRR37*, *TOC1*, *LHY1*, *FT*, *GI*, *CO*, and *Ehd1*. To determine primer efficiencies serial dilutions were constructed from

purified PCR products amplified from cDNA using gene-specific primer sets over a range of 0.05 ng μl^{-1} to 5.0E^{-7} ng μl^{-1}, which were subsequently used in qRT-PCR reactions. The resulting Ct values were used to obtain a standard curve, by which the efficiencies for each primer pair for all genotypes were calculated (Applied Biosystems). Primers whose efficiencies were within 10% between genotypes were used for downstream analysis (Table 5). No-template control qRT-PCR reactions were also run using 18S ribosomal RNA (Applied Biosystems) for 10 ng µl-1 RNA from each sample to verify that there was no genomic DNA contamination. All of these and subsequent reactions were performed on the 7900HT Fast Real-Time PCR System running SDSv2.3 software. Gene-specific reactions were performed using Power SYBR® Green PCRMaster Mix (Applied Biosystems). All control 18S ribosomal RNA reactions were performed using the TaqMan® Universal PCR Master Mix Protocol with the rRNA Probe (VIC™ Probe), rRNA Forward Primer, and rRNA Reverse Primer (Applied Biosystems).

For the time course studies, raw Ct values were collected for each gene, and normalized to 18S ribosomal RNA to obtain ΔCt values. Relative expression was calculated using the Comparative Ct (ΔΔCt) method (189) with the most highly expressed sample used as the calibration sample. For genes in which LD and SD values were to be directly compared (FT, Ehd1, ZCN8, ZCN12) one calibration sample was used for each replicate between both LD and SD samples. Mean values are based on three technical qRT-PCR replicates and three biological replicates for both reference and target genes, ± SEM.

Table 5. Primer sequences used for qRT-PCR analysis.

Gene	Locus ID	Orthologous locus in rice or maize	Forward Sequence	Reverse Sequence	ATx623 (%)	R.07007 (%)	100M (%)	SM100 (%)
PRR37	Sb06g014570	LOC_Os07g49460	AACAGGACGGAACTGGAGAGAGAT	CCAAAGCAATCTTGCTAGAGGCGA	96	98	91	85
FT	Sb10g003940	LOC_Os06g06320	AGCATTTGGGCAAGAGGTGATCTG	AAGTCCCTGGTGTTGAAGTTCTGG	90	96	93	86
CO	Sb10g010050	LOC_Os06g16370	TAGTCCCAGACAACATGGCAACGA	AGGTCAAGTGGAGTGGCATCTGAA	95	87	92	89
Ehd1	Sb01g019980	LOC_Os10g32600	CGTCAGGGAAGCAATGTCCTTCAT	CTTCAGTTGGAAAGCACACATCGC	93	94	92	92
TOC1	Sb04g026190	LOC_Os02g40510	GAGTGCAGATGATTACTGCTCACTTTG	TGCTGCCTTGTTGCCAGTAGAAGA	91	95	88	84
LHY	Sb07g003870	LOC_Os08g06110	GGCCTGCCTCTACCATGAAGTTTA	GCACTGCATTGCAAGGTTTGAAGTCC	91	97	89	86
GI	Sb03g003650	LOC_Os01g08700	ATGCACCCGCTTCCTAGTCATCTT	TTCAGGGCTGTCATGGTTCCTCAT	92	98	92	94
ZCN8	Sb09g025760	GRMZM2G179264	AACTGTCAAAGGGAAGGTGGATCG	GACTAAGCTCTCAACCCTTCAAGTC	88	90	89	90
ZCN12	Sb03g034580	GRMZM2G103666	TGCATGCATGAATATCGTCGTCT	CCCGGGTAGTACATATAAGGTGGT	104	106	107	95
PRR37 intron7	Sb06g014570	LOC_Os07g49460	TGACAGTCACGACAACGAAGCAGA	TCGGCTAACTGATCCAGAGACATTGC	—	—	—	—
PRR37 intron8	Sb06g014570	LOC_Os07g49460	ACCAGGTACGAGCAACAGAAACTG	TTGGTGGGTACTCACAACGTCCAT	—	—	89	92

Gene locus IDs are given based on Phytozome v7.0. Forward and reverse primer sequences and amplification efficiencies for each respective genotype are given. Primer efficiencies were calculated using the standard curve method as described in Materials and Methods.

Absolute quantification at 15 hours after lights on was carried out using the standard curve method (189). PCR-amplified products from each primer set (Ehd1, FT, and Ribosomal 18S mRNA) were purified by gel electrophoresis and sequenced to ensure that no polymorphisms were present between genotypes. The resulting products were subsequently used to construct a dilution series spanning 0.05ng µl-1 to 5.0E-7ng µl-1 for each gene. The number of molecules present in each dilution was calculated from the molecular weight of each product. The dilution series for each gene was then used on a qRT-PCR run in parallel with experimental samples, and absolute transcript abundance was calculated by the ABI HT7900 instrument running SDSv2.3 software. Transcript abundance was then expressed as a ratio of Ehd1 or FT copy number to 18S ribosomal RNA copy number (TaqMan® Universal PCR Master Mix, Applied Biosystems) to obtain relative transcript number. Mean values are based on three technical replicates and three biological replicates for both reference and target genes, ± SD.

Because of difficulties in making primers to differentially detect transcripts varying in only six bases, relative quantification of the splice variant found in intron 7 (Fig. 19*B*) was obtained using fragment analysis as a semi-quantitative method . Fluorescent primers (5-FAM) flanking the six base pair insertion were designed (Table 5). The number of cycles used to amplify each product was based on previous qRT-PCR data to target amplification to the exponential phase. The resulting PCR-amplified fragments were resuspended in a solution of Hi-Di Formamide and GeneScan™ 400HD ROX™ Size Standard and subjected to analysis on the Applied Biosystems 3130xl

Genetic Analyzer. These fragments were also sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) to ensure pure and correct product. The resulting data were then analyzed using Peak Scanner™ Software v1.0 (Applied Biosystems) to obtain the area under the curve for the correctly sized fragment. The area was averaged over three biological replicates ± SEM. The expression of the splice variant at intron 8 (Fig. 19*A*) was analyzed by qRT-PCR as described above.

CHAPTER III

IDENTIFICATION AND CHARACTERIZATION OF *MATURITY LOCUS 6* (Ma_6)

INTRODUCTION

Sorghum [*Sorghum bicolor* (L.) Moench] is a C4 cereal species cultivated widely as a source of food and fodder for millions worldwide (1, 2, 18). Recently, interest has been generated in this grass as a biofuels feedstock, as its drought tolerance allows it to be grown in marginal croplands without displacing designated food crops (2, 6, 23). Sorghum grown for high grain yields and those grown for bioenergy have very different characteristics; for grain production, short, early flowering plants are required for increased yields and mechanized harvest. By contrast, sorghum grown for bioenergy requires very late flowering plants to maximize the production of vegetative biomass throughout the growing season. In sorghum, the time to maturation is controlled largely through response to photoperiod, and as a short day plant, flowering in this species is repressed under long day conditions. However, the extent to which each variety responds to non-inductive photoperiods is regulated by a set of maturity loci, termed Ma_1-Ma_6 (15, 16, 22). Each of these genes acts to repress flowering when dominant, and were utilized in various combinations to confer certain levels of photoperiod sensitivity necessary for the improvement of grain, sweet, and forage sorghum in the U.S., specifically through the sorghum conversion program.

The sorghum conversion program was established in 1963, the purpose of which was to reduce the genetic bottleneck that resulted from extensive selection of short, early flowering varieties by introgressing recessive maturity loci from these photoperiod insensitive lines into diverse late-flowering germplasm (18). Of the maturity loci, Ma_1 is considered the most effective, and was the most extensively used in converting late flowering lines. This gene, which was identified as *PSEUDORESPONSE REGULATOR 37*, inhibits flowering by activating the floral repressor *CO* and repressing floral activators *Ehd1*, *ZCN8*, and *Hd3a* in a manner consistent with the external coincidence model (184). However, during the course of the conversion program, a few varieties were identified that, when crossed to the early flowering ma_1 donor, resulted in offspring that were much later flowering than either parent (22). From these unique genotypes Ma_6 was discovered, and this locus is essential to confer the photoperiod sensitivity required for late flowering bioenergy lines.

In this chapter, the gene underlying Ma_6, was identified as the ortholog of the rice gene *GRAIN NUMBER, PLANT HEIGHT AND HEADING DATE* 7 (*Ghd7*), a major repressor of flowering in long days, using a map-based cloning approach. Sequence analysis from Milo lines with major contributions to the sorghum conversion program as well as from historically important Milo maturity standards revealed a mutation that likely arose very early in sorghum improvement, possibly predating the import of sorghum into the U.S. A second recessive allele was identified that was present in grain and sweet varieties represented most prevalently by the caudatum race. By contrast, multiple functional alleles were uncovered in late flowering bioenergy varieties.

The expression of *Ghd7*, like that of *PRR37*, is gated by the circadian clock and responsive to light, suggesting a mechanism by which these two genes work in tandem to delay flowering under non-inductive conditions. The expression of *CO*, a floral repressor is sorghum, is affected primarily by PRR37, however *Ehd1*, *ZCN8*, and *ZCN12* are repressed to a greater extent under long day conditions in genotypes that posses dominant alleles of both *PRR37* and *Ghd7* than varieties with *PRR37* alone, indicating that Ghd7 down regulates *FT-like* genes synergistically with PRR37, through the *Ehd1* and *CO* pathways, respectively. The data presented in this chapter suggest a model by which these two agronomically important genes act synergistically in the SD plant sorghum to repress *FT*-like genes and flowering through intersecting pathways. Moreover, this study provides novel insight into the selection of early flowering sorghum in the years prior to the initiation of the sorghum conversion program in addition to aiding the development of molecular-based tools that will greatly advance marker assisted breeding efforts in the improvement of all varieties of sorghum.

RESULTS AND DISCUSSION

Map-based Cloning of *Ma_6*

Ma_6 was reported in 1999 as a locus that caused late flowering in the offspring of photoperiod insensitive lines that were crossed to early flowering R.07007 (22). R.07007, which flowers after ~74 days under LD conditions, was crossed to BTx623 (*ma_1*, *Ma_5*, *ma_6*), a photoperiod insensitive variety that flowers after about 56 days in the same conditions. When these parents are crossed, the maturity loci deficiency in each is complemented, producing a photoperiod sensitive F_1, which, because it contains a full set of maturity loci, will not flower for 200 days or more under long day conditions (Fig. 30*A*), Late flowering genotypes such as these have recently generated great interest as a biofuels feedstock, because they continue to accumulate vegetative growth throughout the year and therefore produce much more biomass than parental lines (Fig. 30*B*). Therefore, it has become important to identify the gene underlying the *Ma_6* locus. In order to achieve this goal, BTx623 and R.07007 were used to create a series of mapping populations (Fig. 30*A*), and the photoperiod-insensitive F_1 offspring were grown in short days and selfed

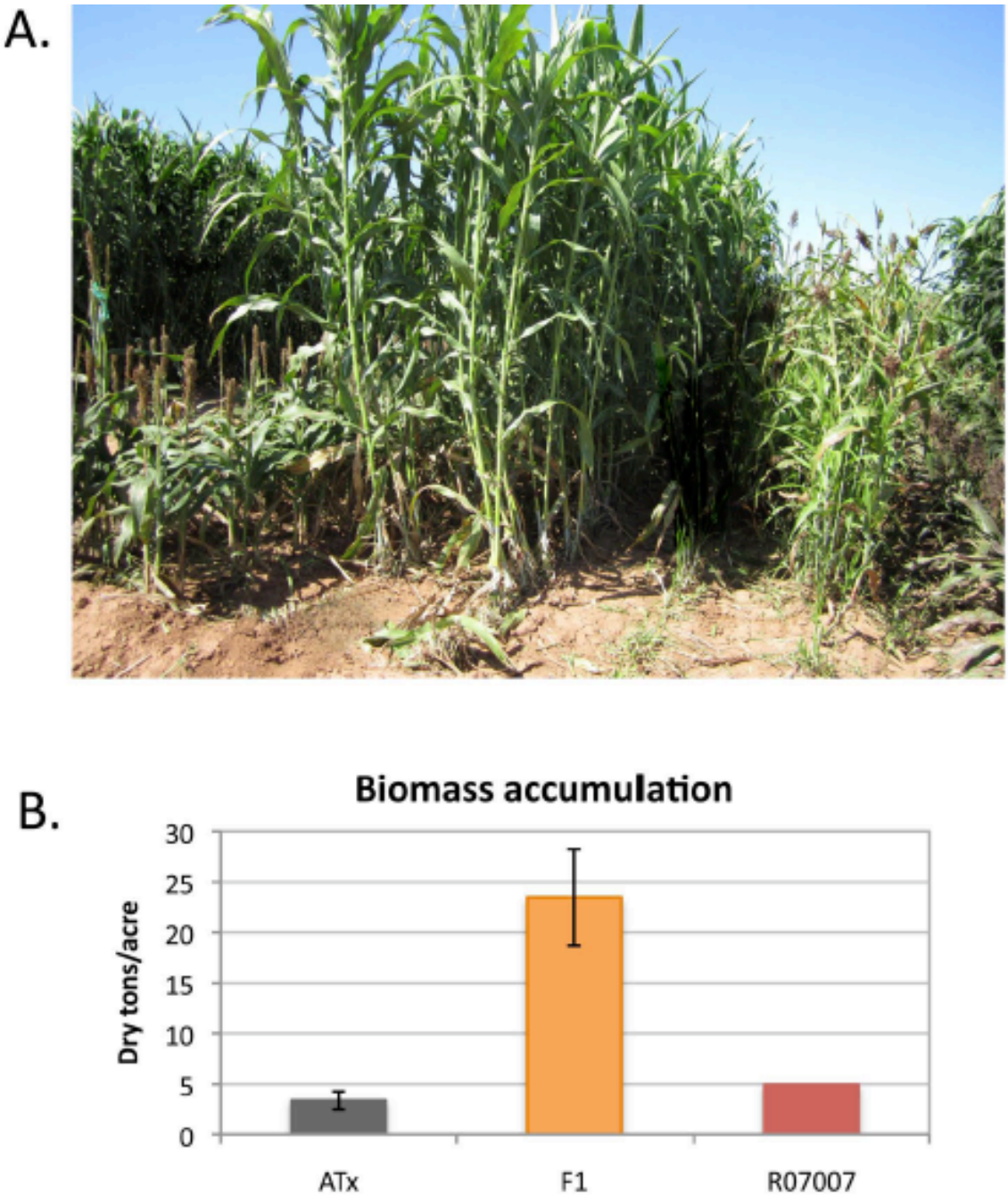

Fig. 30. Phenotypic analysis of field-grown sorghum. (*A*) The ATx623 x R.07007 F_1 (center) is taller than its ATx623 (left) and R.07007 (right) parents. (*B*) Preliminary biomass measurements were made on sorghum test plots (n=3 for ATx623 and the F_1, n=1 for R.07007 due to low germination rates) Plants were grown in 2010 in College Station, TX.

to derive F_2 and F_3 progeny, which were phenotyped for DTF under long day conditions in the greenhouse. High-throughput genotyping of BTx623, R.07007, and individuals from the derived F_2 and F_3 families (n=182) was performed using the Illumina sequencer (Morishige, *et al*., in preparation), and the location of the *Ma*$_6$ QTL was initially mapped to the end of chromosome 6 (Fig. 31). Two other QTL were detected in this analysis that correspond to *Ma*$_5$ (Yang, unpublished) and *Ma*$_1$, of which *Ma*$_1$ was identified in a previous study as *SbPRR37 (22, 184)*. In order to eliminate the effects of secondary loci, and to further define the interval that encodes *Ma*$_6$ additional, recombinants were obtained from a BC_1F_1 (n=1821) population created by backcrossing F_1 plants derived from ATx623 x R.07007 to the male sterile parent ATx623 (Fig. 13*A*). SSRs, SNPs, and INDEL markers were developed spanning the *Ma*$_6$ interval for further refinement of the locus. CAPS markers were developed from known sequence differences in restriction enzyme cut sites that were obtained from genotyping the R.07007 and BTx623 using the Illumina Sequencer. SSR polymorphisms were detected by SSRIT (198), as described above, and SNPs and INDELs were identified through *de novo* sequencing the promoters and intron regions of genes within the *Ma*$_6$ interval. Utilization of these markers for fine mapping in these populations ultimately resulted in a interval delimited by markers SNP1 and SSR5 (Fig. 32), spanning a physical distance of 267kb, within which lie 14 genes (Table 6), as annotated by Phytozome v8.0.

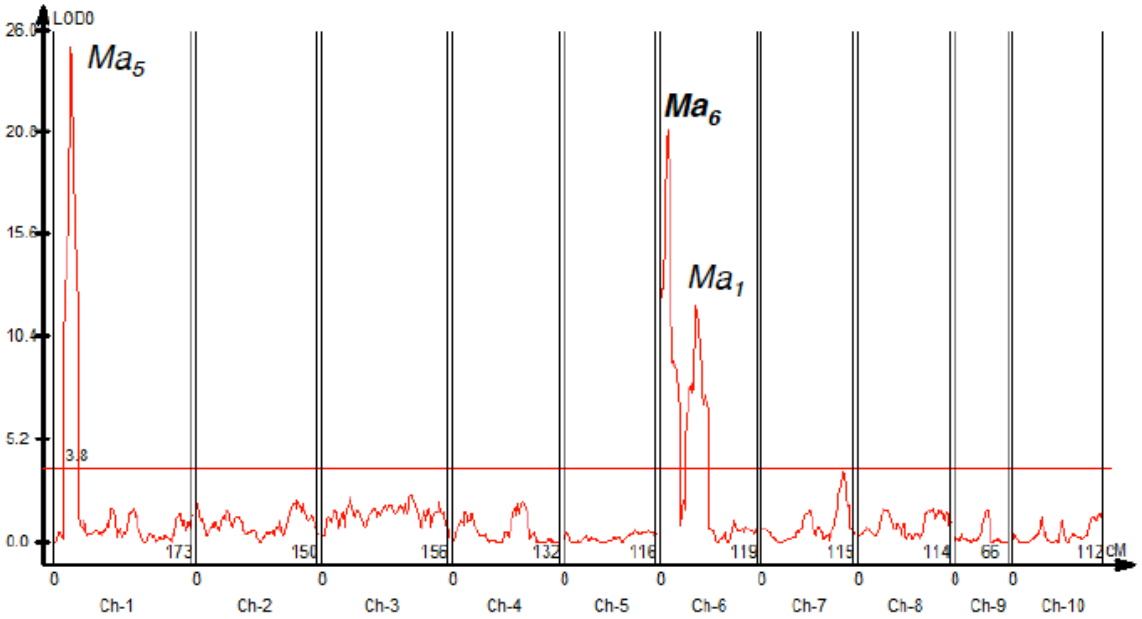

Fig. 31. QTL mapping of *Ma*$_6$. Statistically significant QTL are those with LOD scores that surpass the threshold value (horizontal red line), determined by permutation test at 1000 iterations. Chromosome numbers are given along the abscissa. *Ma*$_1$ and *Ma*$_5$ are known loci; *Ma*$_6$ is presented in bold.

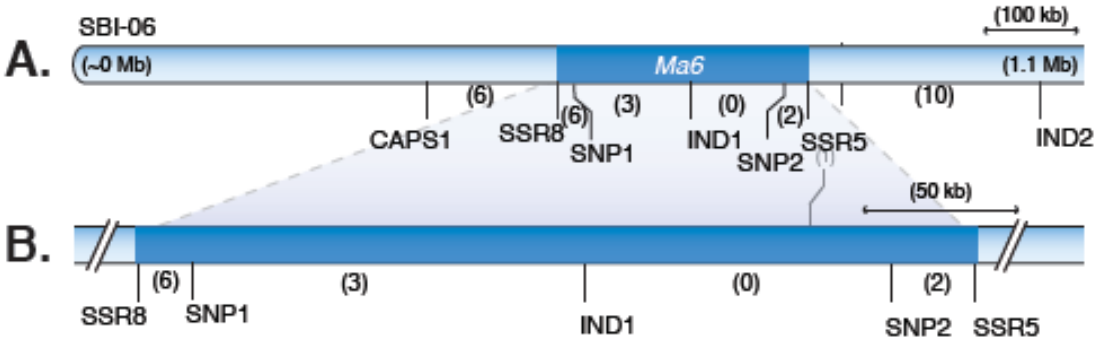

Fig. 32. Fine mapping of *Ma*$_6$. (*A*) The interval was refined to 267kb between SNP1 and SSR5 on the end of chromosome 6. (*B*) Close up of the region in A. IND1 lies in the promoter of the strongest candidate gene, *Grain number, plant height, and heading date* (*Ghd7*) (Table 6). Recombinants are given in parentheses.

Table 6. Genes in the *Ma*$_6$ region.

Gene	Start (b)	Predicted Function
CAPS1	383,829	
SSR8	525976	
Sb06g000495	526,777	nucleobase, nucleoside, nucleotide and nucleic acid metabolism
Sb06g000500	534205	ET TRANSLATION PRODUCT-RELATED
SNP 1		
Sb06g000510	543953	transmembrane transport
Sb06g000520	548793	Predicted transporter (major facilitator superfamily)
Sb06g000530	574537	ET TRANSLATION PRODUCT-RELATED
Sb06g000540	617726	Domain of unknown function (DUF1719)
Sb06g000550	624810	Plant invertase/pectin methylesterase inhibitor
Sb06g000560	662927	Predicted protein
Sb06g000570 (**SNP1**)	670721	Ghd7
Sb06g000580	753105	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amidines
Sb06g000590	758004	Glyoxylate/hydroxypyruvate reductase (D-isomer-specific 2-hydroxy acid dehydrogenase superfamily)
Sb06g000600	761772	Glyoxylate/hydroxypyruvate reductase (D-isomer-specific 2-hydroxy acid dehydrogenase superfamily)
Sb06g000610	769358	Protein of unknown function, DUF573
Sb06g000620	771940	Glyoxylate/hydroxypyruvate reductase (D-isomer-specific 2-hydroxy acid dehydrogenase superfamily)
SNP 2	771940	
Sb06g000630	780552	weakly similar to B0616E02-H0507E05.9 protein
Sb06g000640	792009	Glyoxylate/hydroxypyruvate reductase (D-isomer-specific 2-hydroxy acid dehydrogenase superfamily)
Sb06g000650	793876	Glyoxylate/hydroxypyruvate reductase (D-isomer-specific 2-hydroxy acid dehydrogenase superfamily)
SSR5	798820	
IND2	1,049,495	

Markers and genes used as markers are highlighted in bold type. The physical coordinates and functional annotation of each gene are given as based on Phytozome v8.0

Among these, the strongest candidate for *Ma_6* was Sb06g000570, the sorghum ortholog of rice *Ghd7* as determined by sequence similarity and genomic colinearity across this region if SBI06 (Gramene). Other genes within this interval were sequenced to ensure that no functional polymorphisms existed between the two parents. No obvious functional polymorphisms were detected in ATx623 in the candidate genes sequenced, and so there is no evidence to suggest that the recessive allele contributed by ATx623 is not due to *ghd7*.

Sequencing of the *Ghd7* Gene

To identify functional polymorphisms in *Ghd7* alleles, the gene was sequenced across multiple genotypes. This analysis revealed several neutral SNP polymorphisms, as well as two mutations that result in severe lesions in the protein. The primary functional allele, *GHD7-1*, which is present in R.07007, consists of a 741bp coding region translating to a 246 amino acid product, which contains a CCT domain spanning residues 187–231 (Fig. 33*A*, red; Table 7). Five additional alleles were identified that are presumed dominant based on phenotype and sequencing, containing only slight variations from *GHD7-1*; these are referred to as *GHD7-2* through *GHD7-6* (Table 7). The first recessive allele identified, *ghd7-1*, was found in ATx623 and is characterized by a five base insertion near the end of the first exon, upstream of the CCT domain (Fig. 33*B*). This mutation results in a frameshift and an eventual premature termination of the protein. A second mutation, *ghd7-2*, was characterized by the same five base insertion observed in ATx623, as well as an A → T transversion at position 731 (Table 7).

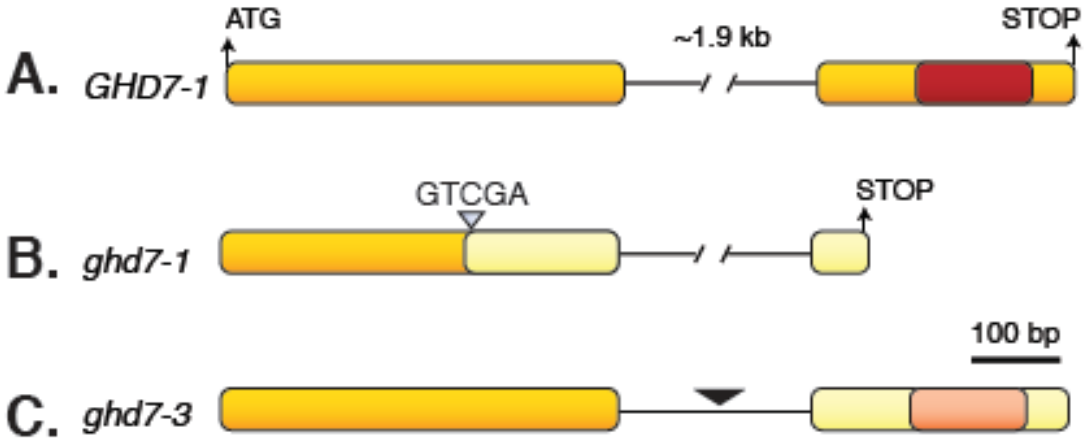

Fig. 33. *Ghd7* allele sequences. (*A*) The dominant *GHD7-1* allele from R.07007 contains a CCT domain (red). (*B*) The *ghd7-1* allele from ATx623 contains a five base insertion in the first exon. This causes a frameshift (light yellow) and a premature stop. (*C*) The *ghd7-3* allele contains a large insertion in the intron, which affects the integrity of the second intron (light yellow, CCT domain, light red).

Table 7. Variation at *Ghd7*.

Allele	58	98	282-286	316	397	409	420-423	512	599	731	738
GHD7-1	C	C	. . . . .	C	A	G	ACGA		T	A	C
GHD7-2	C	C	. . . . .	C	A	G	ACGA		G	A	C
GHD7-3	A	C	. . . . .	C	A	G	ACGA		G	A	C
GHD7-4	C	G	. . . . .	C	A	T	ACGA		G	A	C
GHD7-5	C	C	. . . . .	C	A	G	. . . T		G	A	C
GHD7-6	C	C	. . . . .	A	C	G	. . . T		G		G
ghd7-1	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C
ghd7-2	C	C	GTCGA	C	A	G	ACGA	STOP	G	C	C
ghd7-3	C	C	. . . . .	C	A	G	ACGA		T	A	C

Dominant alleles (*GHD7-1* to *GHD7-6*) contain variation at the single nucleotide level, as well as the deletion of three bases at position 420-423 in *GHD7-5* and *-6* alleles (yellow). Recessive alleles *ghd7-1* and *ghd7-2* contain a five base insertion at position 282-286 that results in a frameshift and a premature stop at position 512. The recessive *ghd7-3* allele contains an insertion in the intron (between position 423 and 512), but is otherwise identical to *GHD7-1*, found in R.07007. The first exon is marked in green, the second in blue.

