Maturity2, a novel regulator of flowering time in *Sorghum bicolor*, increases expression of *SbPRR37* and *SbCO* in long days delaying flowering

Running title: *Maturity2* - a novel regulator of flowering time in *Sorghum bicolor*

Anna L. Casto[1¶], Ashley J. Mattison[1¶], Sara N. Olson[1], Manish Thakran[1], William L. Rooney[2], John E. Mullet[1*]

[1] Biochemistry and Biophysics Department, Texas A&M University, College Station, TX, United States of America

[2] Soil and Crop Science Department, Texas A&M University, College Station, TX, United States of America

* Corresponding author

E-mail: jmullet@tamu.edu (JEM)

¶These authors contributed equally to this work

Abstract

Sorghum bicolor is a drought-resilient facultative short-day C4 grass that is grown for grain, forage, and biomass. Adaptation of sorghum for grain production in temperate regions resulted in the selection of mutations in *Maturity* loci (Ma_1 – Ma_6) that reduced photoperiod sensitivity and resulted in earlier flowering in long days. Prior studies identified the genes associated with Ma_1 (*PRR37*), Ma_3 (*PHYB*), Ma_5 (*PHYC*) and Ma_6 (*GHD7*) and characterized their role in the flowering time regulatory pathway. The current study focused on understanding the function and identity of Ma_2. Ma_2 delayed flowering in long days by selectively enhancing the expression of *SbPRR37* (Ma_1) and *SbCO,* genes that co-repress the expression of *SbCN12*, a source of florigen. Genetic analysis identified epistatic interactions between Ma_2 and Ma_4 and located QTL corresponding to Ma_2 on SBI02 and Ma_4 on SBI10. Positional cloning and whole genome sequencing identified a candidate gene for Ma_2, Sobic.002G302700, which encodes a SET and MYND (SYMD) domain lysine methyltransferase. Nine sorghum genotypes previously identified as recessive for Ma_2 contained the mutated version of Sobic.002G302700 present in 80M (ma_2).

Introduction

Sorghum bicolor is a drought resilient, short-day C4 grass that is grown globally for grain, forage and biomass [1–4]. Precise control of flowering time is critical to achieve optimal yields of sorghum crops in specific target production locations/environments. Sorghum genotypes that have delayed flowering in long days due to high photoperiod sensitivity are high-yielding sources of biomass for production of biofuels and specialty bio-products [3,5]. In

58 contrast, grain sorghum was adapted for production in temperate regions by selecting genotypes
59 that have reduced photoperiod sensitivity resulting in earlier flowering and reduced risk of
60 exposure to drought, heat, or cold temperatures during the reproductive phase. A range of
61 flowering times are found among forage and sweet sorghums [6]. Sweet sorghum genotypes with
62 longer vegetative growth duration have larger stems that have greater potential for sucrose
63 accumulation [6–8].

64 Flowering time is regulated by development, day length, phytohormones, shading,
65 temperature, and the circadian clock [9–11]. In the long-day plant *Arabidopsis thaliana*,
66 circadian and light signals are integrated to increase the expression of *FLOWERING LOCUS T*
67 (*FT*) and flowering in long days. *FT* encodes a signaling protein synthesized in leaves that
68 moves through the phloem to the shoot apical meristem (SAM) where it interacts with
69 *FLOWERING LOCUS D* (*FD*) and reprograms the vegetative shoot apical meristem for
70 reproductive development [12,13]. Expression of circadian clock genes such as *LATE*
71 *ELONGATED HYPOCOTYL* (*LHY*) and *TIMING OF CAB1* (*TOC1*) regulate the expression of
72 the clock output gene *GIGANTEA* (*GI*) and genes in the flowering time pathway [14–16].
73 Photoperiod and circadian clock signals are integrated to control the expression and stability of
74 CONSTANS (CO) an activator of *FT* expression [17]. Under inductive long day (LD)
75 photoperiods, CO promotes the expression of *FT* which induces flowering in *Arabidopsis* [18].