A third recessive allele, *ghd7-3* was characterized by a chromosomal rearrangement that resulted in the insertion of a repetitive segment of DNA 268bp into the intron, as detected in the Rio genotype (Fig. 33*C*; Fig. 34; Table 7). The size and type of this insertion is not known. Blast analysis reveals the presence of this element throughout the genome, but the occurrence with highest similarity (94%) is about 8.3kb (Phytozome v8.0) downstream of the *Ghd7* gene, suggesting that it may have moved into the intron from this nearby location. Both exons of this gene are still intact, however, the expression of this gene is affected in Hegari (*ghd7-3*) (Fig. 35). Though it cannot be ruled out that differences in genetic background result in the expression difference, it is not due directly to lesions in the upstream sequence of the Hegari *ghd7-3* allele (Fig. 36). All three of these mutations (*ghd7-1*,2,*3*) are correlated with reduced photoperiod sensitivity and earlier flowering in an otherwise dominant maturity locus background, as noted by the difference in flowering times between 100M (120 days, Ma_1-Ma_5, ma_6), and the F_1 (ATx623 x 007) or R07020 (both >200 days, Ma_1-Ma_6).

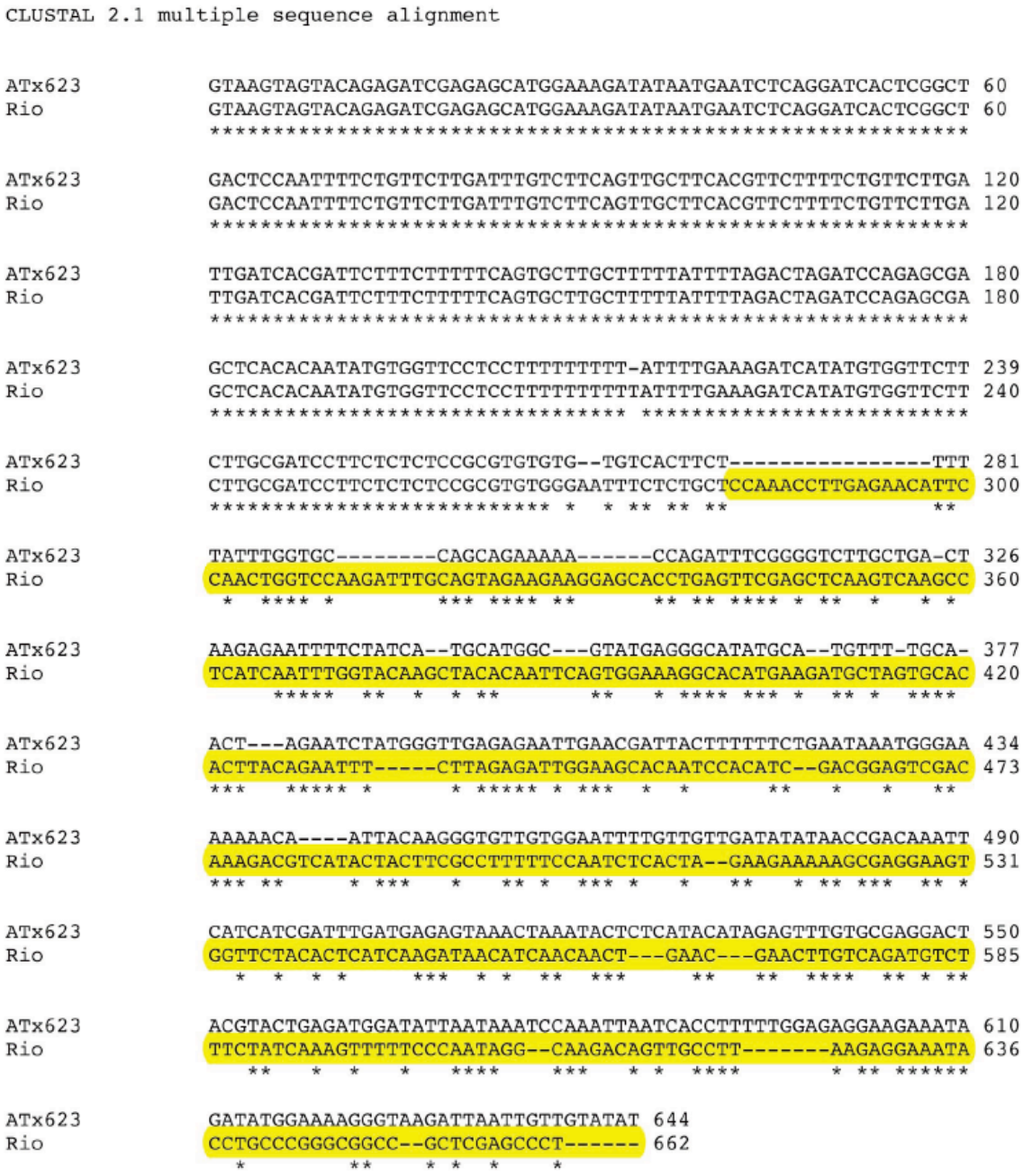

Fig. 34. Alignment of the intron in *ghd7-1* and *ghd7-3* types. ClustalW was used to align a portion of the intron from ATx623 (*ghd7-1*) and Rio, a representative *ghd7-3* genotype. The sequence begins at the first base of the intron sequence. The Rio sequence is highlighted in yellow from the beginning of the insertion until the end of the alignment.

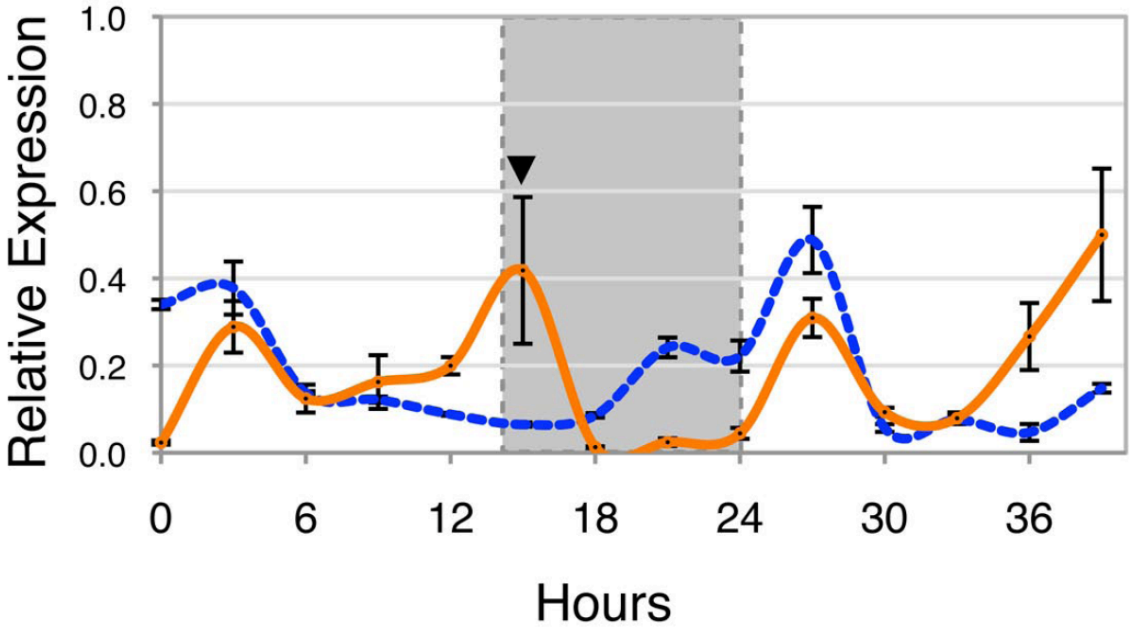

Fig. 35. Expression of the *ghd7-3* allele is altered. Plants were treated under14-h light:10-h dark. The evening peak of expression in *Ghd7* is apparent in the ATx623 x R.07007 F_1 (orange line, arrowhead), but not in Hegari (g*hd7-3*) (blue dashed line). The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample and is based on three biological replicates ± SEM (189).

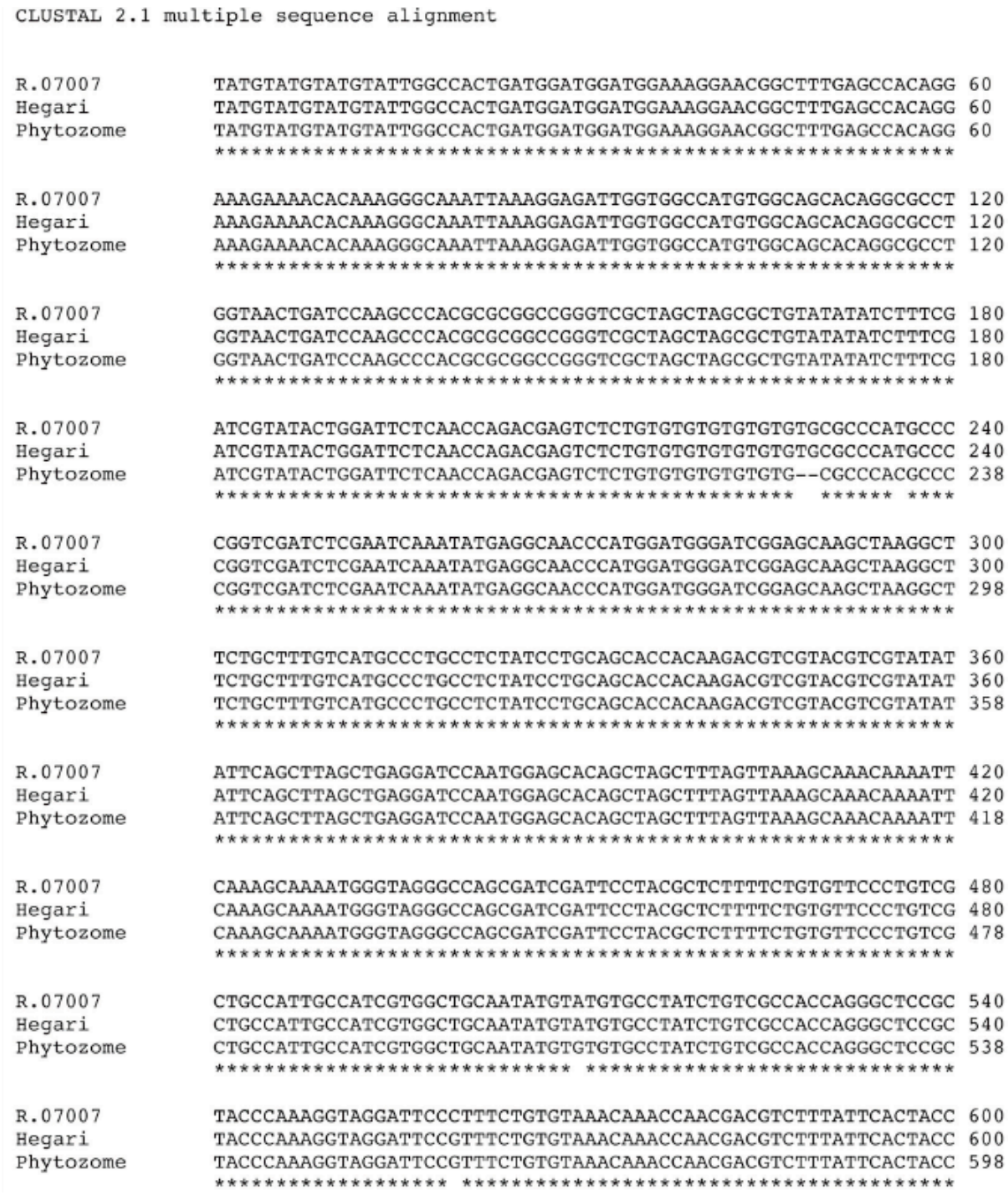

Fig. 36. There are no obvious lesions in the promoter of a *ghd7-3* line. The promoter sequences of R.07007, Hegari, and ATx623 (from Phytozome v8.0) were aligned using ClustalW. R.07007 and Hegari sequences are exactly alike aside from a single SNP at position 560. The ATx623 sequence contains several small differences from R.07007 and Hegari, as well as a large 149b INDEL from position 1400 to 1549. The beginning of the coding sequence is marked in yellow.

```
R.07007         CAAATGCACCATGCATTGCATTGCATCTTCCTGTTCGAGAGTTTTTTAAAAGCCAACACT 660
Hegari          CAAATGCACCATGCATTGCATTGCATCTTCCTGTTCGAGAGTTTTTTAAAAGCCAACACT 660
Phytozome       CAAATGCACCATGCATTGCATTGCATCTTCCTGTTCGAGAGTTTTTTAAAAGCCAACACT 658
                ************************************************************

R.07007         TATATATACTTTCTTTAGGACTAATTGGTAGTAAGCACTTTAACAAGAAATGTACTGGCA 720
Hegari          TATATATACTTTCTTTAGGACTAATTGGTAGTAAGCACTTTAACAAGAAATGTACTGGCA 720
Phytozome       TATATATACTTTCTTTAGGACTAATTGGTAGTAAGCACTTTAACAAGAAATGTACTGGCA 718
                ************************************************************

R.07007         ATAGCTACTACTAGTGTGCTACTCTCGGTGATCAAAAAATACTTGTCATATTTTGCTTTG 780
Hegari          ATAGCTACTACTAGTGTGCTACTCTCGGTGATCAAAAAATACTTGTCATATTTTGCTTTG 780
Phytozome       ATAGCTACTACTAGTGTGCTACTCTCGGTGATCAAAAAATACTTGTCATATTTTGCTTTG 778
                ************************************************************

R.07007         CACGGTATTAATAATTTTATGTGTATTTTGACAACTAATTAACTTTTTGTATAATACAAC 840
Hegari          CACGGTATTAATAATTTTATGTGTATTTTGACAACTAATTAACTTTTTGTATAATACAAC 840
Phytozome       CACGGTATTAATAATTTTATGTGTATTTTGACAACTAATTAACTTTTTGTATAATACAAC 838
                ************************************************************

R.07007         CAAACTTACAATCTCTTAGGCACTACGCCAGAACACGCAATCACTACTAGGTCTTAACAT 900
Hegari          CAAACTTACAATCTCTTAGGCACTACGCCAGAACACGCAATCACTACTAGGTCTTAACAT 900
Phytozome       CAAACTTACAATCTCTTAGGCACTACGCCAGAACACGCAATCACTACTAGGTCTTAACAT 898
                ************************************************************

R.07007         AGGTATATAATATATAACATGCTACACCAGAACACGTAATCACTACTGCCAGGTCACTTA 960
Hegari          AGGTATATAATATATAACATGCTACACCAGAACACGTAATCACTACTGCCAGGTCACTTA 960
Phytozome       AGGTATATAATATATAACATGCTACACCAGAACACGTAATCACTACTGCCAGGTCACTTA 958
                ************************************************************

R.07007         ACACAGGTATATAATAACATGCTTT---GACCTGGTAGTGATAGCGTATTATCACTGCAG 1017
Hegari          ACACAGGTATATAATAACATGCTTT---GACCTGGTAGTGATAGCGTATTATCACTGCAG 1017
Phytozome       ACACAGGTATATAATAACATGCTTTATTGACCTGGCAGTGATAGCGTATTATCACTGTAG 1018
                ************************   ******* ********************** **

R.07007         GGACGACCATATATAATAAACCTATAATGATAACTAGCCCCATCACTACCGATTGTCCTA 1077
Hegari          GGACGACCATATATAATAAACCTATAATGATAACTAGCCCCATCACTACCGATTGTCCTA 1077
Phytozome       GGACGACCATATATAATAAACCTATAATGATAACTAGCCCCATCACTACCGATTGTCCTA 1078
                ************************************************************

R.07007         ATCAATTGCGATGGTGATGGAGGTTTCTTAGCCGACATTGATTGTGTGTACAGGTTAAGG 1137
Hegari          ATCAATTGCGATGGTGATGGAGGTTTCTTAGCCGACATTGATTGTGTGTACAGGTTAAGG 1137
Phytozome       ATCAATTGCGATGGTGATGGAGGTTTCTTAGCCGACATTGATTGTGTGTACAGGTTAAGG 1138
                ************************************************************

R.07007         ACTTCAACTTAGACGAATAGGTTATACCATAAGAAACCCAACCAAGTGAGGTACACACAG 1197
Hegari          ACTTCAACTTAGACGAATAGGTTATACCATAAGAAACCCAACCAAGTGAGGTACACACAG 1197
Phytozome       ACTTCAACTTAGACGAATAGGTTATACCATAAGAAACCCAACCAAGTGAGGTACACACAG 1198
                ************************************************************

R.07007         TTCACATGTGTGTTGAGACTACTTATGCATATTAACGTGAGAATATCCACACTACAACTA 1257
Hegari          TTCACATGTGTGTTGAGACTACTTATGCATATTAACGTGAGAATATCCACACTACAACTA 1257
Phytozome       TTCACATGTGTGTTGAGACTACTTATGCATATTAACGTGAGAATATCCACACTACAACTA 1258
                ************************************************************

R.07007         AAGGATCGGAGCTAGAAGTCAGAGCTAGCAGAAAGGTGGCATCCCTAAGGTCTTGTTTAG 1317
Hegari          AAGGATCGGAGCTAGAAGTCAGAGCTAGCAGAAAGGTGGCATCCCTAAGGTCTTGTTTAG 1317
Phytozome       AAGGATCGGAGCTAGAAGTCAGAGCTAGCAGAAAGGTGGCATCCCTAAGGTCTTGTTTAG 1318
                ************************************************************
```

Fig. 35. Continued.

```
R.07007         TTCGTTAAAATTTTTAGATTTCGTTACTATAGCATTTTCGTTTTTATTTAATAATTATTG 1377
Hegari          TTCGTTAAAATTTTTAGATTTCGTTACTATAGCATTTTCGTTTTTATTTAATAATTATTG 1377
Phytozome       TTCGTTAAAATTTTTAGATTTCGTTACTATAGCATTTTCGTTTTTATTTAATAATTATTG 1378
                ************************************************************

R.07007         TCCAATCATAAACTACCTAGGCT------------------------------------- 1400
Hegari          TCCAATCATAAACTACCTAGGCT------------------------------------- 1400
Phytozome       TCCAATCATAAACTACCTAGGCTCGAAAGATTCGTCTCGCAAATTACATATAAACTGCGT 1438
                **********************

R.07007         ------------------------------------------------------------
Hegari          ------------------------------------------------------------
Phytozome       AATTAGTTATCTTTTTTATCTATATATCTAATGCTTCATGCATGTACCGTAAGATTTGAT 1498


R.07007         ----------------------------------------------------GTTATATTA 1409
Hegari          ----------------------------------------------------GTTATATTA 1409
Phytozome       GTAACGGAGAATTTTGAAAAAAAAAATGGAACTAAACAAGTCCTAAGTTAAGTTATATTA 1558
                                                                   *********

R.07007         TGCACCCAAATGCTTTGGTTTTCTTCGTGCACTCTTGTTTCGTTTCTTTATCTGATCCAC 1469
Hegari          TGCACCCAAATGCTTTGGTTTTCTTCGTGCACTCTTGTTTCGTTTCTTTATCTGATCCAC 1469
Phytozome       TCCACCCAAATGCTTTGGTTTTCTTCGTGCACTCTTGTTTCGTTTCTTTATCTGATCCAC 1618
                * ************************************************************

R.07007         ACATCCAGAGGCAATAACTCCTTGTAGTACTTATCGGTGCAAAGATGCTCTCACACCTAA 1529
Hegari          ACATCCAGAGGCAATAACTCCTTGTAGTACTTATCGGTGCAAAGATGCTCTCACACCTAA 1529
Phytozome       ACATCCAGAGGCAATAACTCCTTGTAGTACTTATCGGTGCAAAGATGCTCTCACACCTAA 1678
                ************************************************************

R.07007         ACTGCAAGCTGCCTAGCTAGCTAGCTTCGATAGATAGATAGATAGATAGGTAGATGAAAC 1589
Hegari          ACTGCAAGCTGCCTAGCTAGCTAGCTTCGATAGATAGATAGATAGATAGGTAGATGAAAC 1589
Phytozome       ACTGCAAGCTGCCTAGCTAGCTAGCTTCGATAGATAGATAGATA----GGTAGATGAAAC 1734
                ******************************************    ************

R.07007         AGAAGGAGTCATAGCCCCTACTGAAATTAAATTAAACAAAGAAATACAACAATAAGCAGT 1649
Hegari          AGAAGGAGTCATAGCCCCTACTGAAATTAAATTAAACAAAGAAATACAACAATAAGCAGT 1649
Phytozome       AGAAGGAGTCATAGCCCCTACTGAAATTAAATTAAACAAAGAAATACAACAATAAGCAGT 1794
                ************************************************************

R.07007         AGCTAAAGAAAAAATAAACCTAAATATGTATTATGCCTAGATACATAACGAAAGCTATGT 1709
Hegari          AGCTAAAGAAAAAATAAACCTAAATATGTATTATGCCTAGATACATAACGAAAGCTATGT 1709
Phytozome       AGCTAAAGAAAAAATAAACCTAAATATGTATTATGTCTAGATACATAACGAAAGCTATGT 1854
                ********************************** ************************

R.07007         GTGTGTATATTTAGAAAAAACTAGAACGTTAGTCGATTTTGAACTAATAATTTATTTAAG 1769
Hegari          GTGTGTATATTTAGAAAAAACTAGAACGTTAGTCGATTTTGAACTAATAATTTATTTAAG 1769
Phytozome       GTGTGTATATTTAGAAAAAACTAGAACGTTAGTCGATTTTGAACTAATAATTTATTTAAG 1914
                ************************************************************

R.07007         CGTTGCTAGCTGACCGGCACGTTATCTCCCTGCTGCCACGGTATACCAGATCGAATCCTG 1829
Hegari          CGTTGCTAGCTGACCGGCACGTTATCTCCCTGCTGCCACGGTATACCAGATCGAATCCTG 1829
Phytozome       CGTTGCTAGCTGACCGGCACGTTATCTCCCTGCTGCCACGGTATACCAGATCGAATCCTG 1974
                ************************************************************

R.07007         TCCATATATATCTATTCTGCCGCGGGCACTGGCTACCAGATTTTGCATGCGATGCACACA 1889
Hegari          TCCATATATATCTATTCTGCCGCGGGCACTGGCTACCAGATTTTGCATGCGATGCACACA 1889
Phytozome       TCCATATATATCTATTCTGCCGCGGGCACTGGCTACCAGATTTTGCATGCGATGCACACA 2034
                ************************************************************
```

Fig. 35. Continued.

```
R.07007         TCTGCTTTTCTGCTGCCTTCCCCTGGCCTTTCCTTTTTCAGTGCTTAACTGCCGCGCGCT 1949
Hegari          TCTGCTTTTCTGCTGCCTTCCCCTGGCCTTTCCTTTTTCAGTGCTTAACTGCCGCGCGCT 1949
Phytozome       TCTGCTTTTCTGCTGCCTTCCCCTGGCCTTTCCTTTTTCAGTGCTTAACTGCCGCGCGCT 2094
                ************************************************************

R.07007         ACCCCCATGGAATTATAACCCTAGCAGCTATAGCTATCAAGCTTTATTTATTCCTTCCTT 2009
Hegari          ACCCCCATGGAATTATAACCCTAGCAGCTATAGCTATCAAGCTTTATTTATTCCTTCCTT 2009
Phytozome       ACCCCCATGGAATTATAACCCTAGCAGCTATAGCTATCAAGCTTTATTTATTCCTTCCTT 2154
                ************************************************************

R.07007         CCTTCCTCTTCTAGCAGCTACACTACACCCTGAACTGCCTCGATCAATATATATATAACT 2069
Hegari          CCTTCCTCTTCTAGCAGCTACACTACACCCTGAACTGCCTCGATCAATATATATATAACT 2069
Phytozome       CCTTCCTCTTCTAGCAGCTACACTACACCCTGAACTGCCTCGATCAATATATATATAACT 2214
                ************************************************************

R.07007         CAATTCAACAAAGTCATCATCTCCCTCCCCAGTTGCCGATCGATCGATCCATCCATCCAT 2129
Hegari          CAATTCAACAAAGTCATCATCTCCCTCCCCAGTTGCCGATCGATCGATCCATCCATCCAT 2129
Phytozome       CAATTCAACAAAGTCATCATCTCCCTCCCCAGTTGCCGATCGATCGATCGATCCATCCAT 2274
                ********************************************** **********

R.07007         CTTTTAATTTGCTCCAAACCTATCCACACAGCAGACCTGAGATCGAGCTGTGCGCGCGAT 2189
Hegari          CTTTTAATTTGCTCCAAACCTATCCACACAGCAGACCTGAGATCGAGCTGTGCGCGCGAT 2189
Phytozome       CTTTTAATTTGCTCCAAACCTATCCACACAGCAGACCTGAGATCGAGCTGTGCGCGCGAT 2334
                ************************************************************

R.07007         CGATCGATCGATCCGTTCACATGTCAGGGCCAGCATGCGGTGTGTGCGGTGCAGCCGCCT 2249
Hegari          CGATCGATCGATCCGTTCACATGTCAGGGCCAGCATGCGGTGTGTGCGGTGCAGCCGCCT 2249
Phytozome       CGATCGATCGATCCGTTCACATGTCAGGGCCAGCATGCGGTGTGTGCGGTGCAGCCGCCT 2394
                ************************************************************

R.07007         GCTGCCGGCACCTCTTCCACACCGGCGAC 2278
Hegari          GCTGCCGGCACCTCTTCCACACCGGCGAC 2278
Phytozome       GCTGCCGGCACCTCTTCCACACCGGCGAC 2423
                *****************************
```

Fig. 35. Continued.

Ghd7 Allelic Distribution in Sorghum Germplasm

Ma_6 is dominant in R.07007, which was introduced from Argentina, while most American sorghum is predicted to be recessive at this locus (22). Therefore, in order to understand the potential sources and distribution of the GHD7 alleles in germplasm, the sequence of this gene was obtained from historical, grain, sweet, and bioenergy varieties. The very late flowering bioenergy varieties R.07018, R.07019, and R.07020 are presumed dominant at all maturity loci, and contain *GHD7-6, -2,* and *-5* alleles,

respectively. These lines are extremely photoperiod sensitive and do not initiate flowering until >175 days under LD conditions (Table 8). Likewise, IS12646, an unconverted line which encodes *GHD7-3*, remains vegetative for >175 days. By contrast, 100M (*ghd7-1*), dominant at Ma_1- Ma_5, flowers around 130 days under 14 hour day lengths. Therefore, the dominant *Ghd7* allele is associated with this substantial increase in flowering time.

Table 8. *Ghd7* alleles in bioenergy lines.

	58	98	282-286	316	397	409	420-423	512	599	731	738	Ma_6	Ma_1
R.07007	C	C	.....	C	A	G	ACGA		T	A	C	*GHD7-1*	*PRR37*
R.07019	C	C	.....	C	A	G	ACGA		G	A	C	*GHD7-2*	*PRR37*
IS12646	A	C	.....	C	A	G	ACGA		G	A	C	*GHD7-3*	*PRR37*
R.07020	C	C	.....	C	A	G	...T		G	A	C	*GHD7-5*	*PRR37*
R.07018	C	C	.....	A	C	G	...T		G		G	*GHD7-6*	*PRR37*

Lines for potential bioenergy feedstock are dominant at both *Ghd7* and *PRR37* (Ma_6 and Ma_1). The first exon is marked in green, the second in blue. Variation from the R.07007 sequence is marked in yellow.

In the 1940s, a set of maturity standards were developed that are nearly isogenic and possess the same genetic background except for variation at specific maturity loci, (199). Four of these genotypes, 100M, 80M, SM100, and 58M were grown in LD and SD, and phenotyped for DTF under each photoperiod by Dr. Daryl Morishige. The difference in flowering time observed under LD and SD was used as a measure of photoperiod sensitivity. 100M, which has the maturity genotype Ma_1-Ma_5, ma_6, is the

most photoperiod sensitive of the lines developed by Quinby. A 73-day difference is observed between LD- and SD-grown 100M. By contrast, SM100, which is recessive at both *Ma*$_6$ and *Ma*$_1$, is completely photoperiod insensitive, flowering with a difference of three days under contrasting photoperiods. Similarly, 58M, recessive at *Ma*$_6$ and *Ma*$_3$, flowers only two days earlier in SD than in LD. It is not surprising that the phenotype of plants isogenic except at *Ma*$_1$ and *Ma*$_3$ (both are *ma*$_6$) should exhibit the same degree of photoperiod sensitivity, as it is predicted that *Ma*$_3$, which is phytochrome B, is epistatic to *Ma*$_1$ (*SbPRR37*). It is thought that PhyB regulates the most upstream component of the light-dependent transcriptional regulation of this *SbPRR37*. 80M, which is recessive at both *Ma*$_6$ and *Ma*$_2$, is delayed by 36 days under LD conditions relative to SD. That 80M exhibits a photoperiod sensitivity between that observed for 100M (*ma*$_6$) and SM100 (*ma*$_1$, *ma*$_6$) is consistent with prior observations. The nearly identical genetic backgrounds of these maturity standards allow for a comparison of photoperiod sensitivity as a result of specific combinations of the first three maturity loci in an otherwise uniform null *ghd7-1* background. However, a standard that is dominant at

Table 9. *Ghd7* alleles in Milo maturity cultivars.

	58	98	282-286	316	397	409	420-423	512	599	731	738	*Ma*$_6$	*Ma*$_1$
100M	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
80M	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
SM100	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-3*
58M	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*

Ghd7 and *PRR37* alleles are given in the two right-most columns for each genotype. The first exon is marked in green, the second in blue. Variation from the R.07007 sequence is marked in yellow.

Ma_6 does not exist among these four lines, or among other maturity genotypes. The absence of a functional *Ghd7* allele in much of the early sorghum germplasm suggests that a mutation in *Ghd7* arose before the development of these lines. Therefore, *Ghd7* was sequenced in these four genotypes in order to verify this assumption, and in fact, all four harbored an identical *ghd7-1* allele, reinforcing this hypothesis (Table 9).

Because the alleles present in the maturity standards were all identical, the *Ghd7* allele in historic Milos involved in the sorghum conversion program was also analyzed. The primary goal of the conversion program was to convert novel germplasm with tall, late-flowering phenotypes to short, early-flowering varieties while retaining the beneficial variation found in exotic lines. This was done by crossing the exotic line to a genotype with recessive alleles at *Ma_1* and *Dw_2* (18). Though several different lines with recessive *Ma_1* were used, BTx406 was used most extensively for this purpose. The pedigree of BTx406 with respect to *Ma_1* has been traced through the genotypes utilized in its development (21) (Fig. 37), and although BTx406 contains the *ghd7-1* allele, the source of this allele in conversion materials is unknown. Therefore, in order to determine if *ghd7* was introduced at a specific point in the construction of BTx406, the lines used to develop this genotype were sequenced at this locus (Table 10). DDYM (Double Dwarf Yellow Milo) and DYM (Dwarf Yellow Milo) possess dominant *PRR37* alleles as predicted, and they also possess recessive *ghd7-1* alleles identical to the one found in EWM (Early While Milo) and the maturity Milos. The *ma_1* and *ma_6* alleles were carried further through SA403 through the crossing of DDWS (Double Dwarf White Sooner

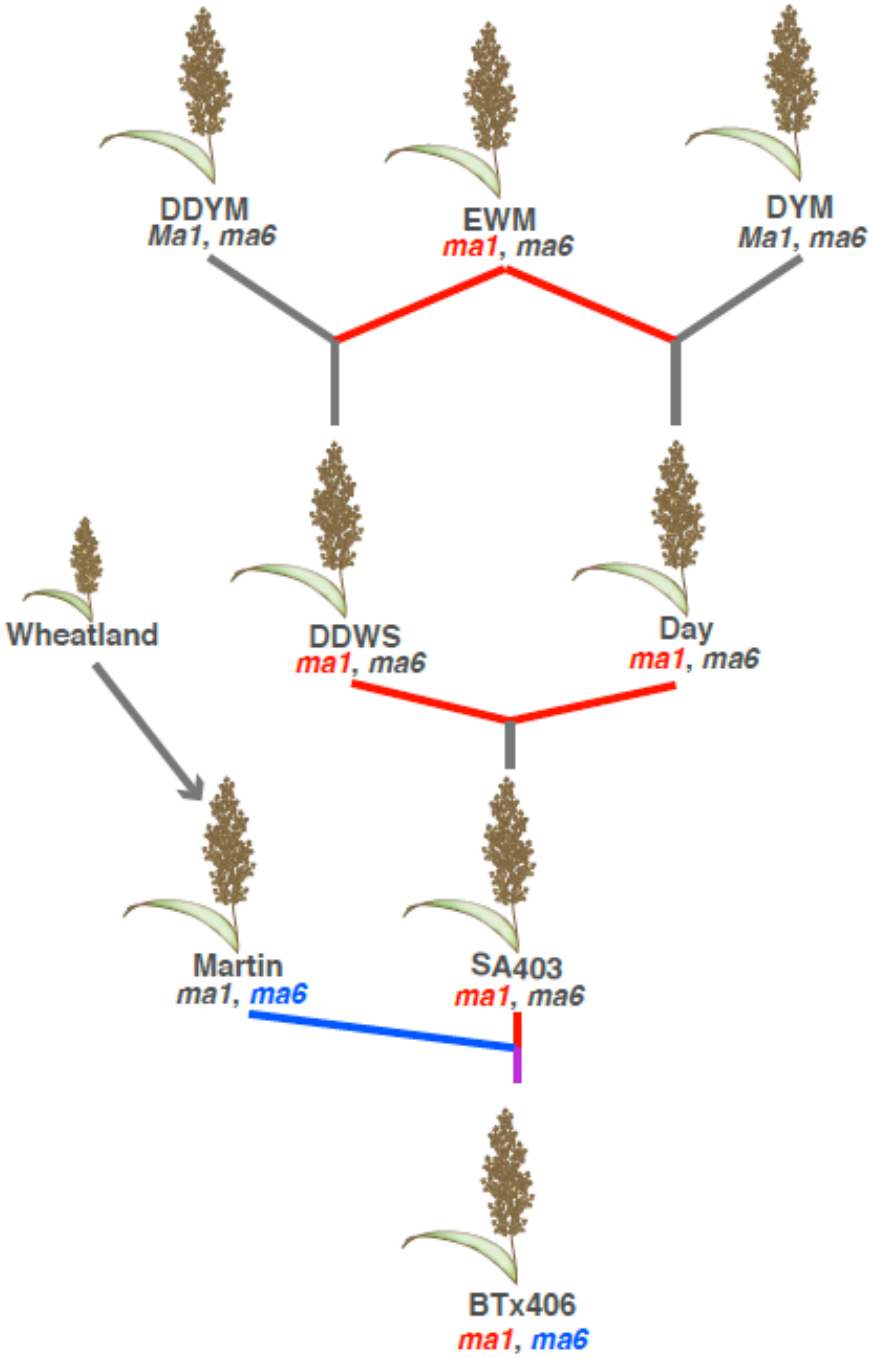

Fig. 37. Pedigree of BTx406. The *ma*$_1$ allele (red) donated from EWM can be traced through DDWS (Double Dwarf White Sooner Milo), Day, and SA403 (red line), though Martin possesses an identical *prr37-1* allele. Based on haplotype analysis (21), it is likely that the *ghd7-1* allele (blue) was obtained from the Martin cultivar (blue line).

Milo) and Day, the offspring of DDYM x EWM and DYM x EWM, respectively. In the final phase of BTx406 development, SA403 (*ma*$_1$, *ma*$_6$) was crossed to Martin (Tx399),a derivative of the Kafir-Caudatum line, Wheatland (Tx398). Haplotype analysis of this region of chromosome 6 suggests that a block of DNA from Martin was transferred to BTx406 in the region spanning the end of SBI06, where *Ghd7* is located (21). Wheatland and Martin possess an allele identical to those contained in the Milos used to derive SA403 and BTx406, reinforcing the hypothesis that mutations in *Ghd7* arose very early in sorghum improvement in the United States, or present in germplasm imported from Africa.

Table 10. *Ghd7* alleles in historical Milo cultivars.

	58	98	282-286	316	397	409	420-423	512	599	731	738	*Ma*$_6$	*Ma*$_1$
EWM	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-1*
DYM	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
DDYM	C	C	GTCGA	C	A	G	ACGA	STOP	G			*ghd7-1*	*PRR37*
Day	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-1*
Martin	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	-
BTx406	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-1*
WTLD	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	-
IS12464c	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	-

Ghd7 and *PRR37* alleles are given in the two right-most columns for each genotype. EWM (Early White Milo); DYM (Dwarf Yellow Milo); DDTM (Double Dwarf Yellow Milo); WTLD (Wheatland). The first exon is marked in green, the second in blue. Variation from the R.07007 sequence is marked in yellow. The dark gray portion represents missing sequence at that position, and gray text in the allele column represents the inferred genotype.

Moreover, though it has been well established that late flowering exotic germplasm was converted at Ma_1, it is important to assess whether Ma_6 represents a second, necessary conversion point. It is hypothesized that conversion to recessive ma_6 is necessary for sufficiently reduced photoperiod sensitivity, but because many lines were Ma_6 recessive prior to conversion, it is difficult to determine this. However, one example that supports this hypothesis is the line IS12646 (*GHD7-3*; Table 8). This line was converted by introgressing the BTx406 ma_1 allele into IS12646, creating IS12646c. IS12646c (Table 10) contains the *prr37-1* and *ghd7-1* alleles from BTx406, consistent with the hypothesis that Ma_6 may have been required for conversion of some lines to obtain sufficiently early flowering for grain production in addition to Ma_1.

Table 11. *Ghd7* alleles in historical Kafir cultivars.

	58	98	282-286	316	397	409	420-423	512	599	731	738	Ma_6	Ma_1
TX BHK	C	C	GTCGA	C	A	G	ACGA	STOP	G	C	C	*ghd7-2*	*prr37-3*
ST BHK	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-2*
CK60	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-3*
COM7078	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-1*
TX7000	C	C	GTCGA	C	A	G	ACGA	STOP				*ghd7-1*	*prr37-2*
ATx623	C	C	GTCGA	C	A	G	ACGA	STOP	G	C	C	*ghd7-2*	*prr37-3*

Ghd7 and *PRR37* alleles are given in the two right-most columns for each genotype. TX BHK (Texas Blackhull kafir); ST BHK (Standard Blackhull kafir); CK60 (Combine kafir 60); COM7078 (Combine 7078). The first exon is marked in green, the second in blue. Variation from the R.07007 sequence is marked in yellow. The dark gray portion represents missing sequence at that position.

The *ghd7-1* allele found in the Milo genotypes likely predates their utilization in the United States, however; many sorghum varieties grown in the United States today were initially founded not only by Milos, but also by Kafirs and Kafir crosses. Therefore, it is also important to analyze historical Kafir varieties at this locus. All of the Kafirs analyzed are recessive at Ma_1 and Ma_6, though the *ghd7* alleles found in Kafirs are not entirely identical to those in Milos (Table 11). ST BHK (Standard Blackhull Kafir), CK60 (Combine Kafir 60), and Combine 7078 (COM7078) encode a *ghd7-1* allele like that in the Milos, however; a SNP at position 738 is present in TX BHK and its derivative ATx623, which represent the *ghd7-2* allele. The *ghd7-2* allele is less common than *ghd7-1* in the Kafir background of those genotypes analyzed, which

Table 12. *Ghd7* alleles in grain sorghum.

	58	98	282-286	316	397	409	420-423	512	599	731	738	Ma_6	Ma_1
Spur Feterita	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-4*
IS27034	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-4*
IS8525J	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-1*
BTx631	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
RTx430	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-1*
M35-1	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
IS3620C	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-1*
BTx642	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
SC56	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*

Ghd7 and *PRR37* alleles are given in the two right-most columns for each genotype. The first exon is marked in green, the second in blue. Variation from the R.07007 sequence is marked in yellow.

suggests that *ghd7-1* may be the older allele. Moreover, much of the grain sorghum varieties grown in the United States at present are derivatives of several races, including Milos, Kafirs, or some mixture of the two. When the sequence of the *Ghd7* gene was analyzed in several grain types, they all had identical copies of the *ghd7-1* allele (Table 12), further suggesting that this allele arose early or prior to sorghum importation and utilization in the United States. It is also important examine the alleles in other sorghum functional groups, namely sweet sorghum. Therefore *Ghd7* was sequenced from several of these cultivars, and the *ghd7-1* allele was predominant among a specific subset of these varieties (Table 13), further suggesting that an early mutation in this gene is associated sorghum development in the U.S.

Table 13. *Ghd7* alleles in sweet sorghum.

	58	98	282-286	316	397	409	420-423	512	599	731	738	Ma_6	Ma_1
SUMAC	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	-
DELLA	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
N100	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-2*
M81E	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*PRR37*
Wray	C	C	GTCGA	C	A	G	ACGA	STOP	G	A	C	*ghd7-1*	*prr37-2*

Ghd7 and *PRR37* alleles are given in the two right-most columns for each genotype. The first exon is marked in green, the second in blue. Variation from the R.07007 sequence is marked in yellow.

A third allele was found within sorghum germplasm, represented in grain and sweet varieties and across various races. This allele, *ghd7-3*, is characterized by an insertion of a predicted repetitive element in the intron of the gene, and is associated

with misexpression of the gene (Fig. 35). Several *ghd7-3* allele variants containing the intronic insertion were identified (Table 14), and among the Kafir group are Red and Pink Kafir, Rio (a sweet variety), and Kalo, with a Kafir x Milo pedigree. However, the most represented race is caudatum, including the No.1 Gambelas (converted and preconverted), DDF (Double Dwarf Feterita), SC170, Hegari, and Grassl. Because this allele is found more frequently in caudatum lines than other races, this suggests that the

Table 14. Varieties with the *ghd7-3* allele.

	58	98	282-286	316	397	409	420-423	512	599	731	738	Ma_6	Ma_1
No. 1 Gambela (C)	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*prr37-1*
RED KAFIR	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*prr37-2*
KALO	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*prr37-2*
DDF	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*prr37-3*
SC170	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*prr37-2*
RIO	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*prr37-2*
PINK KAFIR	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*prr37-2*
Hegari	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*PRR37*
P898012	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*PRR37*
NO. 1 Gambela (P)	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*PRR37*
IS 12666	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*PRR37*
GRASSL	C	C	. . . . .	C	A	G	ACGA		T	A	C	*ghd7-3*	*PRR37*

Ghd7 and *PRR37* alleles are given in the two right-most columns for each genotype. The first exon is marked in green, the second in blue. Variation from the R.07007 sequence is marked in yellow.

mutation arose in imported caudatum germplasm prior to its utilization in U.S. sorghum breeding. Regardless of the origins of the recessive alleles, each is associated with

decreased photoperiod sensitivity, and has been utilized to varying extents in sorghum germplasm.

Ghd7 Expression is Regulated by Light and the Circadian Clock

The expression of *Ghd7* has been shown to be responsive to phytochrome-mediated light signals in rice such that expression is diminished in SD (166). Therefore, in order to gain insight into the molecular mechanism by which Ghd7 represses flowering in sorghum, its expression was measured by qRT-PCR in 100M (Ma_1-Ma_5, ma_6) and the ATx623 x R.07007 F_1 (Ma_1-Ma_6) under LD or SD followed by a period of free-running LL. In rice, a single peak of *Ghd7* expression occurs around dawn in long days (166). However, in sorghum, the expression of *Ghd7* in the F_1 and 100M is bimodal, peaking 3 hours (Fig. 38, *A*-*C*, arrow) and 15 hours (Fig. 38, *A*-*C*, arrowhead) after initial light exposure. In SD, the second peak in expression is diminished. However, when these SD-treated plants are allowed to free run in LL, the evening peak in expression reappears (Fig. 38, *A* and *C*, 39 hours). This expression pattern is quite similar to that observed for *SbPRR37*, and suggests a requirement for light. Therefore, tissue was collected from plants grown in LD and released into DD instead of LL. The mRNA abundance of *Ghd7* peaks normally during the 24 hour LD cycle, but is substantially dampened during the DD cycle, showing a requirement for light in the afternoon induction of *Ghd7* expression (Fig. 38*B*). Taken together, these data strongly suggest that in order to up-regulate Ghd7 expression differentially under a LD photoperiod, light exposure must occur during a specific time gated by the circadian

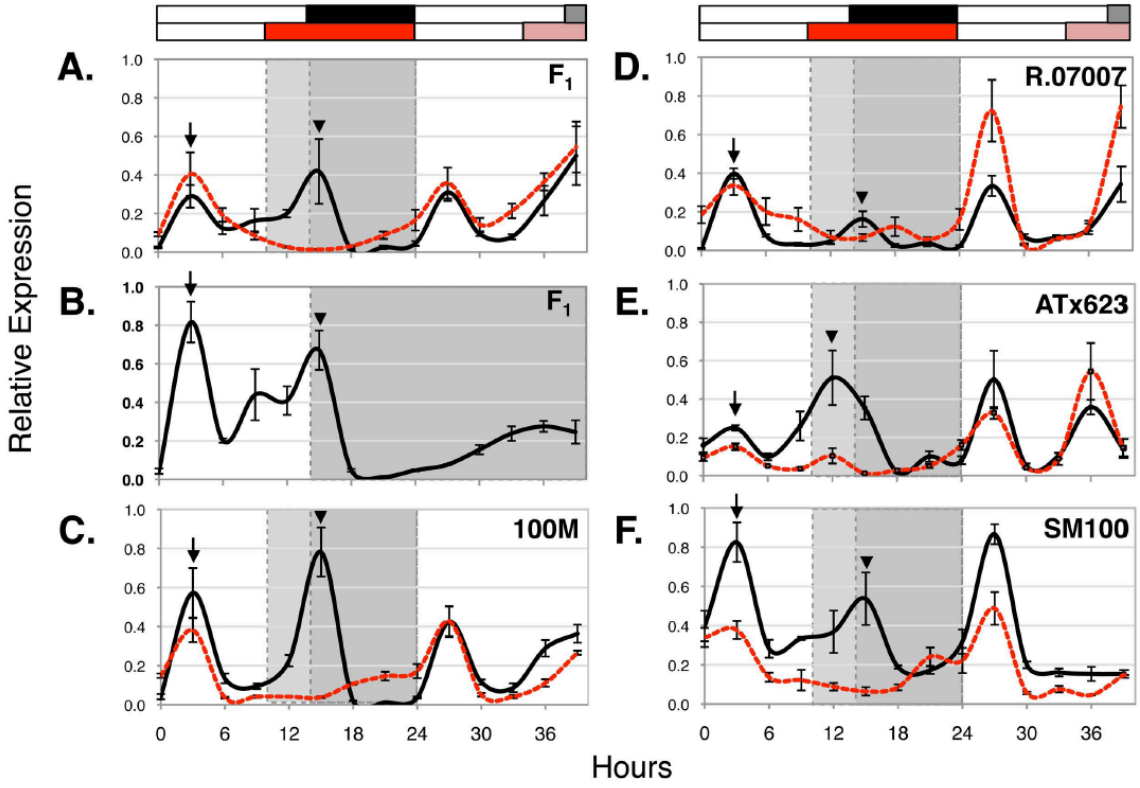

Fig. 38. *Ghd7* expression is clock- and light-dependent. Plants were treated under14-h light:10-h dark (LD, solid line) or 10-hlight:14-h dark (SD, red dashed line) conditions. Relative expression of *Ghd7* was analyzed at 3-h intervals by quantitative RT-PCR. In ATx623 x R.07007 F_1 (*A* and *B*) and 100M plants (*C*), *Ghd7* expression increased in the morning (arrow) and evening (arrowhead) of long days. (*B*) The expression of Ghd7 is diminished in ATx623 x R.07007 F_1 plants grown in 14-h:10-h LD and then released into DD at time 24 h. The evening peak in expression is shifted 3 hours late in (*D*) R.07007 (arrowhead) and 3 hours early in (*E* and *F*) ATx623 and SM100 (arrowhead) relative to the F_1. The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar above the plot indicates the dark period for LD-treated plants; gray bars indicate subjective dark during LL conditions. Red bars indicate darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. Light-gray shading within the plot area indicates darkness for SD treated plants only; dark-gray shading indicates darkness for both LD- and SD-treated plants.

clock. This is consistent with what was observed for *Ghd7* in rice (166) and for sorghum *SbPRR37* (184).

In order to more closely examine the importance of the LD-specific evening peak of Ghd7, its expression was analyzed in R.07007, which has the functional *GHD7-1* allele but flowers early due to a deficiency at other loci. The morning peak in expression occurs normally in this genotype (Fig. 38*D*, arrow), but the evening peak in expression is diminished in LD to levels comparative with SD (Fig. 38*D*, arrowhead). When this genotype is transferred to LL conditions, both peaks reappear. However, in the F_1 and 100M, expression of *Ghd7* begins to increase around 33 hours for the second peak in LL conditions (Fig. 38, *A*-*C*). In R.07007, initial up-regulation of this second LL peak occurs later at 36 hours (Fig. 38*D*). In order for the afternoon peak of *ghd7* expression to occur, the clock-controlled inductive signal must be coincident with light. The period of Ghd7 in R.07007 is lengthened by three hours, which results in the reception of this positive signal during a period of darkness. It is presumed that the expression of *Ghd7* is controlled by a second locus, likely Ma_5, which is recessive in this background, because this aberrant expression is complemented in the F_1. In fact, ATx623, the other parent of the F_1, also exhibits disruptions in the period length of expression, but in the opposite direction. The expression of the first peak *Ghd7* in ATx623 (*ghd7-2*) occurs at 3 hours as observed for all other genotypes (Fig. 38*E*, arrow). However, the second peak occurs at 12 hours instead of 15 (Fig. 38*E*, arrowhead). This 3-hour early shift is reiterated in the expression patterns of LL treated sorghum (Fig. 38). The second LL peak occurs at 36 hours and has already decreased in amplitude by 39 hours in contrast to the F_1.

Moreover, SM100 (*ghd7-1*) also exhibits altered *Ghd7* expression, and though the pattern in LD/SD appears to be similar to the F_1, the second peak never reappears under LL (Fig. 38). The mechanism by which the period shift occurs in the PI genotypes R.07007, ATx623, and SM100 is not understood, but in R.07007, a *Ghd7* dominant line, the decreased afternoon expression of this gene likely underlies some portion of its early flowering phenotype. Additionally, like *PRR37*, *Ghd7* expression is activated by the clock at precise intervals throughout the day, and when paired with light exposure represses flowering, consistent with the external coincidence model.

Ghd7 Affects Downstream Floral Activators

In order to understand the effects Ghd7 has on downstream flowering genes, the expression of *Ehd1*, a downstream target of Ghd7 repression in rice, was analyzed in 100M and the F_1, which vary only at Ma_6 with respect to the maturity loci. In both F_1 and 100M, the expression of *Ehd1* in LD is greatly repressed relative to SD (Fig. 39*A* and *B*). However, in the F_1, *Ehd1* is repressed to a much greater extent than in 100M (Fig. 39*C*), and in fact, at certain time points its expression is undetectable, even at 40 cycles (Fig. 39*A* and *C*, open circles). This suggests that while *Ehd1* is repressed to some extent in 100M, it is repressed to a much greater extent in the *Ghd7* dominant F_1 (Ma_1Ma_6), which delays flowering time more than 80 days relative to 100M. In R.07007, *Ehd1* is repressed under LD (Fig. 39*D*). However, the difference between LD and SD is not as great as that observed in the F_1. By contrast, in ATx623, and especially SM100 (both *ghd7-1*), the expression of *Ehd1* in LD and SD occurs at similar

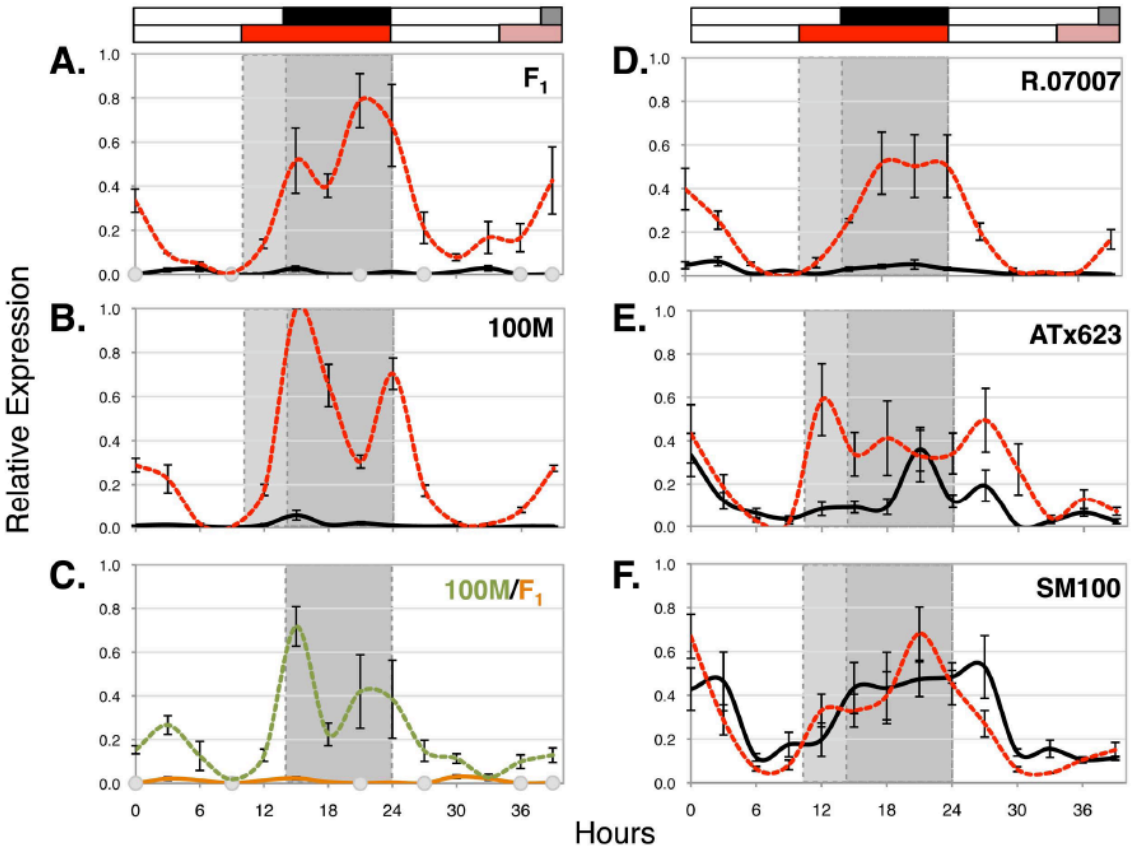

Fig. 39. *Ehd1* expression in strongly repressed in LD in *GHD7* plants. Plants were treated under14-h light:10-h dark (LD, solid line) or 10-hlight:14-h dark (SD, red dashed line) conditions. Relative expression of *Ehd1* was analyzed at 3-h intervals by quantitative RT-PCR. In ATx623 x R.07007 F_1 and 100M (*A* and *B*), *Ehd1* is strongly repressed. However, *Ehd1* is much more strongly repressed in LD-treated F_1 (*Ghd7-1*, *PRR37*, orange line) than in 100M (*ghd7-1*, *PRR37*, green dashed line) (*C*). Open circles in F_1 expression in (*A*) and (*C*) represent points where *Ehd1* expression was too low to be detected in any biological or technical replicate at 40 cycles. (*D*) LD-treated R.07007 plants (*Ghd7-1*, *PRR37*, ma_2ma_5) display some repression in LD, while *Ehd1* is de-repressed to near-SD levels in ATx623 and SM100 (*ghd7-1*, *prr37*) (*E* and *F*). The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar above the plot indicates the dark period for LD-treated plants; gray bars indicate subjective dark during LL conditions. Red bars indicate darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. Light-gray shading within the plot area indicates darkness for SD treated plants only; dark-gray shading indicates darkness for both LD- and SD-treated plants.

levels (Fig. 39, *E* and *F*). Because *Ghd7* encodes a functional protein in R.07007, it is likely that, though it is misexpressed, enough Ghd7 protein is produced to elicit a response to LD.