76 Many of the genes in the *Arabidopsis* flowering time pathway are found in sorghum and
77 other grass species such as *Oryza sativa* (rice) [10] and maize [19], however, the regulation of
78 flowering time in these grasses has diverged from Arabidopsis in several important ways. Both
79 rice and sorghum are facultative short-day (SD) plants. In rice, the expression of the *FT-like* gene
80 *Heading date 3a* (*Hd3a*) is promoted in SD [20]. In sorghum, expression of two different *FT-*

like genes, *SbCN8* and *SbCN12,* is induced when plants are shifted from LD to SD [21,22]. In contrast to *Arabidopsis*, the rice and sorghum homologs of *CO* (rice *Heading date1, OsHd1; SbCO*) repress flowering in LD [10,23]. Rice and sorghum encode two additional grass-specific regulators of flowering Ehd1 and Ghd7. *Early heading date1* (Ehd1) activates the expression of *FT*-like genes, and *Grain number, plant height and heading date7* (Ghd7) represses the expression of *EHD1* and flowering [24,25]. When sorghum is grown in short days, SbEhd1 and SbCO induce the expression of *SbCN8* and *SbCN12*, leading to floral induction [21,22,26,27].

Under field conditions, time to flowering in sorghum varies from ~50 to >150 days after planting (DAP) depending on genotype, planting location and date (latitude/day-length), and the environment. A tall and "ultra-late" flowering sorghum variety called Milo Maize was introduced to the United States in the late 1800s [28]. Shorter and earlier flowering Milo genotypes such as Early White Milo and Dwarf Yellow Milo were selected from the introduced Milo genotype to promote improved grain yield in temperate regions of the US [1,28,29]. Genetic analysis determined that mutations in three independently segregating *Maturity* (*Ma*) loci (Ma_1, Ma_2, Ma_3) were responsible for early flowering times in the Milo genotypes. A cross between Early White Milo ($ma_1Ma_2Ma_3$) and Dwarf Yellow Milo ($Ma_1ma_2ma_3$) was used to construct a set of Milo maturity standards, a series of nearly isogenic lines that differ at one or more of the *Maturity* loci (Quinby and Karper 1945, Quinby 1966, Quinby, 1967). A fourth *Maturity* locus (Ma_4) was discovered in crosses of Milo (Ma_4) and Hegari (ma_4) [30]. More recent studies identified Ma_5 and Ma_6 [31]. Subsequent research showed that all of the Milos are dominant for Ma_5 and recessive for ma_6 [23,26]. In addition to these six *Ma* loci, many other flowering time quantitative trait loci (QTL) have been identified in sorghum [2,32–35].

Additional research has linked several of these QTL to genes such as *SbEHD1* and *SbCO* that are activators of *SbCN8* and *SbCN12* expression, sources of florigen in sorghum.

The genes corresponding to four of the six *Maturity* loci have been identified. Ma_1, the locus with the greatest influence on flowering time photoperiod sensitivity, encodes *SbPRR37*, a pseudo-response regulator that inhibits flowering in LD [21]. Ma_3 encodes phytochome B [36], Ma_5 encodes phytochrome C [23], and Ma_6 encodes *Ghd7* a repressor of flowering in long days [26]. The genes corresponding to Ma_2 and Ma_4 have not been identified but recessive alleles at either locus results in early flowering in long days in genotypes that are photoperiod sensitive (Ma_1) [28]**.** Prior studies also noted that genotypes recessive for Ma_2 flower later in genotypes that are photoperiod insensitive and recessive for Ma_1 and Ma_6 [28].

In this study, the impact of Ma_2 alleles on the expression of genes in the sorghum flowering time pathway was characterized. A QTL corresponding to Ma_2 was mapped and a candidate gene for Ma_2 identified by fine mapping and genome sequencing. The results show that Ma_2 enhances Ma_1 (*SbPRR37*) and *SbCO* expression consistent with the impact of Ma_2 alleles on flowering time in genotypes that vary in Ma_1 alleles.

Methods

Plant growing conditions and populations

The cross of 100M and 80M was carried out by the Sorghum Breeding Lab at Texas A&M University in College Station, TX. F_1 plants were grown in the field in Puerto Rico and self-pollinated to generate the F_2 population used in this study. The 100M/80M F_2 population

was planted in the spring of 2008 at the Texas A&M Agrilife Research Farm in Burleson County, Texas (near College Station, TX).