In rice, Ehd1 directly up-regulates *Hd3a* and *RFT1* expression (163). However, an *RFT1* ortholog has not been identified in the sorghum genome. Moreover, sorghum possesses a set of *FT*-like PEBP family genes similar to the ZCN set found in maize. One of these, *ZCN8*, was determined to be the most likely candidate for florigen in maize because of its transcriptional response to photoperiod (173). Additionally, expression of this gene has been shown to be differentially responsive to photoperiod in sorghum Ma_1 nearly isogenic lines, further implicating it as a component of the florigen signal. Therefore, in order to determine the extent to which Ghd7 influences the transcription of this gene, its expression was analyzed in *Ghd7-1* and *ghd7-1*/*2* backgrounds.

In the ATx623 x R.07007 F_1, *ZCN8* is highly repressed in LD (Fig. 40*A*). Similarly, its expression is greatly reduced in LD-treated 100M (Fig. 40*B*). This repression occurs to a greater extent in the LD-treated F_1 relative to 100M in similar conditions (Fig. 40*C*), though the difference in LD repression between the two genotypes is not as great as that observed for *Ehd1* (Fig. 39*C*). It is not surprising that the repression difference observed for *Ehd1* between the two genotypes does not translate to an equivalent difference in downstream *ZCN8* expression. One property of florigen that exists in all species studied is that it acts as a floral integrator,

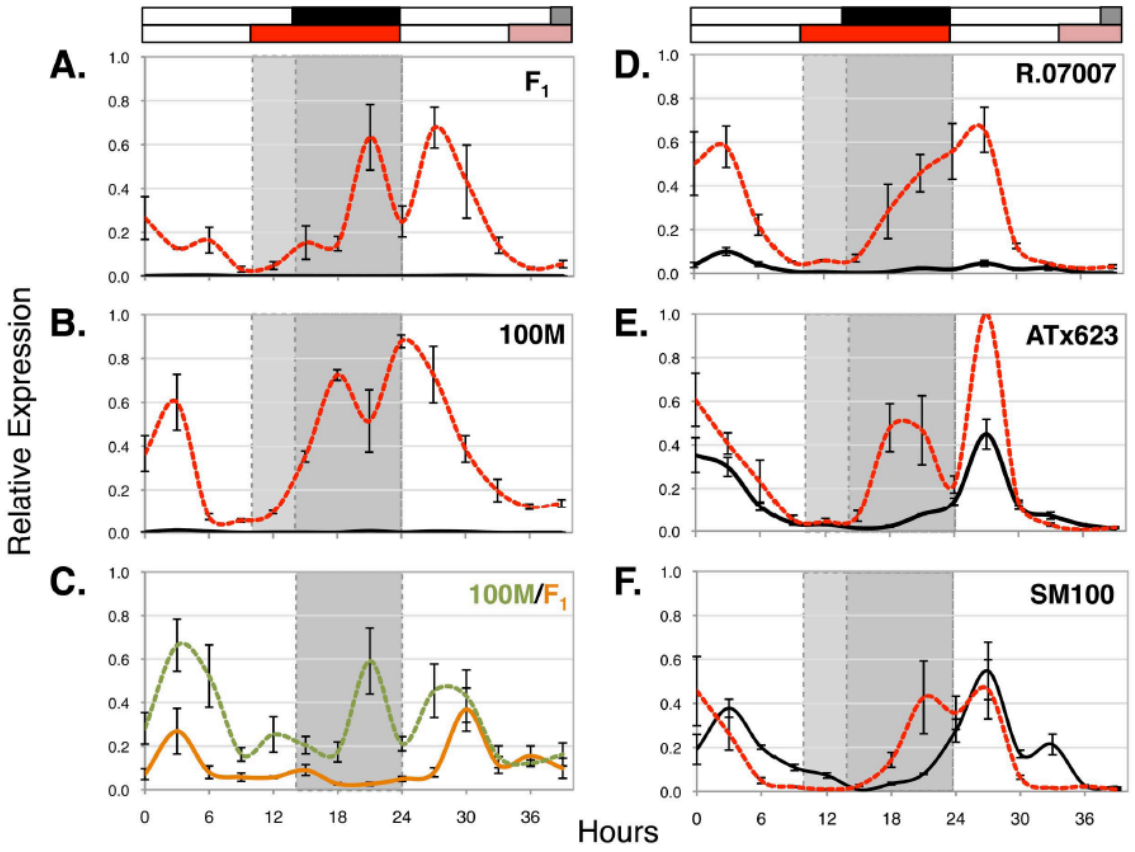

Fig. 40. *ZCN8* is partially repressed by Ghd7 in long days. Plants were treated under14-h light:10-h dark (LD, solid line) or 10-hlight:14-h dark (SD, red dashed line) conditions. Relative expression of *Ehd1* was analyzed at 3-h intervals by quantitative RT-PCR. ZCN8 is repressed in LD in the ATx623 x R.07007 F_1 (*Ghd7-1*) and 100M (*ghd7-1*) (*A* and *B*). However ZCN8 is repressed more strongly in the F_1 (orange line) than in 100M (green dashed line) (*C*), though not to the extent of *Ehd1* (Fig. 39), a gene predicted to be upstream of *FT*-like genes. (*D*) *ZCN8* is repressed in LD in R.07007 (*Ghd7-1*), though not in ATx623 and SM100 (*ghd7-1*) (*E* and *F*). The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar above the plot indicates the dark period for LD-treated plants; gray bars indicate subjective dark during LL conditions. Red bars indicate darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. Light-gray shading within the plot area indicates darkness for SD treated plants only; dark-gray shading indicates darkness for both LD- and SD-treated plants.

transducing input from multiple pathways into one signal that acts in the meristem. Therefore, it is likely that *ZCN8* is transcriptionally regulated not only by the Ghd7-photoperiod dependent pathway, but also by a parallel PRR37/CO-dependent mechanism in addition to the hormone, temperature, and nutrient signaling.

The *Ghd7*-dominant variety R.07007 exhibits repression patterns in LD and SD that mirror those observed for *Ehd1* (Fig. 39*D*). Like *Ehd1*, *ZCN8* is repressed in LD relative to SD, but this difference is not as great as that observed for the F_1 or 100M (Fig. 40*D*). In ATx623 and SM100, *ZCN8* is expressed at similar levels in LD and SD (Fig. 40, *E* and *F*), indicating that the expression of this gene is almost completely de-repressed in genotypes deficient in both PRR37 and Ghd7.

In maize, a second candidate for florigen was identified as *ZCN12* because of its response to photoperiod (173). However, in this species, it was expressed after floral initiation and was therefore thought to participate in floral development. By contrast, in sorghum this gene is expressed in response to short days prior to floral initiation, and with patterns similar to those seen for *ZCN8*. In the F_1 and 100M, this gene is strongly repressed in LD relative to SD (Fig. 41, *A* and *B*), and though the extent of repression in LD-treated F_1 plants is somewhat lower than that of 100M under the same conditions (Fig. 41*C*), this difference is much smaller than that observed for *Ehd1* (Fig. 39*C*). In R.07007, though *ZCN12* is repressed, the expression in LD is higher relative to SD than in either 100M or the F_1 (Fig. 41*D*). Additionally, similar to *ZCN8*, the expression of

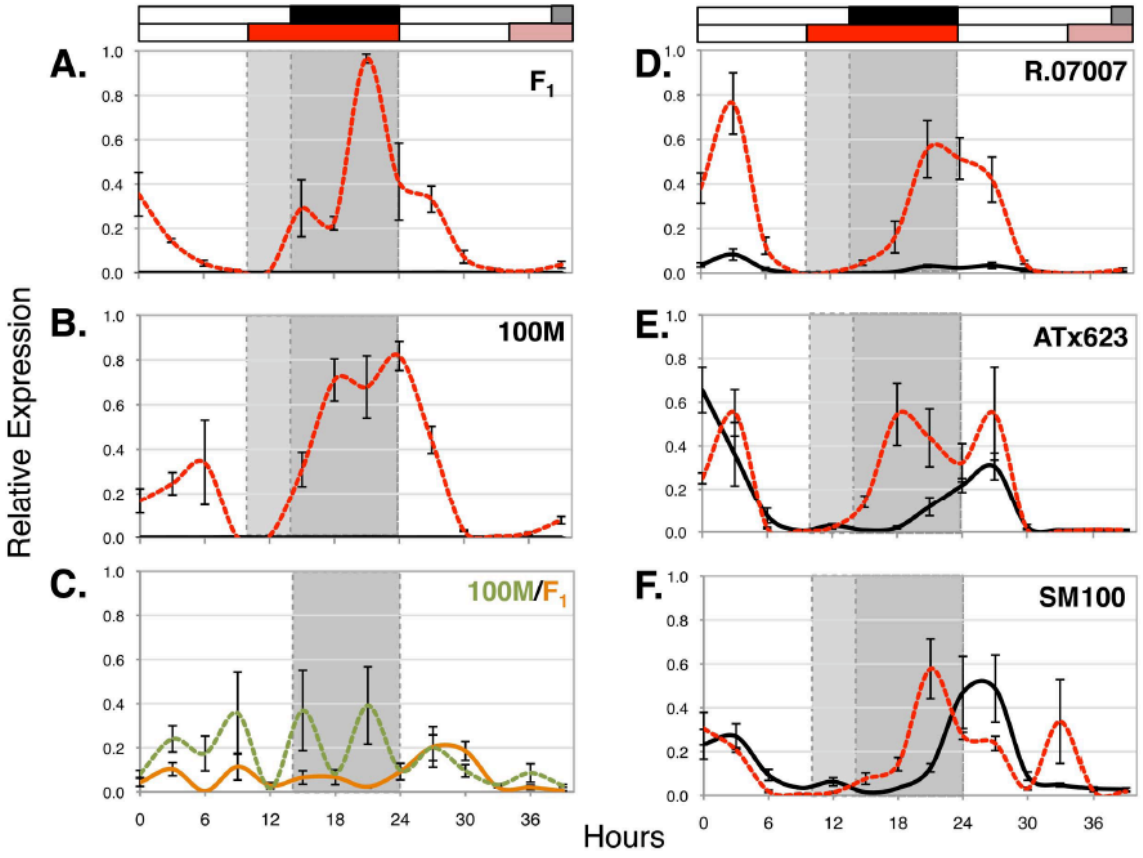

Fig. 41. *ZCN12* is minimally repressed by GHD7. Plants were treated under14-h light:10-h dark (LD, solid line) or 10-hlight:14-h dark (SD, red dashed line) conditions. Relative expression of *ZCN12* was analyzed at 3-h intervals by quantitative RT-PCR. In ATx623 x R.07007 F_1 and 100M (*A* and *B*), *ZCN12* is strongly repressed in LD. *ZCN12* is repressed to a somewhat greater extent in LD-treated F_1 (*Ghd7-1*, *PRR37*, orange line) than in 100M (*ghd7-1*, *PRR37*, green dashed line) (*C*). (*D*) LD-treated R.07007 plants (*Ghd7-1*, *PRR37*, ma_2ma_5) display some repression in LD, while *ZCN12* is de-repressed to near-SD levels in ATx623 and SM100 (*ghd7-1*, *prr37*) (*E* and *F*). The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar above the plot indicates the dark period for LD-treated plants; gray bars indicate subjective dark during LL conditions. Red bars indicate darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. Light-gray shading within the plot area indicates darkness for SD treated plants only; dark-gray shading indicates darkness for both LD- and SD-treated plants.

ZCN12 in LD in *prr37ghd7* genotypes ATx623 and SM100 approaches levels observed in SD (Fig. 41, *E* and *F*), indicating de-repression under either photoperiod. The expression difference in expression patterns between 100M (*ghd7-1*) and SM100 (*prr37-1*, *ghd7-1*) in LD indicate that the repression of this gene occurs in response to SbPRR37. Additionally, the limited difference in LD-induced repression observed between the F_1 (*Ghd7*) and 100m (*ghd7*) suggest that Ghd7 plays a minimal role in repressing this gene.

In rice, *Hd3a* is a major florigen signal in SD (157, 162). Therefore, its expression was analyzed in sorghum in genotypes with dominant and recessive *Ghd7* alleles. In both the ATx623 x R.07007 $F_1$and 100M, *GHD7-1* and *ghd7-1* types, respectively, relative transcript abundance is higher in SD than in LD (Fig. 42, *A* and *B*), however; it is not repressed in LD to the extent observed for *Ehd1*, *ZCN8* or *ZCN12*. The extent to which *Hd3a* is repressed in LD is similar in both 100M and the F_1, again suggesting that Ghd7 plays a limited role in the repression of this gene (Fig. 42*C*). In R.07007, which contains a dominant but misexpressed allele of *Ghd7*, expression in SD is higher relative to LD (Fig. 42*D*). In *prr37ghd7* genotypes ATx623 and SM100, the expression levels of *Hd3a* in LD and SD are very similar, indicating some degree of de-repression in LD (Fig. 42, *E* and *F*). The comparison of *Hd3a* expression in *Ghd7* dominant, *ghd7* recessive, and *prr37ghd7* recessive backgrounds suggest that PRR37 and Ghd7 both act to inhibit its expression, though not necessarily to the extent seen in the other floral activating genes. Moreover, the small difference in relative transcript abundance of *Hd3a* in LD-treated F_1 and 100M suggests that PRR37 may exert more of

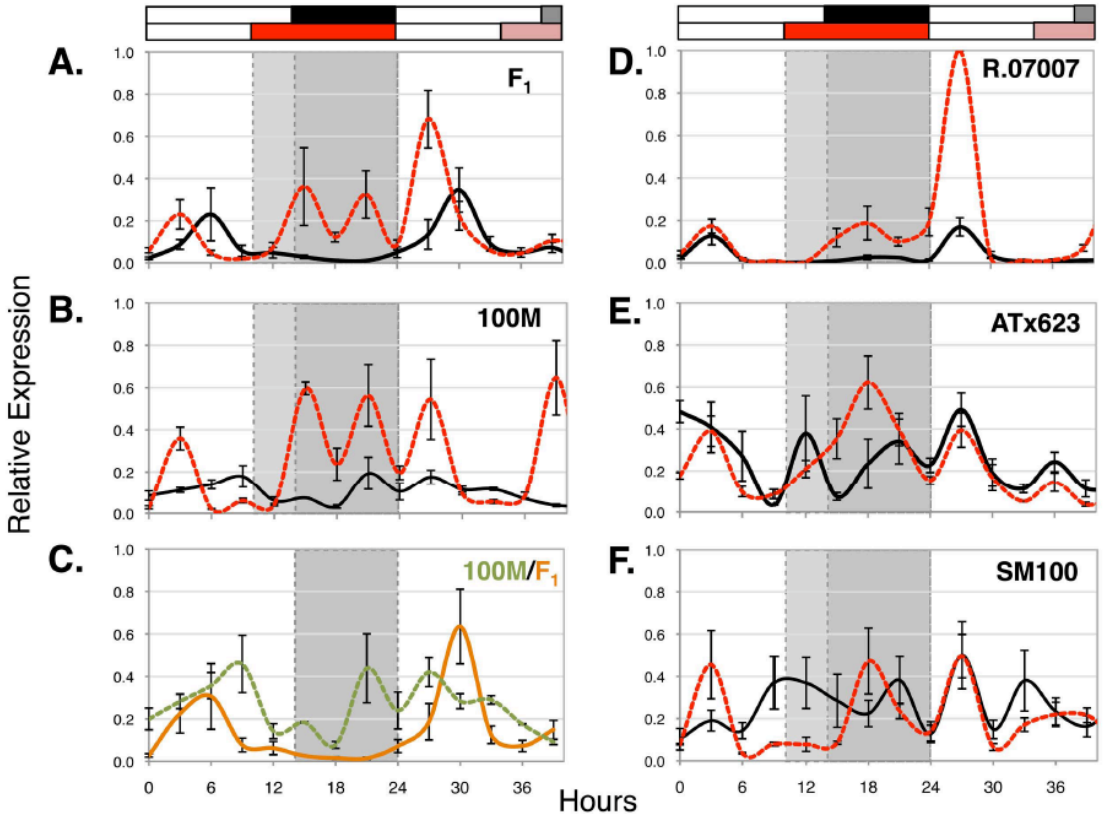

Fig. 42. *Hd3a* is partially repressed by GHD7. Plants were treated under14-h light:10-h dark (LD, solid line) or 10-hlight:14-h dark (SD, red dashed line) conditions. Relative expression of *Hd3a* was analyzed at 3-h intervals by quantitative RT-PCR. In ATx623 x R.07007 F_1 and 100M (*A* and *B*), *Hd3a* is relatively low in LD. The expression of this gene is somewhat more repressed in the F_1 (*Ghd7-1*, *PRR37*, orange line) than in 100M (*ghd7-1*, *PRR37*, green dashed line) (*C*). (*D*) LD-treated R.07007 plants (*Ghd7-1*, *PRR37*, ma_2ma_5) display repression in LD at certain points, while *Hd3a* is de-repressed to near-SD levels in ATx623 and SM100 (*ghd7-1*, *prr37*) (*E* and *F*). The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample and is based on three biological replicates ± SEM (189). The black bar above the plot indicates the dark period for LD-treated plants; gray bars indicate subjective dark during LL conditions. Red bars indicate darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. Light-gray shading within the plot area indicates darkness for SD treated plants only; dark-gray shading indicates darkness for both LD- and SD-treated plants.

an effect on this gene than Ghd7. Additionally, the molecular mechanism underlying the major peak observed in R.07007 under SD is not known (Fig. 42*D*, 27 hours). However, this genetic background contains other recessive maturity genes that likely have an effect on both *Ghd7* and *PRR37* expression, and potentially could have an effect on *Hd3a* via a mechanism independent of these two repressors.

Ghd7 and PRR37 Proteins Do Not Interact *In Vitro*

During map-based cloning of Ma_1 and Ma_6 using the ATx623 x R.07007 BC_1F_1, it was noted that these loci were both required for late flowering in the offspring of this cross. Moreover, the expression patterns of *PRR37* and *Ghd7* mRNA are very similar, with morning and afternoon peaks that occur in response to photoperiod. Sequence analysis of these genes revealed that they both encode proteins with a CCT domain, which have been shown to facilitate interactions with other CCT domain proteins during the formation heterotrimeric complexes consisting of CO and HAP3 and 5 proteins in the context of flowering. Therefore, it could be hypothesized that PRR37 and GHD7 proteins may interact to elicit an inhibitory response in LD. In order to explore this possibility, an *in vitro* co-immunoprecipitation was performed with recombinant PRR37 and Ghd7. PRR37 and T7-tagged Ghd7 were expressed separately *in vitro* using a TnT® T7 Coupled Reticulocyte Lysate System (Promega) in the presence of ^{35}S-Methionine and were then combined in order to allow an interaction. Additionally, known interactors TEN1 and T7-tagged STN1 were provided by the Shippen Lab and used as positive controls. A pull-down of tagged Ghd7 was subsequently performed

using T7 Tag antibody agarose beads. Aliquots from supernatant and bead fractions were run on a 10% acrylamide gel, and the results were viewed using a phosphorimager.

A strong signal associated with T7-tagged Ghd7 was detected in the bead fraction following the pull down (Fig. 43, Lane 7), however; PRR37 appears only in the supernatant fraction (Fig. 43, Lane 6). This result clearly shows that PRR37 and Ghd7

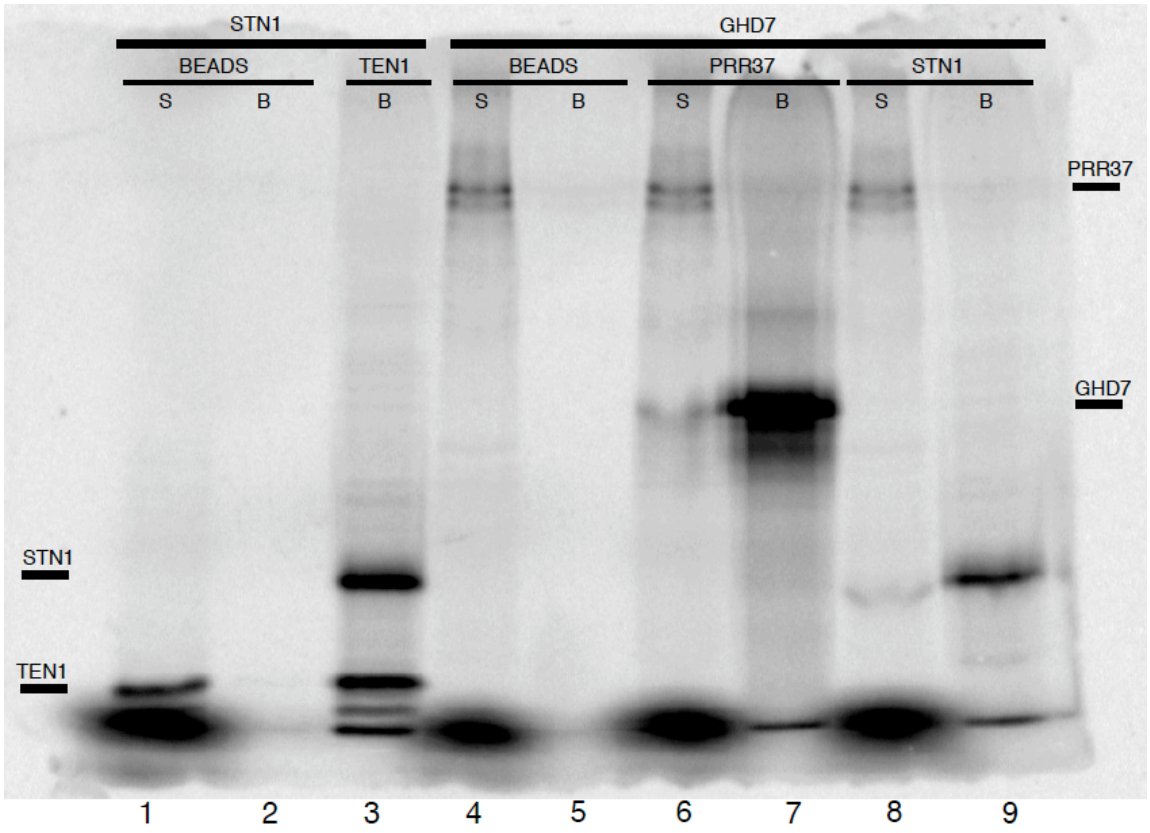

Fig. 43. Ghd7 and PRR37 proteins do not interact *in vitro*. Known interacting pair STN1 and TEN1 shows specific binding. The absence of nonspecific interactions in this pair is demonstrated by the presence of TEN1 in the supernatant fraction in the absence of STN1 (Lanes 1 and 2), and its association with the bead fraction following a pull down of tagged STN1 (Lane 3). PRR37 does not interact nonspecifically with the beads (Lanes 4 and 5) or STN1 (Lanes 8 and 9). However, no interaction with Ghd7 was detected, as demonstrated by the lack of association of PRR37 with the bead fraction following the pull down of tagged Ghd7 (Lanes 6 and 7).

do not interact, at least under these conditions, though it can not be assumed that these proteins do not interact *in vivo*. Because the CCT domain proteins often act in complexes, it may be that this pair requires other components to facilitate and interaction. However, the effects of PRR37 and Ghd7 on the transcriptional regulation of downstream floral activators can be separated to a certain extent in the ATx623 x R.07007 F_1 and 100M (*ghd7-1*), which suggests that these proteins may act separately via distinct pathways that converge on a similar set of flowering genes.

Discussion

It is unclear at what point a mutation arose in Ma_6 during the process of sorghum selection and improvement that gave rise to earlier flowering varieties. However, it is most probable that this variant occurred at a very early stage, as almost all of the historically grown sorghum Milo and Kafir varieties possess an identical *ghd7-1* (or nearly-identical *ghd7-2*) allele. Moreover, it is likely that the *ghd7-3* allele in U.S. germplasm originated in caudatum lines prior to their import into the U.S. Regardless of their origins, these mutations are associated with reduced photoperiod sensitivity and in many cases, recessive alleles of this gene were selected along with Ma_1 throughout the course of the conversion process. Both of these genes lie on chromosome six in a region of relatively low recombination, and haplotype analysis revealed that in many converted varieties, a significant proportion of donor genotype (BTx406) remains in this region of the chromosome (21). Therefore, conversion of the Ma_1 locus likely resulted in the introgression of a major portion of the arm containing both loci.

The expression of *Ghd7* mirrors that of *PRR37*, exhibiting differential expression in LD and SD photoperiods. The afternoon peak of *Ghd7* occurs in the window between the onset of darkness in SD (10 hours) and LD (15 hours), and the disruption of this increase in expression results in reduced photoperiod sensitivity even in a *GHD7-1* background. Moreover, this evening peak is light responsive, though its expression can only be induced during certain times of day set by the circadian clock. The combined input from light and the clock is the most basic principle underlying the external coincident model, and the expression patterns of *Ghd7* suggest a mechanism by which a pathway found in tropical short-day grasses may function according to the model proposed for the long-day plant *Arabidopsis*, but via a separate set of regulatory proteins.

In sorghum varieties with a full set of dominant loci, the expression of *Ehd1* is repressed to near-undetectable levels. By contrast, the repression of *Ehd1* in those types lacking Ghd7 does not occur to the same extent. This indicates that Ghd7 plays a major role in the repression of this gene, consistent with what has been reported in rice (164). The reduction in *Ehd1* is correlated with the reduced expression of *ZCN8*, and to a lesser degree, *ZCN12* and *Hd3a*. *ZCN8* acts as a florigen in maize (173), and as such should receive signals from multiple input pathways; of these Ehd1 represents a component of the photoperiod response. The observation that *ZCN12* and *Hd3a* are repressed in LD, but to a similar extent in *GHD7-1* and *ghd7-1* backgrounds indicates that these are repressed to a greater extent by PRR37 than by Ghd7, likely through the modulation of *CO* expression.

Though Ghd7 and PRR37 are both required to elicit extremely late flowering and are expressed in very similar patterns, preliminary biochemical experiments indicate that these proteins do not interact *in vitro*. Moreover, the Ghd7 protein contains a CCT domain, also found in CO and PRR37. CO and PRR37 have been shown to interact with NF-YB and C proteins (HAP3 and 5) in the context of flowering (53, 132, 133). Moreover, a recently-identified flowering QTL in rice was shown to encode a member of the NF-Y family (160). It is clear that the interactions of CCT domain proteins and NF-Ys make up an important component of the flowering response. However, it is not known how Ghd7 participates in this mechanism. It can be hypothesized that Ghd7 may also interact with these factors, allowing it to repress transcription of *Ehd1* or other floral activators (Fig. 44). Finally, the identification of this locus will not only enhance the photoperiodic flowering time model in the SD plant sorghum, but will also accelerate the selection of sorghum for hybrid production using the $Ma_1/Ma_5/Ma_6$ system to generate biofuels feedstock, in addition to enhancing the established practices for grain, sweet, and forage sorghum breeding.

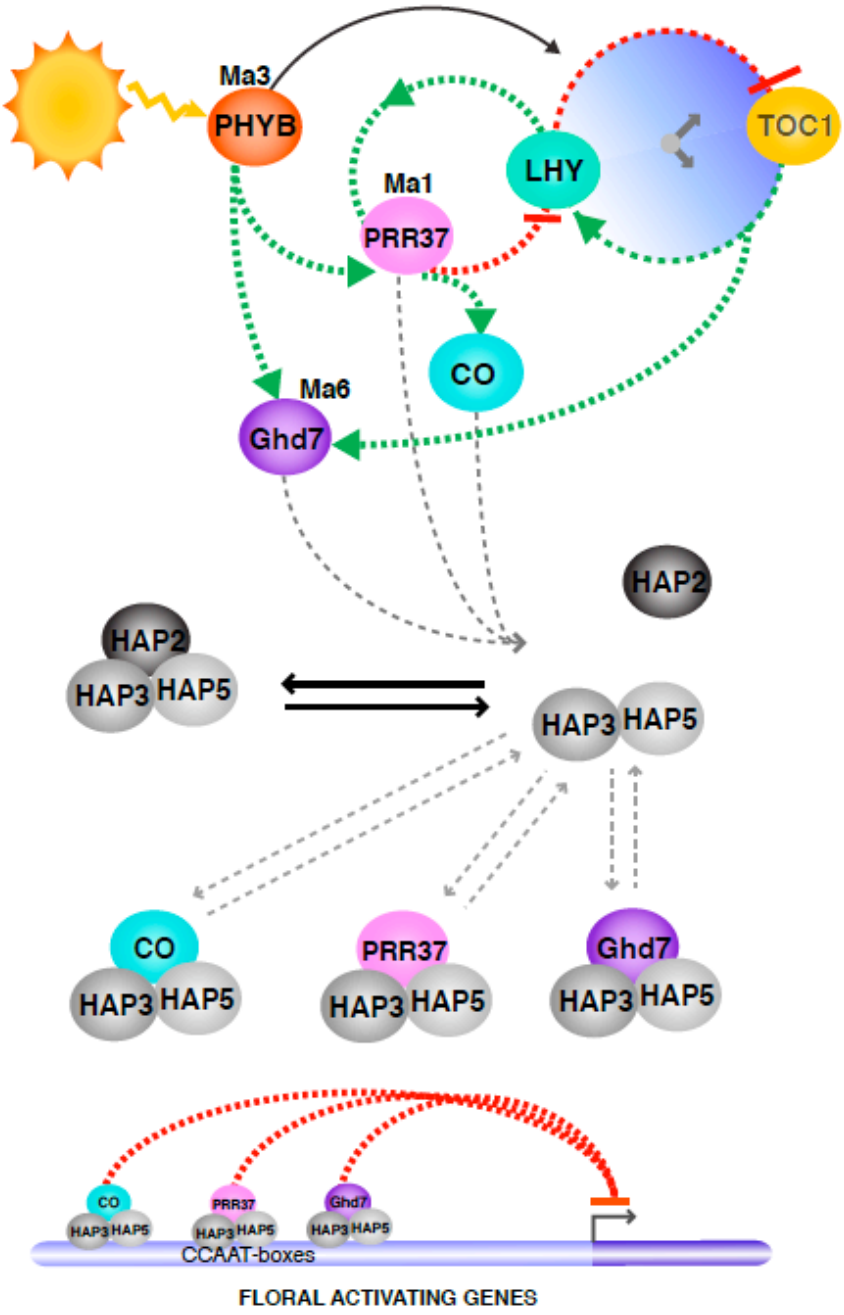

Fig. 44. CO, PRR37, and Ghd7 may interact in HAP complexes. Clock- and light-regulated expression of CO, PRR37, and Ghd7 confers differential expression in LD conditions. Subsequently, each of these proteins may compete with HAP2 (NF-YA) for binding in a trimeric complex with HAPs 3 and 5 and (NF-YB and C). These interactions may facilitate downstream repression of floral activating genes, like Ehd1.

MATERIALS AND METHODS

Plant Materials

The ATx623 x R.07007 BC_1F_1 population (n=1821) was grown under field conditions in the summer of 2003 in College Station, Vega, and Plainview, TX (USA). For flowering date determinations, days to mid-anthesis (pollen shed) were evaluated. BTx623 F_2 plants (n=124) were grown under long days in greenhouse conditions in 2008 and phenotyped for days to flowering (anthesis). Tissue was collected from individuals for genotyping, and seed were collected for planting. Following F_2 family genotyping, selected F_3 progeny (n=255) from that population were grown under the same long day greenhouse conditions in 2010, and scored for flowering at anthesis. Tissue was collected from this generation as well, and 185 combined progeny from both populations were used to in downstream QTL mapping. Preliminary biomass data was collected as dry weight/unit area of plants grown under field conditions.

Map-based Cloning of *Ghd7*

Genomic DNA was isolated from each line using the FastDNA Spin Kit (MP Biomedicals, Solon, OH, USA) and quantified using the NanoDrop1000 instrument. Novel marker discovery was performed using high-throughput genotyping of parental lines and F_2/F_3 offspring using the Illumina sequencer by the method described by Morishige *et al.*, in preparation. Template for sequencing was prepared using 500ng of starting genomic DNA, which was first digested with the restriction enzyme FseI (New

England BioLabs, Inc.), an eight-base cutter with methylation sensitivity. Adapters containing short identification sequences were ligated on to the FseI-cut end of the resulting fragments. Once the id tags were in place, the samples were pooled in groups of 48, precipitated, and sheared to 200-600bp using a biorupter (Diagenode). After fragmentation, size selection of 150-250bp product was performed by electrophoresis through an agarose gel and subsequent gel extraction (QIAGEN). Following this step, the samples were first treated with Bst DNA Polymerase (New England BioLabs, Inc.) to repair the 5' overhang present in the adapter sequences, and then blunt-ended using a Quick Blunting Kit (New England BioLabs, Inc.) to repair any ragged ends remaining from shearing. Finally, A-tails were added to the sheared end using Klenow 3'-5' exonuclease (New England BioLabs, Inc.). A PCR purification (QIAGEN) step was performed following each of these enzyme steps. Once end repair and A-tailing had been performed, a second set of adapters were ligated on to the 3' end of the product with a 5' identification adapter and a 3' general adapter. Ligation reactions were cleaned up using AMPure XP Reagent (Beckman-Coulter) to remove and excess adapter. Subsequently, products were enriched by PCR amplification from the adapters using one standard and one biotinylated primer. The desired products were isolated using magnetic Dynabeads® M-280 Streptavidin beads (Invitrogen), and the captured double stranded DNA was denatured at 98°C to form single stranded products. One strand, biotinylated, remained attached to the beads and was discarded, while the other was carried forward through a second round of PCR (Phusion™ High-Fidelity DNA Polymerase, New England BioLabs, Inc.). The final PCR resulted in double-stranded

products that contained the end sequences necessary for bridge-amplification, and samples at 10nM concentration were submitted for sequencing.

Initial quality assessment and analysis of the raw data was performed by Dr. Patricia Klein. Only sequence tags represented at least three times that also aligned to the physical map were retained. Polymorphisms were detected by comparing the each tag between two parents, ultimately for analysis in individual offspring. A set of individuals and their score at each of the newly identified polymorphic sites resulted from this analysis. The data were formatted and a genetic map was constructed using the MapMaker3.0/Exp script using the Kosambi Mapping Function. Redundant markers (those with identical marker scores for all individuals, and thus separated by 0cM distance) were removed from the analysis to control the overall map size. The remaining markers were checked for proper ordering using the ORDER and RIPPLE statistical checks at a minimum LOD of 2.0. The order of these markers was also compared to the physical map. Any marker falling outside these criteria was removed from further analysis. The genetic map was then combined with phenotype data and was used as input in QTL Cartographer using Kosambi distances. The LOD threshold was set using permutations of the data at 1000 iterations. QTL above this threshold were considered statistically significant.

The locus was refined further in the BTx623 F_2/F_3 and ATx623 x R.07007 BC_1F_1 populations using polymorphisms developed from SSR markers (SSRIT (198)), CAPS markers, SNPS, and INDELS. SNP and INDEL markers were discovered by *de novo* sequencing of regions surrounding predicted open reading frames. PCR primers

flanking polymorphic sequences were designed using PrimerQuestSM (Integrated DNA Technologies, Inc). Products were amplified from genomic DNA and the resulting products were purified and used in sequencing reactions with Big Dye Terminator v3.1 (Applied Biosystems). Capillary sequencing was performed on the ABI 3130xl Genetic Analyzer, and sequence assembly and analysis was carried out using Sequencher® v4.8 (Gene Codes Corporation). All markers used for fine mapping of the Ma_6 region are listed with their physical coordinates in Table 15.

Sequencing *Ghd7* Alleles

Genomic DNA from each of the lines was extracted using the FastDNA Spin Kit (MP Biomedicals, Solon, OH, USA), and amplified using primers flanking the *Ghd7* region. For certain lines, total RNA was extracted and cDNA synthesis was carried out using a SuperScript™ III First-Strand Synthesis System primed with a 9:1 ratio of random hexamer/oligo dT mix(Invitrogen, Carlsbad, Ca, USA). *Ghd7* coding sequence was amplified from both cDNA and genomic DNA, and sequencing reactions were performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) to obtain full sequences and intron/exon junctions these lines. These reactions were precipitated and sent for capillary sequencing on the Applied Biosystems 3130xl Genetic Analyzer. The results were analyzed using Sequencher® v4.8 (Gene Codes Corporation).

Table 15. Markers used for fine mapping the *Ma*$_6$ locus.

Marker	Forward	Reverse primer	SBI-06 location
CAPS1	TGGCATCTACGTGGAACTATTGGC	GCACAAGCTGACGTGTGGATTCAA	383,829
SSR8	TGGAAGGACTCCGATCTTTGTGGT	TGACGTCGAAATGACACCACTCCA	525,976
SNP1	TATATCACGATGACGGCTTGGGCA	TGGGTACGGAGACGAACAATCCAA	543,953
Sb06g000570 (IND1)	TCAGGACAACGATGACCACCAAGA	ATCAACCTCAAAGGTGAGCCCGTT	671,438
SNP2	TGGAGCCTGATCGCCCATTTAT	AAACTCGCCGCGGGAAACTAAA	771,940
SNP8	AGATTTAGCACTATGGGCCAACCC	TTTGTTGCCTGTCTATGTGGCAGC	792,009
SNP9	CGTCATCCTTCTGTTCCGGTACGATT	GCAGTGCATCTTGTCTTTGTCCCA	793,366
SSR5	TCCTGGCAGCAATACCGCATTAGA	TTTGGTCTGGCGTATAGATGGCGT	798,820
IND2	AAATGGATGGGTCCCTTGCCC	GAGAGATGGGAGACAGAAACAAAC	1,049,495

Sequences of forward and reverse primers are given, as well as polymorphism classification. Physical coordinates are given as obtained from Phytozome v7.0.

For the ghd7-3 allele, a protocol for flanking PCR was used to delimit the border of a repetitive sequence inserted into the intron (200). Genomic DNA (5µg) from control genotypes ATx623 (ghd7-1) and R07007 (GHD7), as well as from ghd7-3 genotypes Rio and Hegari, were digested using the blunt end cutting restriction enzymes PvuII, HincIII, DraI and SspI (New England BioLabs, Inc.). These enzymes were selected because they possess cut sites that are present in locations in the intron that would potentially result in polymorphic fragments between genotypes. The resulting fragments were purified using PCR purification columns (QIAGEN). Adapters (Table 16) were then ligated on to the end of the fragments with T4 DNA ligase (Promega), and

Table 16. Adapter and primer sequences for flanking PCR.

Adapter L	5' CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT 3'
Adapter S	3' NH_2-CCC GTC CA-PO_3 5'
AP1	5' GGATCCTAATACGACTCACTATAGGGC 3'
AP2	5' AATACGACTCACTATAGGGCTTCGAGCGGC 3'

All sequences are as previously published (200).

PCR was carried out using Phusion® High-Fidelity DNA Polymerase (New England BioLabs, Inc.) in a reaction with primers complementary to a known sequence in each exon in combination with primers complimentary to the adapter sequence. The resulting PCR products were gel purified (QIAGEN) and subjected to capillary sequencing using the gene specific primer and the adapter primer used for amplification. Sequences were analyzed using Sequencher v4.8.

Expression Analysis.

The expression of *Ghd7* was analyzed on tissue collected from 39 day old plants treated with LD or SD light conditions at three-hour intervals, as described in the Ma_1 study (Fig 18). 100M, SM100, Hegari, ATx623, R07007, and ATx623 x R07007 F_1 plants were grown under long day conditions (14hr days) in a greenhouse in Metro-Mix 200 soil (Sunshine MVP; Sun Gro Horticulture, Canada CM, Ltd.) and were fertilized after two weeks using Peters Professional Allrounder fertilizer (The Scotts Company LLC). After 32 days, the plants were transferred to a long (14 hr) or short (10 hr) day growth chamber for one week at a light intensity of ~300µmol s-1 m-2 at ~50% humidity with 30° C day temperatures and 23° C night temperatures. After being allowed to entrain for a week under a SD or LD photoperiod, three leaves from each of three individual plants of the same genotype were harvested and pooled at three hour intervals for one 24 hour light-dark cycle and two additional 48-hour LL (constant 30° C) or constant DD (constant 23° C).

Tissue homogenization was carried out in liquid nitrogen using a mortar and total RNA was extracted from each sample using the TRI REAGENT® Protocol modified for plant samples (high polysaccharide protocol; Molecular Research Center, Inc.). The RNeasy Mini Kit was used to purify the RNA, and on-column DNase digestion was performed to remove any remaining DNA contamination (QIAgen). Quantification of samples was performed in duplicate using a NanoDrop 1000 Instrument (Thermo Fisher Scientific, Inc), and the values were averaged. RNA integrity for all samples was assessed on a 1% MOPS gel visualized using the Molecular Imager® Gel Doc™ XR

running Quantity One®v4.6.8 software (Bio-Rad Laboratories, Inc.). First strand cDNA was synthesized using the SuperScript™ III First-Strand Synthesis System primed with a 9:1 ratio of random hexamer/oligo dT mix (Invitrogen, Carlsbad, Ca, USA). The reactions were diluted to 10ng/µl cDNA in TE buffer for downstream use in qRT-PCR reactions.

Primers for qRT-PCR-based expression analysis of flowering genes were designed using PrimerQuestSM (Integrated DNA Technologies, Inc) as described in the *Ma1* study. Primer efficiencies were determined by the construction of serial dilutions from purified cDNA-amplified PCR products of target genes over a range of 0.05 ng μl^{-1} to 5.0E^{-7} ng μl^{-1}, which were subsequently used in qRT-PCR reactions. The resulting standard curve was used to calculate the efficiencies for each primer pair for all genotypes(Applied Biosystems). The primers used for Ghd7 expression (Forward: TCAGGACAACGATGACCACCAAGA; Reverse: ATCAACCTCAAAGGTGAGCCCGTT) were 91-94% efficient; the efficiencies for clock and downstream flowering genes are reported in Table 5. Only primers whose efficiencies did not vary more than 10% between genotypes were used for qRT-PCR analysis. The absence of genomic DNA contamination was verified by running no-template control qRT-PCR reactions on 18S ribosomal RNA (Applied Biosystems) from each sample (10 ng μl^{-1} RNA). All reactions were performed on the 7900HT Fast Real-Time PCR System running SDSv2.3 software. Gene-specific reactions were performed using Power SYBR® Green PCRMaster Mix (Applied Biosystems). Control reactions were performed using TaqMan® Universal PCR Master Mix in reactions with 18S

ribosomal RNA and an rRNA Probe (VIC™ Probe), rRNA Forward Primer, and rRNA Reverse Primer (Applied Biosystems).

Raw Ct values were collected for each gene, and calibrated to 18S ribosomal RNA to obtain ΔCt values. Relative expression was calculated using the Comparative Ct (ΔΔCt) method (189) with the most highly expressed sample used as the calibration sample between both LD and SD samples. Mean values are based on three technical replicates and three biological replicates for both reference and target genes, ± SEM.

In Vitro Co-Immunoprecipitation

The coding sequences of *PRR37* and *GHD7-1* were amplified from cDNA using primers designed to introduce a 5' BamHI restriction site and a 3' XhoI or NotI site (Table 17). The resulting products were used in a double digest with BamHI-HF™

Table 17. Cloning primers for Ghd7 and PRR37 Co-immunoprecipitation.

	Forward	Reverse
Ghd7	BamHI	NotI
	GATATC**GGATCC**ATGTCAGGGCCAGCATGCGG	CTGCAT**GCGGCCGC**GTAGTCAAATTAACCAGTGCC
PRR37	BamHI	XhoI
	GGTCGC**GGATCC**ATGATGCTTCGGAATAACAAC	CATCTG**CTCGAG**CTAGAGGCGAAGGTGGAGTC

The enzyme used for each is indicated, and the restriction site for each is marked in red.

and either XhoI or NotI-HF™ (New England BioLabs, Inc.), and subsequently gel purified to remove digested primer fragments. Additionally, the vectors pET-28a(+) and

PCITE-4c(+) were digested using the same set of enzymes and gel purified using a Gel Extraction Kit (QIAGEN). The *Ghd7* coding region was ligated into pET-28a(+) (Novagen), which incorporates a T7 tag, and PRR37 was ligated into pCITE-4a(+) (Novagen). The orientation and integrity of the insert was verified by direct capillary sequencing of the plasmid, and the expression of protein was tested using a cell-free TnT® T7 Coupled Reticulocyte Lysate System (Promega) in conjunction with the FluoroTect™ GreenLys in vitro Translation Labeling System (Promega), which incorporates fluorescence into the protein for visualization via the Molecular Imager PharosFX™ System (Bio-Rad) (Fig. 45).

Once quality plasmids expressing each protein were obtained, an *in vitro* Co-Immunoprecipitation was performed to test for interaction. Both proteins, in addition to positive interacting controls STN1 (pET-28a(+)) and TEN1 (pCITE-4a(+)), were expressed separately in RRL in the presence of ^{35}S Methionine.

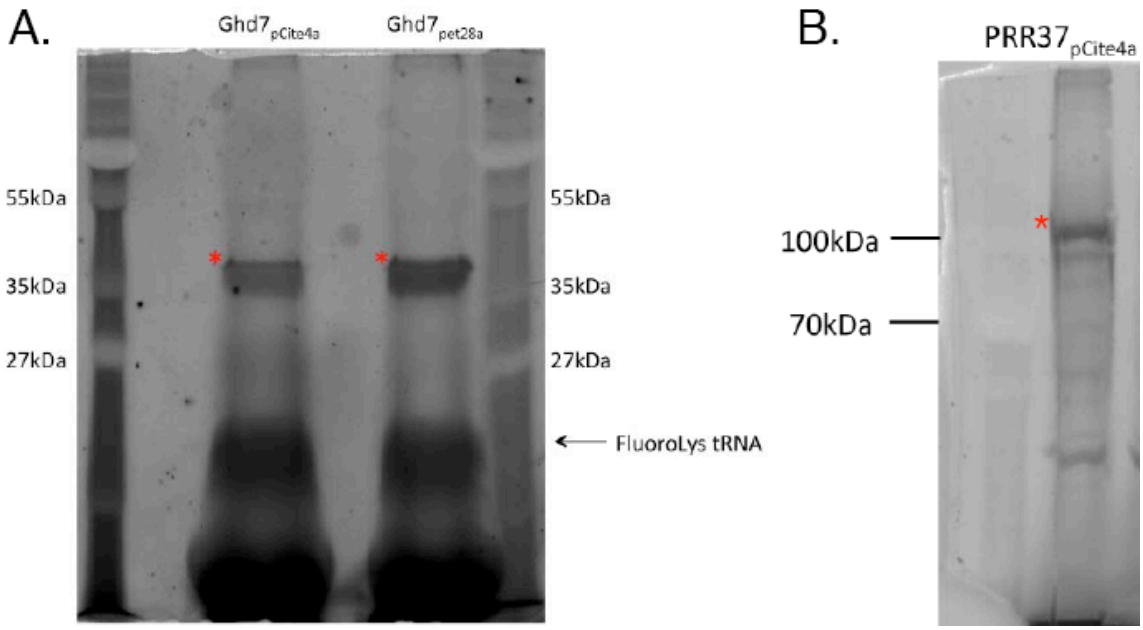

Fig. 45. Test expression of proteins for co-immunoprecipitation. (A) Ghd7 is expressed around 38kDa frompCITE4a and pET28a. (B) PRR37 is expressed around 100kDa from pCITE4a. Proteins were expressed in the presence of FluoroLys tRNA, and run on a 10% acrylamide gel. Expression and visualization of Ghd7 was performed by Andrew Nelson.

Expression was allowed to continue for 90 minutes, and then terminated using cyclohexamide. The protein pairs for which interactions were to be tested were combined into the same tube at a ratio of 3:1 pET-28a(+):pCITE-4a(+) expression product, and allowed to form any potential interactions. Following a two hour incubation, prepared T7•Tag® Antibody Agarose beads (Novagen®) were added to each potential interacting pair, and binding was carried out overnight. Following a series of wash steps, the beads were recovered by centrifugation. SDS loading dye was added to the bead-bound protein, and the samples were boiled briefly to release the protein from the beads. The resulting supernatant was loaded on a 10% acrylamide gel. After drying,

the gel was used to expose a storage phosphor screen overnight, and the screen was visualized using the Molecular Imager PharosFX™ System (Bio-Rad).

CHAPTER IV

CONCLUSIONS

SUMMARY

Initial studies of photoperiodic flowering in *Sorghum bicolor* began decades ago, and through the tenacity of a few geneticists, six agronomically important maturity loci were discovered that repress flowering under long day conditions. However, the identities of these loci had remained largely unknown until recently. The genetic and physiological effects of each locus were well-characterized in a set of maturity standards. However, it was not until 1995 that the position of Ma_1 was initially mapped to a location on chromosome six. Since the mapping of Ma_1, only one other maturity locus had been elucidated; Ma_3 encodes phytochrome B (24). In the following fifteen years, several generations of scientists have worked together to identify the genes underlying the remaining loci, with emphasis on Ma_1 as the most effective floral repressor of the six. In 2011, I identified this gene as *PSEUDORESPONSE REGULATOR 37* using a map based cloning method, and its role within the sorghum photoperiod pathway was characterized.

Several alleles of *PRR37* exist within the sorghum germplasm. Two recessive alleles, *prr37-1* and *prr37-3* are severe; resulting in a frameshift (*prr37-1*) and a premature stop. A third allele, *prr37-2*, contains a single amino acid substitution at a conserved lysine, and genotypes that possess this allele exhibit a reduced delay in flowering in long days relative to genotypes that contain *PRR37*. Other mutations of

varying strength have been found among diverse sorghum accessions (Klein *et al.*, in preparation), and it is likely that the strengths of each allele contribute to the differences in photoperiod sensitivity observed between sorghum varieties.

The expression of *PRR37* is regulated by light, but gated by the circadian clock so that it occurs in a bimodal manner; mRNA abundance peaks in the morning and the afternoon. The afternoon expression is dependent on light exposure, mediated via PHYB (Yang *et al*., in preparation), and is absent in short days, consistent with the *Arabidopsis* external coincidence model (121, 176). As a central repressor in this pathway, PRR37 inhibits flowering both by up-regulating the expression of floral repressors, and down-regulating the expression of floral activators. CO, a repressor in sorghum, exhibits expression patterns that mimic those observed for PRR37, and is expressed in a bimodal fashion with an initial peak that occurs during the night and a second peak that occurs at dawn. This second peak is associated with delayed flowering, and is absent in SD conditions in *PRR37* genotypes and in both LD and SD in *prr37* genotypes. Moreover, in *Arabidopsis*, *PRR7*, the ortholog of *SbPRR37*, is expressed as a single morning-phase peak (193), suggesting that the evening-specific peak may be a special feature of grass species.

Additionally, the expression of the floral activator genes *Ehd1*, *Hd3a*, *ZCN8*, and *ZCN12* is down-regulated in the presence of functional PRR37. In the case of *Hd3a*, *ZCN8*, and *ZCN12*, this repression may be direct, but probably also occurs through CO. Repression of *Ehd1* occurs partially through PRR37, contributing to the downstream

inhibitory effects on the other flowering genes, though the main effect on this gene is primarily due to Ghd7, which was identified as Ma_6.