The cross of Hegari and 80M was made in the greenhouse at Texas A&M University in College Station, TX. F_1 plants were confirmed and self-pollinated to generate the F_2 population used in this study. The Hegari/80M F_2 population (n = 432) was planted in the spring of 2011 in the greenhouse in 18 L nursery pots in a 2:1 mixture of Coarse Vermiculite (SunGro Horticulture, Bellevue, WA) to brown pasture soil (American Stone and Turf, College Station, TX). All subsequent generations of Hegari/80M for fine mapping were grown in similar conditions. Greenhouse-grown plants were watered as needed and fertilized every two weeks using Peters general purpose 20-20-20 (Scotts Professional).

For circadian gene expression experiments, 100M and 80M genotypes were planted in MetroMix 900 (Sungro Agriculture) in 6 L pots, and thinned to 3 plants/pot after 2 weeks. Plants were grown in the greenhouse under 14 h days until 30 days after planting (DAP). After 30 days, the plants were moved into growth chambers and allowed to acclimate for 3 days. The growth chamber was set to 30°C and 14/10h L/D for the 3 days of entrainment and the first 24 h of tissue collection. The lights were changed to constant light for the second 24 h of tissue collection.

QTL mapping and multiple-QTL analysis

DNA was extracted from leaf tissue for all individuals described above as described in the FastDNA Spin Kit manual (MP Biomedicals). All individuals in each mapping or HIF population were genotyped by Digital Genotyping using FseI digestion enzyme as described in Morishige et al [37]. DNA fragments were sequenced using the Illumina GAII platform and the

147 reads were mapped back to the sorghum reference genome (v1.0, Phytozome v6). Genetic maps
148 were created using MapMaker 3.0B with the Kosambi function [38]. QTL were mapped using
149 WinQTLCartographer (v2.5.010) using composite interval mapping with a 1.0 cM walk speed
150 and forward and backward model selection [39]. The threshold was set using 1000 permutations
151 and $\alpha = 0.05$. Upon release of v3.1 of the sorghum reference genome, the QTL coordinates were
152 updated [40].

153 To look for possible gene interactions multiple-QTL analysis was used in the Hegari/80M
154 F_2 population. A single QTL analysis using the EM algorithm initially identified two primary
155 additive QTL which were used to seed model selection. The method of Manichaikul et al. [41]
156 was employed for model selection as implemented in R/qtl for multiple-QTL analysis [42].
157 Computational resources on the WSGI cluster at Texas A&M were used to calculate the
158 penalties for main effects, heavy interactions, and light interactions. These penalties were
159 calculated from 24,000 permutations for flowering time to find a significance level of 5% in the
160 context of a two-dimensional, two-genome scan.
161

162 Fine mapping of the *Ma_2* QTL

163 All fine mapping populations for the *Ma_2* QTL were derived from F_2 individuals from the
164 Hegari/80M population. The genetic distance spanning the *Ma_2* locus is 2 cM corresponding to a
165 physical distance of ~1.8 Mbp, so 1000 progeny would be required to obtain 20 recombinants
166 within the *Ma_2* QTL region. Six individuals that were heterozygous across the *Ma_2* QTL were
167 self-pollinated to generate six heterogeneous inbred families (HIFs) totaling 1000 F_3 individuals.
168 These individuals were grown out in the greenhouse, and flowering time was recorded. They
169 were genotyped by Digital Genotyping as described above [37]. Two F_3 individuals that had

useful breakpoints with a heterozygous genotype on one side of the breakpoint were grown and self-pollinated to generate an additional round of HIFs (F_4, n = 150) that were planted in the spring of 2013 and analyzed as described above. No new breakpoints were identified in the F_4 generation, so this process was repeated again to generate F_5 plants in the spring of 2014.