In the sorghum breeding program, it was observed that offspring from certain crosses flowered much later than either of the parents. This effect could not be attributed to any of the original four maturity loci, and thus in 1999 Ma_5 and Ma_6 were reported (22). Ma_6 was shown to be required for the extreme photoperiod sensitivity, a desirable trait in sorghum intended as bioenergy feedstock, and while many commercial sorghum varieties were dominant at Ma_5, very few possessed dominant alleles of Ma_6. As interest in producing high biomass lines increased, the importance of identifying this gene grew accordingly. In 2012, I identified this locus using a map-based cloning approach as the sorghum ortholog of rice *Grain yield, heading date, and plant height 7* (*Ghd7*).

Multiple functional alleles of *Ghd7* (*GHD7-1* to *GHD7-6*) were identified in lines with increased photoperiod sensitivity, in addition to three nonfunctional alleles found in earlier flowering varieties. The primary alleles found in grain and maturity Milos as well as many historical kafirs were characterized by the insertion of five bases in the coding region upstream of the CCT domain (*ghd7-1* and *-2*). A third allele, *ghd7-3*, characterized by an insertion in the intron region that likely results in inefficient mRNA production, was present in several lines derived from the caudatum race.

Interbreeding between races and the conversion of sorghum in the U.S. at this locus in addition to Ma_1 has made it difficult to determine the origin of the nonfunctional alleles present in current sorghum breeding materials. Haplotype analysis of

chromosome six indicated that the portion of the exotic germplasm retained on this chromosome after conversion by BTx406 was not as high as predicted (21), and recessive alleles of ma_1 and ma_6 were likely introgressed together as a major portion of that chromosome during conversion. Moreover, it is likely that mutations in Ghd7 occurred prior to the import of most sorghum varieties into the United States, and could have potentially played a role even in the domestication and dispersal of sorghum away from the equatorial region in Africa. It is hypothesized that the presence of the Ghd7 system in addition to the CO pathway enables tropical grasses to sense the very small changes in day length that occur at the equator. Sorghum grown at higher latitudes exhibits extremely delayed flowering in response to several factors including photoperiod, and many of these varieties cannot be propagated with a full set of dominant maturity loci.

Ghd7 expression is regulated by the circadian clock in a light-dependent manner, similar to what was observed for *PRR37*. The differential expression of *Ghd7* in LD and SD is consistent with what is observed in rice, as well as with the external coincidence model. In LD, *Ghd7* expression occurs in two peaks; one at 3 hours, the other at 15 hours after initial light exposure. By contrast, in SD, only the morning peak is expressed, however; when SD-treated plants are exposed to constant light, this afternoon peak reappears during the first LL day, suggestive of a light requirement. This is reinforced by the observation that LD-treated plants show an immediate dampening of Ghd7 expression when transferred to constant dark. The afternoon peak of *Ghd7* occurs in the window between the onset of darkness in SD (10 hrs) and LD (15 hrs), and any

disruption in this peak is associated with reduced photoperiod sensitivity, as seen in R.07007, a line that misexpresses an otherwise functional (*GHD7-1*) allele. The dual input from light and the circadian clock lies at the heart of the external coincident model, and the regulation of *Ghd7* is consistent with this model.

Ghd7 acts as a strong repressor of *Ehd1* in rice (164). In accordance, when sorghum Ghd7 is dominant, *Ehd1* is repressed to near-undetectable levels. Though PRR37 can also reduce *Ehd1* expression, Ghd7 plays the largest role in repressing this gene. Moreover, decreased *Ehd1* expression translates downstream to reduced expression of florigen genes *ZCN8*, *ZCN12*, and *Hd3a*. The inhibitory effects imposed on these floral activators by Ghd7 is not as drastic as that observed for *Ehd1*, indicating that, while Ghd7 plays a role in the repression of these genes in LD, its action is only one part of a multi-component repression system, including PRR37 and CO, among others.

PRR37 and *Ghd7* are transcriptionally regulated in a similar manner, and both are required for extreme photosensitivity, though they do not affect the expression of each other. Moreover, the gene products of each encode CCT domain proteins, which are known to form heterotrimeric complexes with other CCT domain proteins (53, 132, 133), thus it was hypothesized that these two proteins may interact. This does not occur under *in vitro* conditions, but due to the heterotrimeric nature of CCT-domain/HAP complexes, it cannot be ruled out that this type of interaction does not occur *in vivo*. However, these preliminary results further suggest that PRR37 and Ghd7 act independently from each other to inhibit flowering.

The molecular mechanism of upstream regulation of PRR37 and Ghd7 is not well understood, though it is hypothesized that they are regulated independently of each other by the same regulatory pathway. Phytochrome B is involved in the light-induced expression of both genes, and in maturity standard 58M (ma_3^R), recessive at *PHYB*, no afternoon expression is observed under LD or SD conditions. It is thought that the other phytochromes are involved as well, though their role is not presently clear. Moreover, the remaining unidentified maturity loci (Ma_2, Ma_4, Ma_5), may play a role in maintaining the proper expression of PRR37 and Ghd7. R.07007, which is deficient at (ma_5), exhibits altered expression patterns of both of these genes, resulting in decreased photoperiod sensitivity. R.07007 is dominant for Ma_1 and Ma_6, therefore when all of the maturity loci are complemented in the ATx623 x R.07007 F_1, normal expression patterns and function of Ma6 are restored. Additionally, there may be an auto-regulatory component between the mRNA and the protein, because genotypes with recessive alleles of each show a shortened period, as observed for ATx623.

The results of these two studies provide a novel perspective into the mechanism of photoperiodic regulation of flowering time in the SD-grass sorghum. The specific bimodal expression patterns of light and clock regulation of these genes is uniquely observed in sorghum, while simultaneously fitting into the external coincidence model. Within the sorghum-specific model, the connection between the maturity loci, their biochemical function, and their relationship to each other is becoming clear. From this information, an initial framework for the sorghum flowering pathway can be assembled (Fig. 46). Furthermore, in these studies, agronomically important alleles of PRR37,

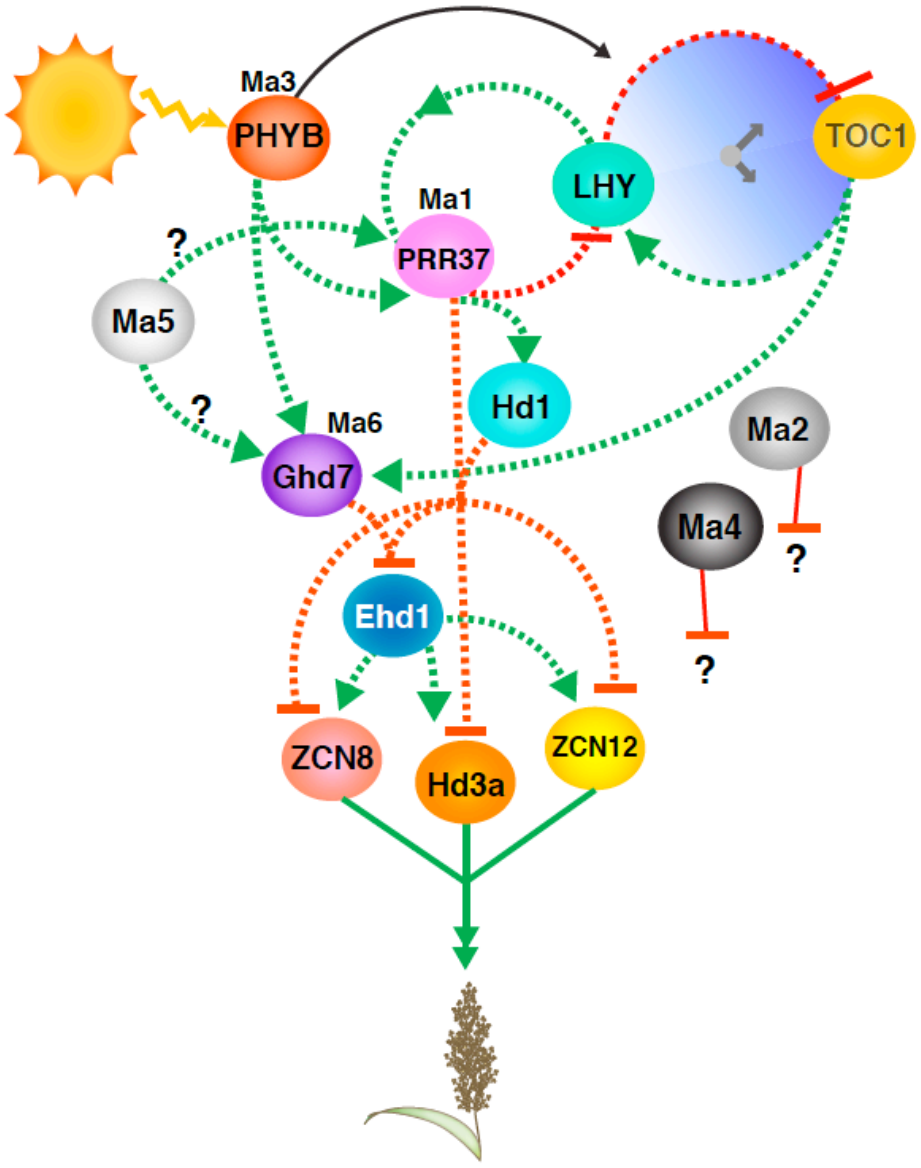

Fig. 46. The sorghum flowering pathway. The expression of major floral repressors PRR37 (*Ma*$_1$) and Ghd7 (*Ma*$_6$) is controlled by the clock and light via phytochrome B (*Ma*$_3$), in addition to potential regulation by other maturity loci. Both of these genes repress *Ehd1*, and the florigens *Hd3a*, *ZCN8*, and *ZCN 12*, though PRR37 works through *CO* while Ghd7 works primarily through *Ehd1*. Dotted lines denote transcriptional activation (green) or repression (red). Solid lines in the same color scheme represent general inductive (green) or repressive (red) input. Question marks represent predicted action.

encoded by a maturity locus critical for the domestication and utilization of sorghum for grain production in America and temperate regions worldwide, was identified (1, 179). Additionally, Ghd7, a gene historically recessive in the majority of grain sorghum, now essential for the production of late-flowering bioenergy lines, was identified as Ma_6. Therefore, in addition to major contributions to the model of photoperiodic flowering in the short day plant sorghum, this information will accelerate breeding efforts for sorghum improvement by providing a way for plant breeders to more precisely control flowering.

FUTURE DIRECTIONS

The understanding of flowering has grown substantially in *Arabidopsis* in the past several decades and many of the factors that control this process have been identified and well-characterized in rice as well. However, many more questions remain. In sorghum specifically, it is clear that the identities of the remaining maturity loci should be uncovered, and the connections between the underlying genes clarified. Additionally, though factors have been reported that repress flowering in LD, more and more evidence suggests that regulation of flowering in LD and SD may require distinct pathways, and identifying factors that control flowering under SD should be a priority in sorghum. However, in addition to these questions, there are other more general phenomena that can be can be investigated to the benefit of flowering research across multiple species.

Identifying the molecular basis of the difference between long day and short day plants has been an overarching theme in flowering research. The answer to this question can probably not be answered with a single "magic bullet," but instead is likely to have multiple parts. One of the key differences between *Arabidopsis*, long day temperate grasses, and short day tropical grass species is the existence of the *Ghd7-Ehd2-Ehd1* pathway. However, it is not known whether this pathway or components thereof exist widely in short day plants other than rice, maize, and sorghum, which are fairly well-characterized with respect to photoperiod sensitivity and closely related. Therefore, it would be interesting to determine where the Ghd7 pathway arose in the evolution of the grasses with respect to each species' photoperiod sensitivity and geographical location. An initial step towards this goal would be to probe the genomes of many more temperate and tropical grasses for the presence of *Ghd7*, *Ehd1*, and *Ehd2*. One approach to this would be to select a set of representatives from each of the twelve recognized subfamilies within the main *Poaceae* family, including such species as oat, bamboo, sugarcane, wild rice, and many others. Eventually, it would be interesting to expand this analysis to tropical SD plants outside of this family. These species may not possess *Ghd7* specifically, but may have some other mechanism for the repression of flowering under LD. Understanding the sources of variation in photoperiodic response would add valuable information to the current model of flowering time control in multiple species. Moreover, with the decreasing cost and time involvement of next generation sequencing it will become possible to obtain entire genome sequences for direct comparison of all of the genes involved floral initiation.

A second point of LD and SD photoperiod control occurs at CO, the main regulator of FT in *Arabidopsis*. In this LD species, CO acts as an activator when days are sufficiently long, binding directly to the *FT* promoter through interactions with certain members of the HAP3/5 family. In rice, however, the activating function of Hd1 (CO) is limited to SD conditions. In LD, differential phosphorylation of the Hd1 protein results in the repression of Hd3a (FT). The molecular basis of the phosphorylation-dependent functional switch is not understood, however it is possible that it could affect which HAP proteins CO binds. It is not known whether certain HAP proteins are intrinsically repressors or activators, although it has been shown that in wheat, CO competes with NF-YA (HAP2) for a place in the heterotrimeric complex with NF-YB and C (HAP3/5). If upon modification, the rice Hd1 protein interacts with a different set of HAP3/5 proteins, this could modulate Hd1's function. Therefore, an important next step would be to determine first, which HAP proteins serve as interacting partners for CO under SD activating conditions. This has not been determined in rice, and therefore would advance the knowledge base for this species. Second, it would be important to characterize the binding of Hd1 to HAP proteins in LD. These experiments could eventually be coupled with others that determine the target promoters of the complexes with respect to downstream flowering signals, and the transcriptional responses of these genes in order to verify the functional significance of any difference in complex formation and downstream effects on flowering.

Upstream of CO is its activator GI, which is a very interesting protein that plays many versatile roles in both flowering and the circadian clock. This protein mediates

several blue light responses by interacting with blue light photoreceptors FKF1 and ZTL, which play important roles in the stability of flowering and clock related proteins. Additionally, it has been shown that both blue light and GI have input into the miRNA172 pathway, and although it has not been shown, it is possible that GI mediates this blue light response. Moreover, GI is predicted to form a circadian clock "gate" around the expression of *Ghd7* and *Ehd1*, regulating the time of day during which each gene is receptive to inductive red and blue light signals, respectively (166). This gating is thought to occur not by GI directly, but by certain unidentified downstream components. The phenomenon of circadian gating has been recognized for a relatively long while, though its molecular basis is still quite enigmatic. Because of the tight association between the clock and photoperiodic flowering, and because several important flowering genes are regulated in this way, understanding this process would have broader implications for flowering time as well. Moreover, with respect to sorghum, the expression of both *Ghd7* and *PRR37* are gated by the clock, so that the response to light only occurs at specific times of day.

One initial step would be to identify factors downstream of GI that may participate in this gating. GI could control other factors in a variety of ways, though it commonly works as a molecular "glue" that reinforces the interactions of F-box proteins and the SCF. Therefore, screening for additional GI interacting partners under blue or red light may reveal whether *Ehd1* and *Ghd7* expression is gated by GI-mediated protein degradation. Similar experiments have been performed in *Arabidopsis* using epitope-tagged GI and candidate proteins that exhibit similar expression patterns to GI (201),

though not under different light conditions. Additionally, screening for genes with differential expression in the *Osgi-1* (166) mutant may also indicate which factors may be playing a role in the formation of gates controlling gene expression, in addition to genes that are subject to GI-dependent regulation.

With the advent of high throughput techniques and the "–omics" era, one important question arises: How should all of this data be handled? While genome, transcriptome, proteome, and other such projects have obvious merits and represent substantial achievements, the standard methods for organizing and presenting the resulting data have to present been unwieldy, and perhaps even intimidating for those interested in gleaning only a small "needle" of information from the general haystack. However, one emerging area for which this type of information may be particularly suited is *in silico* biological modeling. Modeling of gene regulatory networks requires information about temporospatial transcriptional control and the break down of *trans*-acting gene products, the dynamics of protein-protein interactions, and other molecular data (202). High-throughput techniques provide a way to collect this kind of information for many components all at once, greatly accelerating one of the main challenges of biological modeling: collecting enough information with sufficient detail to explain a phenotypic phenomenon. New developments in the technology behind this modeling have also led to improvements in methods for establishing statistical constraints so that more realistic models can be generated automatically from simplified inputs of complex networks.

The complexity of the floral regulatory network in *Arabidopsis* presents an excellent target for this type of analysis, and in fact, certain aspects of floral organ development have already begun to be studied in this way (203, 204). Moreover, the generation of *in silico* representations of gene networks has allowed a connection to be made from genotype to phenotype in such a way that certain traits can be predicted from a novel genotype. The modeling of leaf elongation rates in maize and traits associated with the stay-green trait (drought response) in sorghum are already being put to use (205). Because maturation time is extremely important in obtaining high yields in sorghum, modeling the process of floral initiation could potentially have a very real impact on molecular breeding efforts and crop improvement.

Much is known about the regulation of flowering in rice and maize, and as improved genomics techniques accelerate the identification of important floral regulators in sorghum, its pathway is rapidly approaching the point at which modeling can be a practical next step. One important piece of information is still missing, however. In order to accurately predict flowering time *in silico*, the nature of sorghum florigen must be determined. Characterization of sorghum florigen can be performed in a few parts.

First, one must know how many components make up florigen. In maize, a family of 26 genes have been identified bioinformatically as having similarity to FT or FT-like proteins, and these were eventually narrowed down into one or two candidates by analyzing the timing and distribution of each gene's expression (34, 173). A similar analysis should be performed in sorghum using RT-PCR techniques in a variety of tissues and developmental time points in order to determine to which particular genes

effort should be focused. All known florigen molecules are expressed in the leaves in response to inductive conditions, and travel to the meristem where they initiate a phase change. Thus, candidates that act as florigen should be expected to be expressed in leaves in response to short day treatment. Second, the level of florigen required to initiate flowering should be determined. One way to approach this is to take advantage of the fact that the border between what a plant perceives as a "long day" and a "short day" is very thin (166). Generally speaking, a difference in as little as 20 minutes can determine whether or not a plant flowers in response to photoperiod. In sorghum, this precise day switch in day length has already been studied to some extent (206). This could be characterized further by growing plants in incrementally shorter photoperiods and quantifying the levels of florigen protein at each of these. When the border between inductive and repressive photoperiods is determined, it is likely that this level of florigen marks the tipping point between vegetative and reproductive growth in response to photoperiod. Additionally, it would be important to verify that these florigen proteins bind similar promoter sequences to what was observed in rice and *Arabidopsis*. This would be essential for forming the proper *in silico* connections between genes. Though several specific components would have to be elucidated, modeling of the flowering phenomenon is feasible and could greatly contribute to molecular breeding efforts. Predicting the flowering time of a given genotype in a given environment would allow breeders to select promising varieties from a very early stage, allowing for much more rapid improvement of maturity-dependent sorghum traits.

REFERENCES

1. Smith CW & Frederiksen RA (2000) History of Cultivar Development in the United States: From "Memoirs of A.B. Maunder-Sorghum Breeder". *Sorghum: Origin, History, Technology, and Production*, eds Smith CW & Frederikson RA (John Wiley & Sons, New York), pp 191-223.

2. Dahlberg JA, Berenji J, & Sikora Vc (2001) Assessing sorghum [*Sorghum bicolor* (L) Moench] germplasm for new traits: food, fuels & unique uses. *Maydica* 56:85-92.

3. Maunder AB (2008) Sorghum Worldwide. *Sorghum and Millets Diseases*, (Iowa State Press), pp 11-17.

4. Rooney L*, et al.* (2010) Sorghum: An Ancient, Healthy and Nutritious Old World Cereal. ed Henley EC (United Sorghum Checkoff Program).

5. Doggett H (1988) *Sorghum* (Wiley, New York: Longman) 2nd Ed.

6. Borrell A, Jordan D, Mullet JE, Henzell B, & Hammer G (2006) Drought Adaptation in Sorghum. *Drought Adaptation in Cereals*, ed Ribaut J-M (The Hayworth Press, Inc., Binghamton), pp 335-400.

7. Wiedenfeld RP (1984) Nutrient requirements and use efficiency by sweet sorghum. *Energy in Agriculture* 3(0):49-59.

8. Board on Science and Technology for International Development (1996) Sorghum. *Lost Crops of Africa: Volume I: Grains*, (The National Academies Press), pp 127-144.

9. Foreign Agricultural Service/USDA, Office of Global Analysis (November 2010) *Grain: World Markets and Trade* (United States Department of Agriculture) Circular Series FG 11-10, retrieved from: http://www.fas.usda.gov/psdonline/psdAvailability.aspx.

10. Kimber CT (2000) Origins of Domesticated Sorghum and Its Early Diffusion to India and China. *Sorghum: Origins, History, Technology, and Production*, eds Smith CW & Frederiksen RA (John Wiley & Sons Inc., New York), pp 3-98.

11. Harlan JR & Stemler ABL (1976) The Races of Sorghum in Africa. *Origins of African Plant Domestication*, eds Harlan J, de Wet JMJ, & Stemler ABL (Mouton Press, The Hague, Netherlands), pp 465-478.

12. Quinby JR & Martin JH (1954) *Sorghum Improvement* pp 1-383.

13. Quinby JR, Kramer NW, Stephens JC, Lahr KA, & Karper RE (1958) Sorghum Production in Texas. ed Sta. TAE.

14. Multani DS*, et al.* (2003) Loss of an MDR Transporter in Compact Stalks of Maize *br2* and Sorghum *dw3* Mutants. *Science* 302:81-84.

15. Quinby JR & Karper RE (1945) Inheritance of three genes that influence time of floral initiation and maturity date in milo. *Agron J* 37:916-936.

16. Quinby JR (1966) Fourth maturity gene locus in sorghum. *Crop Sci* 6:516-518.

17. Morgan PW & Finlayson SA (2000) Physiology and Genetics of Maturity and Height. *Sorghum: Origin, History, Technology, and Production*, eds Smith CW & Frederiksen RA (John Wiley & Sons, Inc., New York City), pp 227-259.

18. Dahlberg JA (2000) Collection, Conversion, and Utilization of Sorghum. *Sorghum: Origins, History, Technology, and Production*, eds Smith CW & Frederiksen RA (John Wiley & Sons, Inc., New York), pp 309-328.

19. Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. *5th Grain Sorghum Reseach and Utilization conference*, pp 53-55.

20. Lin YR, Schertz KF, & Paterson AH (1995) Comparative Analysis of QTLs Affecting Plant Height and Maturity Across the Poaceae, in Reference to an Interspecific Sorghum Population. *Genetics* 141:391-411.

21. Klein RR*, et al.* (2008) The Effect of Tropical Sorghum Conversion and Inbred Development on Genome Diversity as Revealed by High-Resolution Genotyping. *The Plant Genome [A Supplement to Crop Sci]* 48:S12-S26.

22. Rooney WL & Aydin S (1999) Genetic Control of a Photoperiod-Sensitive Response in Sorghum bicolor (L.) Moench. *Crop Sci* 39:397-400.

23. Shoemaker CE & Bransby DI (2010) Sustainable Alternative Fuel Feedstock Opportunities, Challenges and Roadmaps for Six U.S. Regions. *Sustainable Feedstocks for Advance Biofuels Workshop*, eds Braun R, Karlen D, & Johnson D, pp 149-159.

24. Childs KL*, et al.* (1997) The sorghum photoperiod sensitivity gene, Ma(3), encodes a phytochrome B. *Plant Physiology* 113(2):611-619.

25. Paterson AH*, et al.* (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457(7229):551-556.

26. Schnable PS*, et al.* (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 326(5956):1112-1115.

27. Paterson AH (2008) Genomics of Sorghum. *Int. J. Plant Genomics* 2008.

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

29. Wu G*, et al.* (2009) The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* 138:750-759.

30. Aukerman MJ & Sakai H (2003) Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its *APETALA2*-Like Target Genes. *Plant Cell* 15:2730-2741.

31. Jung J-H*, et al.* (2007) The *GIGANTEA*-Regulated MicroRNA172 Mediates Photoperiodic Flowering Independent of *CONSTANS* in *Arabidopsis*. *Plant Cell* 19:2736-2748.

32. Michaels SD & Amasino RM (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11:949-956.

33. Kobayashi Y, Kaya H, Goto K, Iwabuchi M, & Araki T (1999) A Pair of Related Genes with Antagonistic Roles in Mediating Flowering Signals. *Science* 286:1960-1962.

34. Danilevskaya ON, Meng X, Hou Z, Ananiev EV, & Simmons CR (2008) A Genomic an Expression Compendium of the Expanded *PEBP* Gene Family from Maize. *Plant Physiol* 146:250-264.

35. Corbesier L*, et al.* (2007) FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* 316:1030-1033.

36. Jaeger KE & Wigge PA (2007) FT protein acts as a long-range signal in Arabidopsis. *Curr Biol.* 17:1050-1054.

37. Mathieu J, Warthmann N, Kuttner F, & Schmid M (2007) Export of FT protein from phoem companion cells is sufficient for floral induction in Arabidopsis. *Curr Biol.* 17:1055-1060.

38. Simpson GG (2004) The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Curr Opin Plant Biol* 7:1-5.

39. Macknight R*, et al.* (1997) *FCA*, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. *Cell* 89:737-745.

40. Simpson GG, Dijkwel PP, Quesada V, Henderson I, & Dean C (2003) FY is an RNA 30 end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell* 113:777-787.

41. Schomburg F, Patton D, Meinke D, & Amasino R (2001) *FPA*, a gene involved in floral induction in *Arabidopsis*, encodes a protein containing RNA-recognition motifs. *Plant Cell* 13:1427-1436.

42. Liu F, Marquardt S, Lister C, Swiezewski S, & Dean C (2010) Targeted 3′ Processing of Antisense Transcripts Triggers *Arabidopsis FLC* Chromatin Silencing. *Science* 327:94-97.

43. Ausin I, Alonso-Blanco C, Jarillo JA, Ruiz-Garcia L, & Martinez-Zapater JM (2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nat Genet* 36:162-166.

44. He Y, Michaels SD, & Amasino RM (2003) Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science* 302:1751-1754.

45. Lee I*, et al.* (1994) Isolation of *LUMINIDEPENDENS*: a gene involved in the control of flowering time in *Arabidopsis*. *Plant Cell* 6:75-83.

46. Lim M*, et al.* (2004) A new Arabidopsis gene, FLK, encodes an RNA binding protein with K homology motifs and regulates flowering time via FLOWERING LOCUS C. *Plant Cell* 16:731-740.

47. Mockler TC*, et al.* (2004) Regulation of flowering time in Arabidopsis by K homology domain proteins. *Proc Natl Acad Sci* 34:12759-12764.

48. Michaels SD & Amasino RM (2000) Memories of winter: vernalization and the competence to flower. *Plant Cell Environ* 23:1145-1153.

49. Sung S & Amasino RM (2004) Vernalization and epigenetics: how plants remember winter. *Curr Opin Plant Biol* 7:4-10.

50. Kim D-H, Zografos BR, & Sung S (2010) Vernalization-Mediated VIN3 Induction Overcomes the LIKE-HETEROCHROMATIN PROTEIN1/POLYCOMB REPRESSION COMPLEX2-Mediated Epigenetic Repression. *Plant Physiol* 154(949-957):949.

51. Wood C*, et al.* (2006) The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc Natl Acad Sci* 103:14631–14636.

52. Kim D-H, Doyle MR, Sung S, & Amasino RM (2009) Vernalization: Winter and the Timing of Flowering in Plants. *Annu. Rev. Cell Dev. Biol.* 25:277-299.

53. Li C, Distelfeld A, Comis A, & Dubcovsky J (2011) Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes. *Plant J.* 67:763-773.

54. Wellmer F & Riechmann JL (2010) Gene networks controlling the initation of flower development. *Trends in Genetics* 26:519-527.

55. Santner A, Calderon-Villalobos LIA, & Estelle M (2009) Plant hormones are versatile chemical regulators of plant growth. *Nat Chem Biol* 5:301-307.

56. Yamaguchi S (2008) Gibberellin Metabolism and its regulation. *Ann Rev Plant Biol* 59:225-251.

57. Hou X*, et al.* (2008) Global Identification of DELLA Target Genes during Arabidopsis Flower Development. *Plant Physiol* 147:1126-1142.

58. Hartweck LM (2008) Gibberellin signaling. *Planta* 229:1-13.

59. Eriksson S, Böhlenius H, Moritz T, & Nilsson O (2006) GA4 Is the Active Gibberellin in the Regulation of LEAFY Transcription and Arabidopsis Floral Initiation. *Plant Cell* 18:2172-2182.

60. Hisamatsu T & King RW (2008) The nature of floral signals in Arabidopsis. II. Roles for *FLOWERING LOCUS T* (*FT*) and gibberellin. *J Exp Bot* 59:3821-3829.

61. Cheng H*, et al.* (2004) Gibberellin regulates Arabidopsis floral development via suppression of DELLA protein function. *Development* 131:1055-1064.

62. Peng J (2009) Gibberellin and jasmonate crosstalk during stamen development. *J Integr Plant Biol* 51:1064-1070.

63. Achard P*, et al.* (2007) The plant stress hormone ethylene controls floral transition via DELLA-dependent regulation of floral meristem-identity genes. *Proc. Natl. Acad. Sci* 104:6484-6489.

64. Wang KL-C, Li H, & Ecker JR (2002) Ethylene biosyntesis and Siganling Networks. *Plant Cell* Supplement 2002:S131-S151.

65. Hua J & Meyerowitz EM (1998) Ethylene Responses Are Negatively Regulated by a Receptor Gene Family in *Arabidopsis thaliana*. *Cell* 94:261-271.

66. Hua J, Chang C, Sun Q, & Meyerowitz EM (1995) Ethylene insensitivity conferred by Arabidopsis ERS gene. *Science* 269:1712-1714.

67. Chang C, Kwok SF, Bleecker AB, & Meyerowitz EM (1993) *Arabidopsis* ethylene response gene *ETR1*-similarity of product to 2-component regulators. *Science* 262:539-544.

68. Sakai H*, et al.* (1998) ETR2 is an ETR1-like gene involved in ethylene signaling in Arabidopsis. *Proc Natl Acad Sci* 95:5812-5817.

69. Kyriakis JM*, et al.* (1992) Raf-1 activates MAP kinase-kinase. *Nature* 358:417-421.

70. Woodward AW & Bartel B (2005) Auxin: Regulation, Action, and Interaction. *Ann Bot-London* 95:707-735.

71. Heisler MG*, et al.* (2005) Patterns of Auxin Transport and Gene Expression during Primordium Development Revealed by Live Imaging of the Arabidopsis Inflorescence Meristem. *Current Biology* 15(21):1899-1911.

72. Grebe M (2005) Growth by Auxin: When a Weed Needs Acid. *Science* 310:60-61.

73. Feraru E & Friml J (2008) PIN Polar Targeting. *Plant Physiology* 147(4):1553-1559.

74. Leyser O (2010) The Power of Auxin in Plants. *Plant Physiology* 154(2):501-505.

75. Krizek BA (2011) Auxin regulation of *Arabidopsis* flower development involves members of the AINTEGUMENTA-LIKE/PLETHORA (AIL/PLT) family. *J Exp Bot*.

76. Kaufmann K*, et al.* (2009) Target Genes of the MADS Transcription Factor SEPALLATA3: Integration of Developmental and Hormonal Pathways in the *Arabidopsis* Flower. *PLoS Biol* 7:e1000090.

77. Stepanova Anna N & Alonso Jose M (2011) Bypassing Transcription: A Shortcut in Cytokinin-Auxin Interactions. *Developmental cell* 21(4):608-610.

78. Su Y-H, Liu Y-B, & Zhang X-S (2011) Auxin–Cytokinin Interaction Regulates Meristem Development. *Molecular Plant*.

79. Perilli S, Moubayidin L, & Sabatini S (2010) The molecular basis of cytokinin function. *Curr Opin Plant Biol* 13:21-26.

80. Sakakibara H (2006) Cytokinins: Activity, Biosynthesis, and Translocation. *Ann Rev Plant Biol* 57:431-449.

81. Bernier G (2011) My favourite flowering image: the role of cytokinin as a flowering signal. *Journal of Experimental Botany*.

82. Corbesier L*, et al.* (2003) Cytokinin levels in leaves, leaf exudate and shoot apical meristem of Arabidopsis thaliana during floral transition. *Journal of Experimental Botany* 54(392):2511-2517.

83. D'Aloia M*, et al.* (2011) Cytokinin promotes flowering of Arabidopsis via transcritional activation of the *FT* paralogue *TSF*. *Plant J* 65:972-979.

84. Bernier G, Kinet JM, & Sachs RM (1981) *Transition to reproductive growth* (CRC Press).

85. Corbesier L, Lejeune P, & Bernier G (1998) The role of carbohydrates in the induction of flowering in Arabidopsis thaliana: comparison between the wild type and a starchless mutant. *Planta* 206:131-137.

86. Rolland F, Moore B, & Sheen J (2002) Sugar Sensing and Signaling in Plants. *Plant Cell* 14:S185-S205.

87. Moore DB & Sheen J (1999) Plant sugar sensing and signaling-a complex reality. *Trends Plant Sci* 4:250.

88. Smeekens S (2000) Sugar-Induced Signal Transduction in Plants. *Annu. Rev. Plant Biol.* 51:49-81.

89. Baroja-Fernández E*, et al.* (2011) Sucrose synthase activity in the sus1/sus2/sus3/sus4 Arabidopsis mutant is sufficient to support normal cellulose and starch production. *Proc Natl Acad Sci* 109:321-326.

90. Seo PJ, Ryu J, Kang SK, & Park C-M (2011) Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in *Arabidopsis*. *Plant J* 65:418-429.

91. Colasanti J, Yuan Z, & Sundaresan V (1998) The *indeterminate* Gene Encodes a Zinc finger Protein and Regulates a Leaf-Generated signal Required for the Transition to Floweirng in Maize. *Cell* 93:593-603.

92. Turgeon R & Medville R (2004) Phloem Loading. A Reevaluation of the Relationship between Plasmodesmatal Frequencies and Loading Strategies. *Plant Physiology* 136(3):3795-3803.

93. de Mairon J (1729) Observation botanique. *Hist. Acad. Roy. Sci*:35-36.

94. McClung RC (2006) Plant Circadian Rhythms. *Plant Cell* 18:792-803.

95. Jarillo JA, Capel J, & Cashmore AR (2004) Physiological and Molecular Characteristics of Plant Circadian Clocks. *Molecular Biology of Circadian Rhythms*, ed Seghal A (John Wiley & Sons, Inc., Hoboken, NJ), pp 183-212.

96. Darwin C & Darwin F (1880) *The Power of Movement in Plants* (London).

97. Kloppstech K (1985) Diurnal and circadian rhythmicity in the expression of light-induced nuclear messenger RNAs. *Planta* 165(502-506).

98. Nagy F, Kay SA, & Chua N-H (1988) A circadian clock regulates transcription of the wheat Cab-1 gene. *Genes Dev.* 2:376-382.

99. Millar AJ & Kay SA (1991) Circadian control of *cab* gene trancription and mRNA accumulation in *Arabidopsis*. *Plant Cell* 3:541-550.

100. Harmon FG, Imaizumi T, & Kay SA (2005) The plant circadian clock: review of a clockwork *Arabidopsis*. *Endogenous Plant Rhythms*, eds Hall AJW & McWatters H (Blackwell Publishing Ltd, Oxford, UK), pp 1-18.

101. Millar AJ, Short SR, Chua N-H, & Kay SA (1992) A Novel Circadian Phenotype Based on Firefly Luciferase Expression in Transgenic Plants. *The Plant Cell* 4:1075-1087.

102. Somers DE, Webb AA, Pearson M, & Kay SA (1998) The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in *Arabidopsis thaliana*. *Development* 125:485-494.

103. Strayer C*, et al.* (2000) Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. *Science* 289:768-771.

104. Schaffer R*, et al.* (1998) The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photperiodic control of flowering. *Cell* 93:1219-1229.

105. Zagotta MT*, et al.* (1996) The *Arabidopsis* ELF3 gene regulates vegetative photomorphogenesis and the photoperiodic induction of flowering *Plant J* 10:691-702.

106. Kenigsbuch D & Tobin EM (1995) A region of the *Arabidopsis Lhcb1*3* promoter that binds to CA-1 activity is essential for high expression and phytochrome regulation. *Plant Physiol* 108:1023-1027.

107. Wang ZYT, E.M. (1998) Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* 93:1207-1217.

108. Albadi D*, et al.* (2001) Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* 293:880-883.

109. Daniel X, Sugano S, & Tobin EM (2004) CK2 phosphorylation of CCA1 is necessary for its circadian oscillator function in *Arabidopsis*. *Proc Natl Acad Sci* 12:3292-3297.

110. Stratmann T & Mas P (2008) Chromatin, photoperiod and the Arabidopsis circadian clock: A question of time. *Semin Cell Dev Biol* 19:554-559.

111. Hanao S & Davis SJ (2005) Pseudo-repsonse regulator genes 'tell' the time of day: multiple feedbacks in the circadian system of higher plants. *Endogenous Plant Rhythms*, eds Hall AJW & McWatters H (Blackwell Publishing Ltd, Oxford).

112. Matsushika A, Makino S, Kojima M, & Mizuno T (2000) Circadian Waves of Expression of the APRR1/TOC1 Family of Pseudo-Response Regulators in Arabidopsis thaliana: Insight into the Plant Circadian Clock. *Plant and Cell Physiology* 41(9):1002-1012.

113. Farré EM, Harmer SL, Harmon FG, Yanovsky MJ, & Kay SA (2005) Overlapping and Distinct Roles of *PRR7* and *PRR9* in the *Arabidopsis* Circadian Clock. *Curr Biol.* 15:47-54.

114. Pruneda-Paz JL, Breton G, Para A, & Kay SA (2009) A functional genomics approach reveals CHE as a component of the *Arabidopsis* circadian clock. *Science* 323:1481-1485.

115. Locke JCW*, et al.* (2005) Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol Syst Biol* 1:2005.0013.

116. Más P, Kim W-Y, Somers DE, & Kay SA (2003) Targeted degradation of TOC1 by ZTL modulates circadian function in *Arabidopsis thaliana*. *Nature* 426:567-570.

117. Kim W-Y*, et al.* (2007) ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nature* 449(7160):356-360.

118. Fujiwara S*, et al.* (2008) Post-translational Regulation of the *Arabidopsis* Circadian Clock through Selective Proteolysis and Phosphorylation of Pseudo-response Regulator Proteins*. *J Biol Chem* 283:23073-23082.

119. Hotta CT*, et al.* (2007) Modulation of environmental responses of plants by circadian clocks. *Plant, Cell & Environment* 30(3):333-349.

120. Somers DE, Devlin PF, & Kay SA (1998) Phytochromes and Cryptochromes in the Entrainment of the Arabidopsis Circadian Clock. *Science* 282(5393):1488-1490.

121. Pittendrigh CS & Minis DH (1964) The entrainment of circadian oscillations by light and their role as photoperiodic clocks. *Am Nat* 198:261-293.

122. Putterill J, Robson F, Lee K, Simon R, & Coupland G (1995) The CONSTANS gene of arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* 80(6):847-857.

123. Turnbull C (2011) Long-distance regulation of flowering time. *J Exp Bot* 62:4399-4413.

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

125. Imaizumi T, Schults TF, Harmon FG, Ho LA, & Kay SA (2005) FKF1 F-Box Protein Mediates Cyclic Degradation of *CONSTANS* in *Arabidopsis*. *Science* 309:293-297.

126. Fornara F*, et al.* (2009) Arabidopsis DOF Transcription Factors Act Redundantly to Reduce *CONSTANS* Expression and Are Essential for a Photoperiodic Flowering Response. *Dev Cell* 17:75-86.

127. Nakamichi N*, et al.* (2007) Arabidopsis Clock-Associated Pseudo-Response Regulators PRR9, PRR7 and PRR5 Coordinately and Positively Regulate Flowering Time Through the Canonical CONSTANS-Dependent Photoperiodic Pathway. *Plant and Cell Physiology* 48(6):822-832.

128. Sawa M, Nusinow DA, Kay SA, & Imaizumi T (2007) FKF1 and GIGANTEA Complex Formation Is Required for Day-Length Measurement in *Arabidopsis*. *Science* 318:261-265.

129. Suarez-Lopez P*, et al.* (2001) CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature* 410(6832):1116-1120.

130. Zuo Z, Liu H, Liu B, Liu X, & Lin C (2011) Blue Light-Dependent Interaction of CRY2 with SPA1 Regulates COP1 activity and Floral Initation in Arabidopsis. *Curr Biol.* 21:841-847.

131. Valverde F*, et al.* (2004) Photoreceptor Regulation of CONSTANS Protein in Photoperiodic Flowering. *Science* 303:1003-1006.

132. Wenkel S*, et al.* (2006) CONSTANS and the CCAAT Box Binding Complex Share a Functionaly Important Domain and Interact to Regulate Flowering of Arabidopsis. *Plant Cell* 18:2971-2984.

133. Gusmaroli G, Tonelli C, & Mantovani R (2002) Regulation of novel members of the Arabidopsis thaliana CCAAT-binding nuclear factor Y subunits. *Gene* 283(1–2):41-48.

134. Siefers N*, et al.* (2009) Tissue-Specific Expression Patterns of Arabidopsis NF-Y Transcription Factors Suggest Potential for Extensive Combinatorial Complexity. *Plant Physiology* 149(2):625-641.

135. Castillejo C & Pelaz S (2008) The Balance between CONSTANS and TEMPRANILLO Activities Determines *FT* Expression to Trigger Flowering. *Curr Biol.* 18:1338-1343.

136. Sawa M & Kay SA (2011) GIGANTEA directly activates *Flowering Locus T* in *Arabidopsis thaliana*. *Proc Natl Acad Sci* 108:11698-11703.

137. Laux T, Mayer KFX, Berger J, & Jürgens G (1996) The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. *Development* 122:87-96.

138. Long JA, Moan EI, Medford JI, & Barton MK (1996) A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis. *Nature* 379(6560):66-69.

139. Gordon SP, Chickarmane VS, Ohno C, & Meyerowitz EM (2009) Multiple feedback loops through cytokinin signaling control stem cell number within the Arabidopsis shoot meristem. *Proceedings of the National Academy of Sciences* 106(38):16529-16534.

140. Lee J & Lee I (2010) Regulation and function of SOC1, a flowering pathway integrator. *Journal of Experimental Botany* 61(9):2247-2254.

141. Liu C*, et al.* (2008) Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. *Development* 135(8):1481-1491.

142. Ferrandiz C, Gu Q, Martienssen R, & Yanofsky MF (2000) Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* 127(4):725-734.

143. Liljegren SJ, Gustafson-Brown C, Pinyopich A, Ditta GS, & Yanofsky MF (1999) Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 Specify Meristem Fate. *The Plant Cell Online* 11(6):1007-1018.

144. Krizek BA & Fletcher JC (2005) Molecular Mechanisms of Flower Development: An Armchair Guide. *Nat Rev Genet* 6(688-698).

145. Coen ES & Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353(6339):31-37.

146. Kaufmann K*, et al.* (2010) Orchestration of Floral Initiation by APETALA1. *Science* 328:85-89.

147. Riechmann JL & Meyerowitz EM (1997) MADS domain proteins in plant development. *Biol. Chem.* 378:1079-1101.

148. Wang H, Caruso LV, Downie AB, & Perry SE (2004) The Embryo MADS Domain Protein AGAMOUS-Like 15 Directly Regulates Expression of a Gene Encoding an Enzyme Involved in Gibberellin Metabolism. *The Plant Cell Online* 16(5):1206-1219.

149. Weigel D & Meyerowitz EM (1994) The ABCs of Floral Homeotic Genes. *Cell* 78:203-209.

150. Pinyopich A*, et al.* (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424(6944):85-88.

151. Pelaz S, Ditta GS, Baumann E, Wisman E, & Yanofsky MF (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405(6783):200-203.

152. Ditta G, Pinyopich A, Robles P, Pelaz S, & Yanofsky MF (2004) The SEP4 Gene of Arabidopsis thaliana Functions in Floral Organ and Meristem Identity. *Current Biology* 14(21):1935-1940.

153. Theißen G (2001) Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology* 4(1):75-85.

154. Cheng Y, Kato N, Wang W, Li J, & Chen X (2003) Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in Arabidopsis thaliana. *Dev Cell* 4:53-66.

155. Li J & Chen X (2003) PAUSED, a Putative Exportin-t, Acts Pleiotropically in Arabidopsis Development But Is Dispensable for Viability. *Plant Physiology* 132(4):1913-1924.

156. Yano MK*, et al.* (2000) Hd1, a major photoperiod senstitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. . *Plant Cell* 12:2473-2483.

157. Kojima S*, et al.* (2002) Hd3a, a Rice Ortholog of the Arabidopsis FT Gene, Promotes Transition to Flowering Downstream of Hd1 under Short-Day Conditions. *Plant and Cell Physiology* 43(10):1096-1105.

158. Tsuji H, Taoka K-I, & Shimamoto K (2010) Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. *Current Opinion in Plant Biology* 14:1-8.

159. Izawa T*, et al.* (2002) Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. *Genes & Development* 16(15):2006-2020.

160. Yan W-H*, et al.* (2011) A major QTL, *Ghd8*, Plays Pleiotropic Roles in Regulation Grain Productivity, Plant Height, and Heading Date in Rice. *Mol Plant* 4:319-330.

161. Hayama R & Coupland G (2004) The Molecular Basis of Diversity in the Photoperiodic Flowering Responses of Arabidopsis and Rice. *Plant Physiology* 135(2):677-684.

162. Komiya R, Ikegami A, Tamaki S, Yokoi S, & Shimamoto K (2008) Hd3a and RFT1 are essential for flowering in rice. *Development* 135:767-774.

163. Doi K*, et al.* (2004) Ehd1, a B-type response regulator in rice, confers short-day promotion of flower and controls FT-like gene expression independently of Hd1. *Gene Dev* 18:926-936.

164. Xue W*, et al.* (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat Genet* 40(6):761-767.

165. Peng L-T, Shi Z-Y, Li L, Shen G-Z, & Zhang J-L (2008) Overexpression of transcription factor OsLFL1 delays flowering time in Oryza sativa. *Journal of Plant Physiology* 165(8):876-885.

166. Itoh H, Nonoue Y, Yano M, & Izawa T (2010) A pair of floral regulators sets critical day length for *Hd3a* florigen expression in rice. *Nat Genet* 42:635-639.

167. Matsubara K*, et al.* (2011) *Ehd3*, encoding a plant homeodomain finger-containing protein, is a critical promoter of rice flowering. *Plant J* 66:603-612.

168. Higgins JA, Bailey PC, & Laurie DA (2010) Comparative Genomics of Flowering Time Pathways Using *Brachypodium distachyon* as a Model for the Temperate Grasses. *PLoS ONE* 5(4):e10065.

169. Smartt J (1990) Evolution. *Grain Legumes: Evolution and Genetic Resources*, (Cambridge University Press, Cambridge, UK), p 125.

170. 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 and Cell Physiology* 46(8):1175-1189.

171. 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.

172. Chardon F & Damerval C (2005) Phylogenomic analysis of the PEBP gene family in cereals. *J Mol Evol* 61:2169-2185.

173. 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* 23:942-960.

174. Rooney WL, Blumenthal J, Bean B, & Mullet JE (2007) Designing sorghum as a dedicated bioenergy feedstock. *Biofuel Bioprod Bior* 1(2):147-157.

175. Quinby JR (1974) *Sorghum improvement and the genetics of growth* (Texas A&M Univ. Press, College Station).

176. Bünning E (1960) Circadian rhythms and the time measurement in photoperiodism. *Cold Spring Harbor Symp Quant Bio*, pp 249-256.

177. Izawa T (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *J Exp Bot* 58(12):3091-3097.

178. Matsubara K*, et al.* (2008) Ehd2, a Rice Ortholog of the Maize INDETERMINATE1 Gene, Promotes Flowering by Up-Regulating Ehd1. *Plant Physiol* 148:1425-1435.

179. Craufurd PQ*, et al.* (1999) Adaptation of sorghum: characterisation of genotypic flowering responses to temperature and photoperiod. *Theor Appl Genet* 99:900-911.

180. Quinby JR (1967) The Maturity Genes of Sorghum. *Advances in Agronomy*, ed Norman AG (Academic Press), Vol 19, pp 267-305.

181. Stephens J, Miller FR, & Rosenow DT (1967) Conversion of alien sorghums to early combine genotypes. *Crop Sci* 7:396.

182. Turner A, Beales J, Faure S, Dunford RP, & Laurie DA (2005) The Pseudo-Response Regulator Ppd-H1 Provides Adaptation to Photoperiod in Barley. *Science* 310:1031-1034.

183. Beales J, Turner A, Griffiths S, Snape JW, & Laurie DA (2007) A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). *Theor Appl Genet* 115:721-733.

184. Murphy RL*, et al.* (2011) Coincident light and clock regulation of *pseudoresponse regulator protein 37* (*PRR37*) controls photoperiodic flowering in sorghum. *Proc Natl Acad Sci* 108:16469-16474.

185. Saitou N & Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.

186. Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

187. Tamura K, Dudley J, Nei M, & Kumar S (2007) MEGA4: Molecular Evolution Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.

188. Whitney IE*, et al.* (2011) Genetic modulation of horizontal cell number in the mouse retina. *Proc Natl Acad Sci* (108):9697-9702.

189. Bookout AL & Mangelsdorf DJ (2003) Quantitative real-time PCR protocol for analysis of nuclear receptor signaling pathways. *Nucl Recept Signal* 1:e012.

190. Farré EM & Kay SA (2007) PRR7 protein levels are regulated by light and the circadian clock in Arabidopsis. *The Plant Journal* 52:548-560.

191. Miller TA, Muslin EH, & Dorweiler JE (2008) A maize *CONSTANS*-like gene, *conz1*, exhibits distinct diurnal expression patterns in varied photoperiods. *Planta* 227:1377-1388.

192. Ito S*, et al.* (2009) A Genetic Study of theArabidopsis Circadian Clock with Reference to the TIMING OF CAB EXPRESSION 1 (TOC1) Gene. *Plant Cell Physiol* 50(2):290-303.

193. Mizuno T & Nakamichi N (2005) *Pseudo*-Response Regulators (PRRs) or *True* Oscillator Components (TOCs). *Plant Cell Physiol* 46(5):677-685.

194. Hanumappa M, Pratt LH, Cordonnier-Pratt MM, & Deitzer G (1999) A photoperiod-insensitive barley line contains a light-labile phytochrome B. *Plant Physiol* 119:1033-1040.

195. Takano M*, et al.* (2005) Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. *Plant Cell* 17(12):3311-3325.

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

197. Menz MA*, et al.* (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP, RFLP, and SSR markers. *Plant Mol Biol* 48:483-499.

198. Temnykh S*, et al.* (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11:1441-1452.

199. Morgan PW, Guy LW, & Pao C-I (1986) Genetic Regulation of Development in *Sorghum bicolor*. *Plant Physiol* 83:448-450.

200. Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, & Lukyanov S (1995) An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res* 23:1087-1088.

201. Rubio V & Deng XW (2007) Standing on the Shoulders of GIGANTEA. *Science* 318(5848):206-207.

202. Knabe JF, Wegner K, Nehaniv CL, & Schilstra MJ (Genetic Algorithms and Their Application to In Silico Evolution of Genetic Regulatory Networks.), Vol 673, pp 297-321.

203. Kaufmann K, Nagasaki M, & Jáuregui R (2010) Modelling the molecular interactions in the flower developmental network of *Arabidopsis thaliana*. *In Silico Biol* 10.

204. Folliard A, Traoré PCS, Vaksmann M, & Kouressy M (2004) Modeling of sorghum response to photoperiod: a threshold–hyperbolic approach. *Field Crops Research* 89(1):59-70.

205. Hammer G*, et al.* (2006) Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science* 11:587-593.

206. Pao C-I & Morgan PW (1986) Genetic Regulation of Development in *Sorghum bicolor*. *Plant Physiol* 82:575-580.

APPENDIX A

PRELIMINARY MAPPING AND CHARACTERIZATION OF *MATURITY LOCUS 7* (*Ma_7*)

INTRODUCTION

In a 1999 study by Rooney and Aydin, two maturity loci, *Ma_5* and *Ma_6*, were reported that caused extremely late flowering in the offspring of otherwise early-flowering parents (A1). One of these loci, *Ma_6*, was identified as the sorghum ortholog of the rice *GRAIN NUMBER, PLANT HEIGHT AND HEADING DATE* 7 (*Ghd7*) gene. However, initial mapping attempts of the *Ma_5* locus by Jeff Brady in 2003 revealed that in R.07007, this was actually two genes located on two different chromosomes (A2). One of these mapped to chromosome one, while the other mapped to a location on chromosome two, and were named *Ma_5* (previously *Ma_5-2*) and *Ma_7* (previously *Ma_5-1*), respectively. Both of these loci contribute to the repression of floral initiation; when neither are functional, flowering occurs early, and, by contrast, when both are functional flowering occurs extremely late. When either one or the other is functional, an intermediate flowering time is observed. Moreover, in order to confer the extremely photoperiod sensitive phenotype, these genes must be present in a dominant *Ma_1* and *Ma_6* background. In *$ma_1Ma_5ma_6Ma_7$* genotypes, very early flowering is still observed. This suggests that these genes interact genetically with *Ma_1* (*Ghd7*) and *Ma_6* (*PRR37*), and that their primary effect lies upstream of these genes The major focus of this study was to identify the gene underlying the *Ma_7* locus, and to determine how it may affect the known maturity genes.

RESULTS AND DISCUSSION

Mapping of *Ma$_7$*

Initial mapping of this locus was carried out by Jeff Brady in a RTx436 x R.07007 BC_1F_1 population created by backcrossing the F_1 offspring to R.07007. This population was set out in fields in College Station, TX in April of 2003, and phenotyped for days to anthesis (n=2915). Preliminary fine mapping was carried out by Karen Harris in a subset of the population between markers 59L10 and *Xtxp428*, and in 2009, the remaining population (n=2443) was screened at these markers to increase the number of available recombinants (Fig. A1). These recombinants were used for further fine mapping. SSR markers located by SSRIT (A3) and SNPs developed by *de novo* sequencing of the regions directly upstream of predicted open reading frames were utilized (Table A1), and this resulted in a locus delimited by SSR4 and SSR12 (Fig. A2). This region is ~258kb, and contains 36 candidate genes (Table A2). Among these genes is an ortholog of *FUSCA5* (*FUS5*), the CSN7 component of the COP9 signalosome

Table A1. Primers for markers used in *Ma$_7$* Mapping.

	Forward	Reverse	SBI02
59L10	TGGACTAAACTCGCCAGGAG	GAACCTGGAGCTCGGGTAGT	67,934,257
SSR 4	GTTGCACTTCCTGCATCT	CGATTGGAAGAAACCGTGCAAC	68,094,785
SNP1	TCCCTTGGATTGGAATGGAGGGTT	GTGCCTTATGGCGAATGCATGGAA	68,142,221
SNP2	CGCGCCTTGTTCGCTTGAGTTTAT	GCGATCACAATGCTAGCGTGACTT	68,292,934
SSR 12	TGCTGCCTTGGGAGTTGTTT	CAGACACTAGGGAAGACTCCTTCT	68,353,149
SSR 15	TGCATGCATGGCAAGATTTACA	ACTGTTGATTCAGTACATCACATCTCT	68,372,647
Txp428	CACTGGCCAAGGTTTCACTT	CATGGAATGCAACATAGCAA	68,393,144

Primers are given with their physical coordinates on chromosome 2 based on Phytozome v8.0

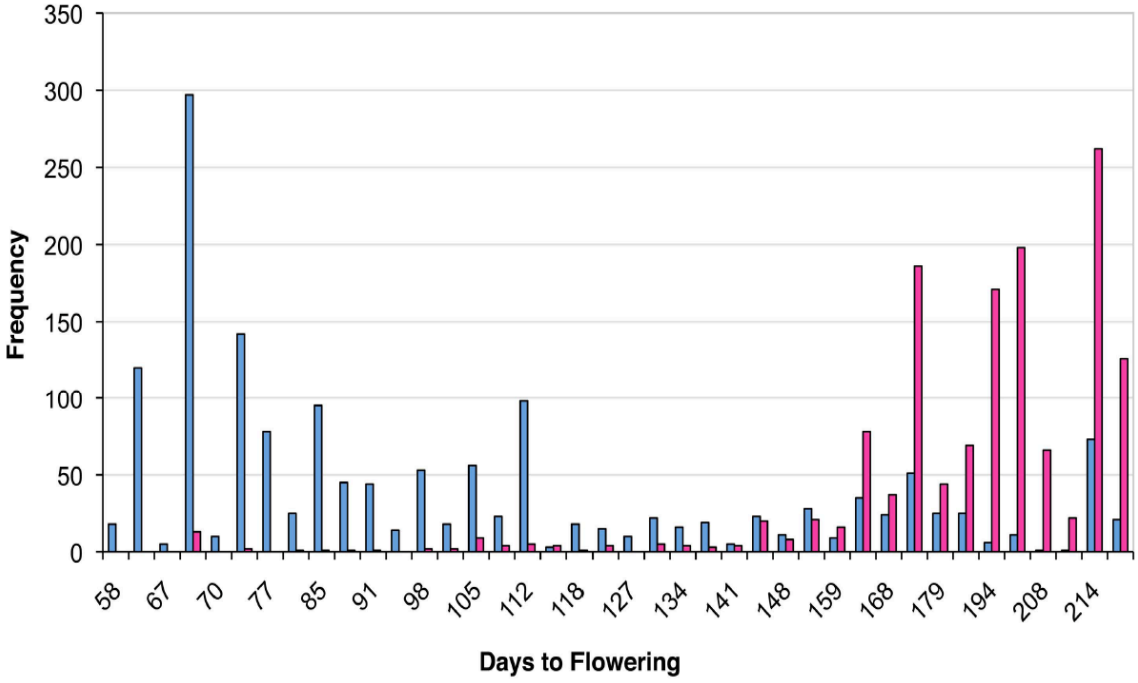

Fig. A1. Genotype at *Xtxi428* in the RTx436 x R.07007 BC_1F_1 population. Plants that are ma_7ma_7 (R.07007 allele) are shown in blue. Plants with the Ma_7ma_7 genotype (heterozygotes) are shown in pink.

(Sb02g033680). The COP9 signalosome (CSN) is a complex of eight subunits that functions in the ubiquitin-protease pathway to regulate diverse cellular processes (A4). This gene was chosen as the best candidate because the CSN modulates many aspects of development, including light responses that could potentially affect the photoperiod pathway (A5). The CSN is also known to be transcriptional repressor of genes that are turned off in the dark (A6). Additionally, while many null alleles of CSN are lethal, weak alleles exhibit other phenotypes including early flowering and defects in circadian rhythms (A7). Thus, in order to determine whether Ma_7 may encode CSN7 (*FUS5*), the

expression of this gene in response to photoperiod was analyzed, and *ma*$_7$ seedlings were phenotyped for growth in red light.

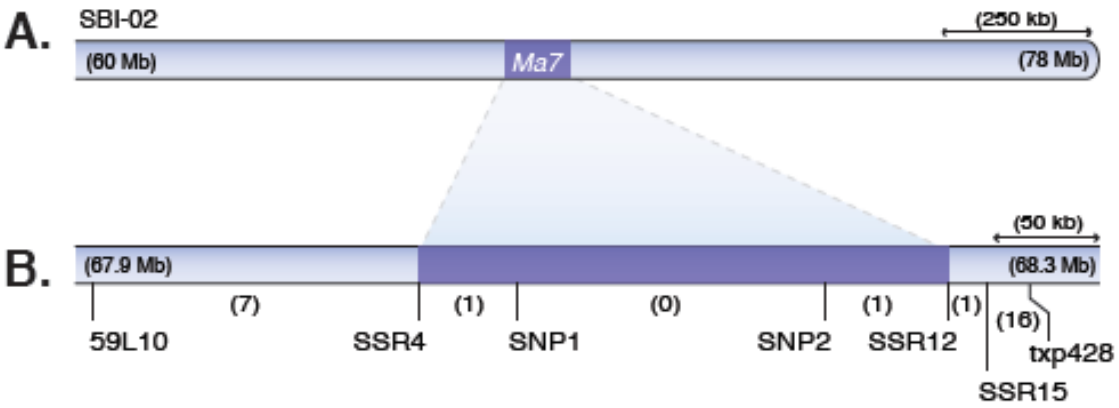

Fig. A2. Fine mapping of the *Ma*$_7$ locus. (*A*) *Ma*$_7$ locus mapped to a region on chromosome 2 in a BC_1F_1 population (n = 2,443) derived from RTx436 and R.07007 (*B*) *Ma*$_7$ mapped to an ~250-kb region delimited by SSR4 and SSR12. Recombination events are shown in parentheses, physical coordinates are at the end of each chromosome segment, and the *Ma*$_7$ locus is shaded in purple.

Table A2. Genes present in the ~250-kb interval mapped in the RTx436 by R.07007 BC_1F_1 population.

Sorghum Gene	Start	End	Best *At* hit	Best *At* TAIR10 hit defined	Best rice hit	Best rice hit defined
Sb02g033460	68086393	68087447	AT3G15351		LOC_Os07g30220	expressed protein
Sb02g033470	68089083	68095840	AT4G17380	MUTS-like protein 4	LOC_Os07g30240	mutS family domain IV containing protein, expressed
Sb02g033480	68101916	68103773	AT4G17380	MUTS-like protein 4	LOC_Os07g30240	mutS family domain IV containing protein, expressed
Sb02g033490	68107371	68108601	AT4G17380	MUTS-like protein 4	LOC_Os07g30240	mutS family domain IV containing protein, expressed
Sb02g033500	68109887	68110624			LOC_Os07g32870	expressed protein
Sb02g033510	68112494	68114742	AT5G07630	lipid transporters	LOC_Os07g30250	RFT1, putative, expressed
Sb02g033515	68116586	68118193	AT5G07630	lipid transporters		
Sb02g033520	68119870	68127225	AT2G04230	FBD, F-box and Leucine Rich Repeat domains containing protein	LOC_Os12g36000	expressed protein
Sb02g033530	68143796	68146148	AT4G16720	Ribosomal protein L23/L15e family protein	LOC_Os03g40180	60S ribosomal protein L15, putative, expressed
Sb02g033540	68151510	68156982	AT5G61530	small G protein family protein / RhoGAP family protein	LOC_Os07g30300	small G protein family protein, putative, expressed
Sb02g033550	68158726	68160814	AT3G60400	Mitochondrial transcription termination factor family protein	LOC_Os07g22670	mTERF domain containing protein, expressed
Sb02g033560	68162346	68165715			LOC_Os07g33090	expressed protein
Sb02g033570	68166723	68168186	AT1G22360	UDP-glucosyl transferase 85A2	LOC_Os07g30469	indole-3-acetate beta-glucosyltransferase, putative, expressed
Sb02g033580	68172202	68174750	AT1G22360	UDP-glucosyl transferase 85A2	LOC_Os07g30469	indole-3-acetate beta-glucosyltransferase, putative, expressed
Sb02g033590	68176892	68178609	AT3G16250	NDH-dependent cyclic electron flow 1	LOC_Os07g30670	2Fe-2S iron-sulfur cluster binding domain containing protein, expressed
Sb02g033600	68183962	68185002	AT3G55500	expansin A16	LOC_Os12g36040	expansin precursor, putative, expressed
Sb02g033610	68185785	68188048	AT3G09430		LOC_Os07g49510	expressed protein
Sb02g033620	68189250	68198270	AT3G16940	calmodulin binding;transcription regulators	LOC_Os07g30774	calmodulin-binding transcription activator 5, putative, expressed
Sb02g033630	68206092	68210802	AT5G51220	ubiquinol-cytochrome C chaperone family protein	LOC_Os07g30790	ubiquinol-cytochrome C chaperone family protein, putative, expressed
Sb02g033640	68210952	68214946	AT3G60370	FKBP-like peptidyl-prolyl cis-trans isomerase family protein	LOC_Os07g30800	immunophilin, putative, expressed
Sb02g033650	68219221	68224519	AT4G38050	Xanthine/uracil permease family protein	LOC_Os07g30810	nucleobase-ascorbate transporter, putative, expressed
Sb02g033660	68225294	68228583	AT1G67325	Ran BP2/NZF zinc finger-like superfamily protein	LOC_Os07g30820	zinc finger family protein, putative, expressed

Table A2. Continued

Sorghum Gene	Start	End	Best *At* hit	Best *At* TAIR10 hit defined	Best rice hit	Best rice hit defined
Sb02g033670	68231152	68232088	AT2G19830	SNF7 family protein	LOC_Os07g30830	SNF7 family protein, putative, expressed
Sb02g033680	68233702	68238431	AT1G02090	Proteasome component (PCI) domain protein	LOC_Os07g30840	proteasome subunit, putative, expressed
Sb02g033690	68243464	68249907	AT1G02120	GRAM domain family protein	LOC_Os07g30940	expressed protein
Sb02g033700	68251429	68253190	AT2G35660	FAD/NAD(P)-binding oxidoreductase family protein	LOC_Os07g30960	monooxygenase, putative, expressed
Sb02g033710	68254058	68256297	AT4G09320	Nucleoside diphosphate kinase family protein	LOC_Os07g30970	nucleoside diphosphate kinase, putative, expressed
Sb02g033720	68259898	68261175	AT5G63370	Protein kinase superfamily protein	LOC_Os09g27350	RIO1 family protein, expressed
Sb02g033730	68267823	68269100	AT5G63370	Protein kinase superfamily protein	LOC_Os09g27350	RIO1 family protein, expressed
Sb02g033740	68275961	68277238	AT5G63370	Protein kinase superfamily protein	LOC_Os09g27350	RIO1 family protein, expressed
Sb02g033750	68279283	68281384				
Sb02g033755	68287250	68290811	AT3G17380	TRAF-like family protein	LOC_Os10g33830	MATH domain containing protein, expressed
Sb02g033760	68289845	68292019	AT1G48590	Calcium-dependent lipid-binding (CaLB domain) family protein	LOC_Os07g31720	GTPase activating protein, putative, expressed
Sb02g033770	68294743	68295897	AT5G44660		LOC_Os07g31660	expressed protein
Sb02g033780	68296525	68301648	AT5G04460	RING/U-box superfamily protein	LOC_Os07g31650	expressed protein
Sb02g033790	68326064	68327876	AT1G30870	Peroxidase superfamily protein	LOC_Os07g31610	peroxidase precursor, putative, expressed
Sb02g033800	68328490	68332181	AT2G14530	TRICHOME BIREFRINGENCE-LIKE 13	LOC_Os07g31550	powdery mildew resistant protein 5, putative, expressed
Sb02g033810	68353762	68357848	AT4G20140	Leucine-rich repeat transmembrane protein kinase	LOC_Os07g31500	leucine-rich repeat receptor protein kinase EXS precursor, putative, expressed

The physical coordinates and functional annotation of each gene are given as based on Phytozome v8.0

Expression of *FUS5*

The expression of *CSN7* (*FUS5*) was analyzed in ATx623 (Ma_7) R.07007 (ma_7) and the ATx623 x R.07007 F_1 (Ma_7ma_7) over a course of 96 hours. Plants were grown in LD for 32 days and then transferred to SD or LD treatment for an additional week. Following treatment under the respective photoperiod, tissue was collected for each genotype every three hours for a single day/night cycle followed by 72 hours of constant light (Fig. 18). The expression of *CSN7* does not appear to be circadian clock-regulated, nor does it show any distinct response to light. Plants grown under 10-hour SDs (Fig. A3, red dashed line) exhibited higher expression relative to LD-treated plants (Fig. A3, solid black line) in all three genotypes. However, this increased expression continues in LL conditions, and therefore does not seem to be immediately repressed by light exposure.

Red Light Response of ma_7 Genotypes.

The CSN regulates responses to light and other stimuli both through SCF-mediated protein degradation and through transcriptional regulation (A6), and in weak COP9 mutants, an extremely dwarfed hypocotyl is often observed in dark grown seedlings relative to wild type (A8). Therefore, in order to characterize this response in ma_7 plants, ATx623, R.07007, and ATx623 x R.07007 F_1 seedlings were germinated under red light conditions. The total and mesocotyl lengths were measured and the ratio was compared between genotypes.

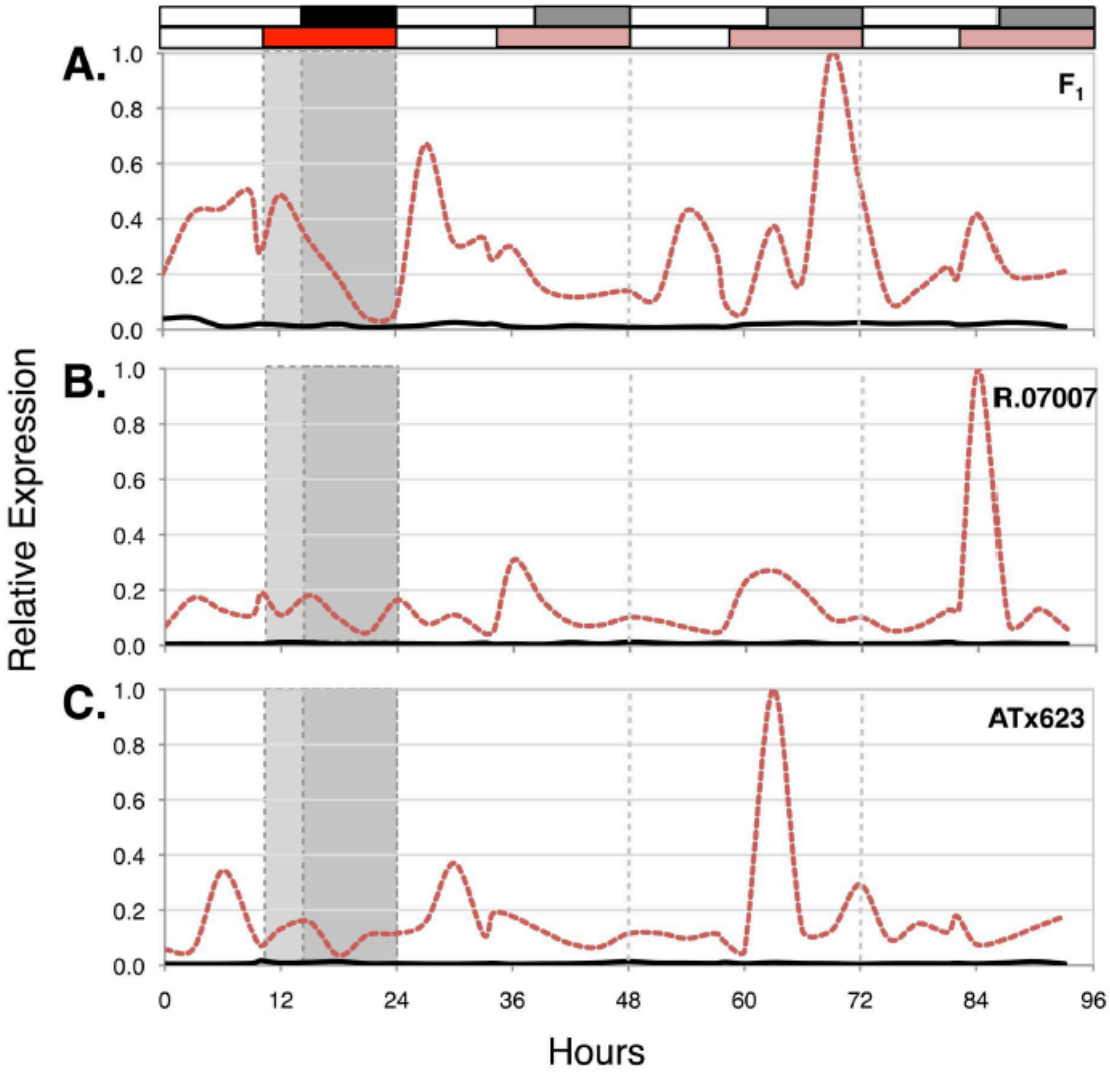

Fig. A3. *FUS5* expression analysis. The ATx623 x R.07007 F_1 (*A*) R.07007 (*B*) and ATx623 (*C*) do not exhibit circadian regulation. The ordinate represents expression normalized to 18S ribosomal RNA expression and relative to a calibrator sample (A9). The black bar above the plot indicates the dark period for LD-treated plants; gray bars indicate subjective dark during LL conditions. Red bars indicate darkness for SD-treated plants; pink indicates subjective dark during LL conditions. Open bars denote light periods. Light-gray shading within the plot area indicates darkness for SD treated plants only; dark-gray shading indicates darkness for both LD- and SD-treated plants.

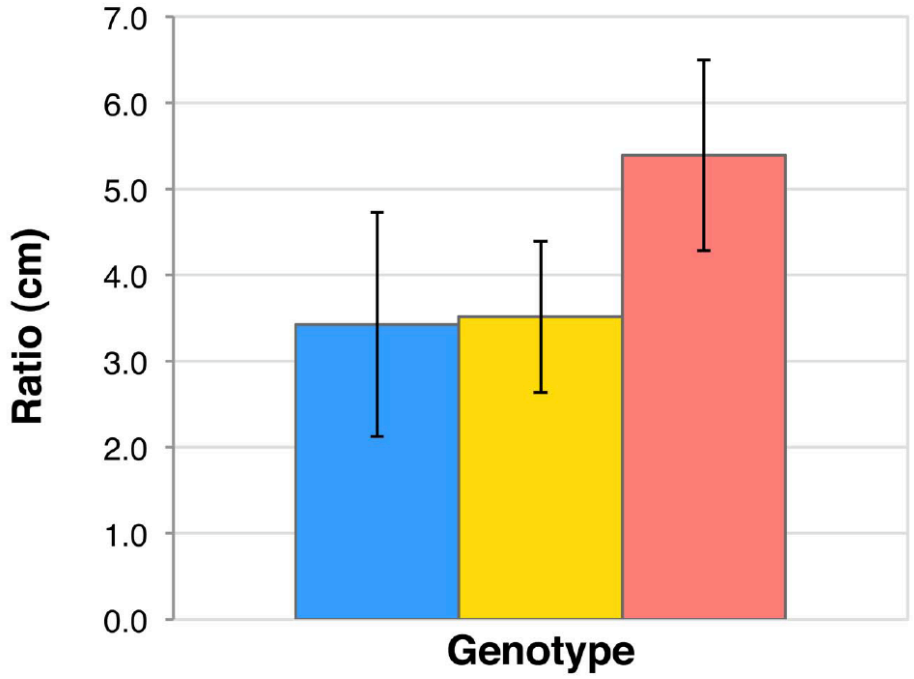

Fig. A4. Red light-grown *Ma*$_7$ and *ma*$_7$ genotypes. ATx623 (blue), R.07007 (red), and the ATx623 x R.07007 F_1 (yellow) seedlings were germinated in under red light conditions and grown for (5 days??). After this period, plants were removed and the mesocotyl and total lengths were measured in cm. The ordinate represents the mesocotyl:total length ratio ± SD; n=30.

In ATx623 (*Ma*$_7$) and the F_1 (*Ma*$_7$*ma*$_7$), the length ratio is ~3.4, while in R.07007 (*ma*$_7$), it is ~5.4, indicating that the mesocotyl is proportionally longer in this genotype (Fig. A4). However, this difference is not significant (Student's T-test, $p>0.05$), nor is this response with what is observed for *Arabidopsis*. Furthermore, it would be important to germinate these seedlings under dark conditions, because a primary phenotype observed in *cop9* mutants is the de-repression of light responsive genes under conditions without light.

Discussion

The locus that encodes *Ma*$_7$ was ultimately refined to ~250kb on chromosome 2, which spans a region containing 36 genes. Though COP9 subunit 7 (*FUS5*) is present in this interval, other genes cannot be ruled out. The expression and red light growth analysis do not provide unequivocal evidence that *CSN7* is *Ma*$_7$. However, in addition to direct sequencing of this gene and other candidates, certain other experiments could be performed that more decisively answer this question. In the initial characterization of COP9, it was observed that the expression of the *cab1* gene is under COP9-mediated light control, and the mutant exhibits repressed *cab1* expression in light-grown seedlings. Therefore, the expression of this gene and other light-responsive genes should be analyzed to determine if there are any differences in the *ma*$_7$ background. Additionally, the accumulation of anthocyanins is greater in dark-grown *cop9* mutants than in wild type *Arabidopsis* seedlings. These types of experiments could be used to further distinguish whether or not the CSN7 gene underlies *Ma*$_7$.

In addition to identifying the *Ma*$_7$ gene, it will be important to characterize the interaction between it and the other maturity loci. In R.07007 (*Ma*$_1$*ma*$_5$*Ma*$_6$*ma*$_7$), the expression of *PRR37* (*Ma*$_1$) and *Ghd7* (*Ma*$_6$) is altered, resulting in decreased photoperiod sensitivity. It can be hypothesized that the expression of PRR37 and Ghd7 is regulated by either *Ma*$_5$, *Ma*$_7$, or both, and when R.07007 is crossed to ATx623, dominant at *Ma*$_5$ and *Ma*$_7$, the expression patterns are restored. In addition to the potential effects of *Ma*$_7$ on *PRR37* and *Ghd7*, a connection exists between *Ma*$_7$ and *Ma*$_2$. *Ma*$_7$ maps to a position on chromosome 2 that coincides with the region identified as the

Ma$_2$ locus in crosses with 80M, the *Ma*$_2$ maturity standard (Olson *et al.*, unpublished). Therefore it was determined that the *Ma*$_7$ locus likely encodes the same gene as the *Ma*$_2$ locus. In addition to *Ma*$_2$ breeding materials, the resources initially used to define the *Ma*$_7$ locus can be used to further define *Ma*$_2$.

REFERENCES

A1. Rooney WL & Aydin S (1999) Genetic Control of a Photoperiod-Sensitive Response in Sorghum bicolor (L.) Moench. *Crop Sci* 39:397-400.

A2. Brady JA (2006) Sorghum Ma5 and Ma6 maturity genes. Ph.D. (Texas A&M University, College Station).

A3. Temnykh S*, et al.* (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11:1441-1452.

A4. Wei N, Serino g, & Deng X-W (2008) The COP9 signalosome: more than a protease. *Cell* 33:592-600.

A5. Wei N & Deng XW (2003) THE COP9 SIGNALOSOME. *Annual Review of Cell and Developmental Biology* 19(1):261-286.

A6. Chamovitz DA (2009) Revisiting the COP9 signalosome as a transcriptional regulator. *EMBO Rep* 10(4):352-358.

A7. Leyser O & Day S (2002) Light. *Mechanisms in Plant Development*, (John Wiley & Sons-Blackwell Science Ltd, Oxford), pp 138-164.

A8. Wei N & Deng X-W (1992) COP9: A New Genetic Locus Involved in Light-Regulated Development and Gene Expression in Arabidopsis. *Plant Cell* 4:1507-1518.

A9. Bookout AL & Mangelsdorf DJ (2003) Quantitative real-time PCR protocol for analysis of nuclear receptor signaling pathways. *Nucl Recept Signal* 1:e012.

VITA

Name: Rebecca Lea Murphy

Address: Texas A&M University, Dept. of Biochemistry/Biophysics, 2128 TAMU, College Station, TX 77843-2128

Email Address: rmurphy46@neo.tamu.edu

Education: B.S., Biology, Centenary College of Louisiana, 2006
Ph.D., Biochemistry, Texas A&M University, 2012

Publications:

- **Murphy, R.L.**, Morishige, D.T., Brady, J.A., Rooney, W.L., Miller, F.R., Klein, P.E., and Mullet, J.E. The Identification of *Maturity Locus 6* (In preparation, expected summer 2012)
- **Murphy, R.L.**, Klein, R.R., Morishige, D.T., Brady, J.A., Rooney, W.L., Miller, F.R., Dugas, D.V., Klein, P.E., and Mullet, J.E. (2011). Coincident light and clock regulation of *pseudoresponse regulator protein 37* (*PRR37*) controls photoperiodic flowering in sorghum. *Proc. Natl. Acad. Sci. USA*. **108**, 16469-16474 September 2011
- Mullet, J.E., Rooney, W.L., Klein, P.E., Morishige, D., **Murphy, R.**, and Brady, J.A. (2010). Discovery and Utilization of Sorghum Genes (*Ma5*/*Ma6*). US. Pub. No.: US 2012/0024065 A1. Filed Jul. 21, 2009. Issue date: Jan. 28, 2010. January 2010