Circadian gene expression analysis

For the circadian gene expression analysis, 30-day-old plants were placed in a growth chamber set to 14 h days for the first 24 h and constant light for the second 24 h at 30°C. Plants were entrained for 3 d before beginning tissue collection. Leaf tissue was collected and pooled from 3 plants every 3 h for 48 h. The experiment was repeated three times for a total of three biological replicates. RNA was extracted from each sample using the Direct-Zol™ RNA Miniprep Kit (Zymo Research) according to the kit instructions. cDNA was synthesized using SuperScript III kit for qRT-PCR (Invitrogen) according to the kit instructions. Primers for sorghum flowering pathway genes were developed previously, and primer sequences are available in Murphy et al [21]. Primer sequences for *Ma_2* are available in S1 Table. Relative expression was determined using the comparative cycle threshold (C_t) method. Raw C_t values for each sample were normalized to C_t values for the reference gene *SbUBC* (Sobic.001G526600). Reference gene stability was determined previously [43]. $\Delta\Delta C_t$ values were calculated relative to the sample with the highest expression (lowest C_t value). Relative expression values were calculated with the $2^{-\Delta\Delta Ct}$ method [44]. Primer specificity was tested by dissociation curve analysis and gel electrophoresis of qRT-PCR products.

Ma_2 phylogenetic analysis

Protein sequences of the closest homologs of Ma_2 were identified using BLAST analysis. Protein sequences were aligned using MUSCLE [45] and visualized using Jalview [46]. Evolutionary trees were inferred using the Neighbor-Joining method [47] in MEGA7 [48]. All positions containing gaps and missing data were eliminated.

Ma_2 DNA sequencing and whole genome sequence analysis

Whole genome sequence reads of 52 sorghum genotypes including 100M and 80M were obtained from Phytozome v12. Base quality score recalibration, INDEL realignment, duplicate removal, joint variant calling, and variant quality score recalibration were performed using GATK v3.3 with the RIG workflow [49]. Sobic.002G302700 was sequenced via Sanger sequencing in the genotypes in Table 1 according to the BigDye Terminator Kit (Applied Biosystems). Primers for template amplification and sequencing are provided in S1 Table.

Results

Effects of *Ma_2* alleles on flowering pathway gene expression

The recessive *ma_2*-allele in 80M (*Ma_1**$\mathbf{ma_2}$**$Ma_3Ma_4Ma_5ma_6$*) was previously reported to cause 80M to flower earlier than100M (*Ma_1**$\mathbf{Ma_2}$**$Ma_3Ma_4Ma_5ma_6$*) in long days [28]. To help elucidate how *Ma_2* modifies flowering time, we investigated the impact of *Ma_2* alleles on the expression of genes in sorghum's flowering time pathway. Gene expression was analyzed by

qRT-PCR using RNA isolated from 100M (Ma_2) and 80M (ma_2) leaves collected every 3 hours for one 14h light/10h dark cycle and a second 24-hour period of constant light.

SbPRR37 is a central regulator of photoperiod sensitive flowering in sorghum that acts by repressing the expression of *SbCN* (*FT*-like) genes in LD [21]. *SbPRR37* expression in 100M and 80M grown in long days peaked in the morning and again in the evening as previously observed [21] (Fig 1). The amplitude of both peaks of *SbPRR37* expression was reduced in 80M (ma_2) compared to 100M (Ma_2) (Fig 1A). *SbCO* also shows peaks of expression in the morning (dawn) and in the evening (~14h) [21] (Fig 5C). Analysis of *SbCO* expression in 100M and 80M showed that both peaks of *SbCO* expression were reduced in 80M compared to 100M (Fig 1B).

Fig 1. Circadian expression of genes regulating flowering in *S. bicolor* in 100M and 80M under long days.
(A) Expression of *SbPRR37* in 100M (solid black lines) and 80M (dashed red lines). The expression peaks of *SbPRR37* are reduced in 80M. This is consistent with earlier flowering in 80M because *SbPRR37* represses the expression of the sorghum *FT*-like genes. (B) Expression of *SbCO* in 100M and 80M. Expression peaks of *SbCO* are also reduced in 80M. This is consistent with earlier flowering in 80M because under long days *SbCO* is a repressor of flowering. All expression values are normalized to *SbUBC* and are the mean of 3 biological replicates.

SbCN8, SbCN12, and *SbCN15* are homologs of *AtFT* that encode florigens in sorghum [22]. Expression of *SbCN8* and *SbCN12* increases when sorghum plants are shifted from LD to SD, whereas *SbCN15* shows minimal response to day length [21,26]. SbPRR37 and SbCO are co-repressor of the expression of *SbCN8* and *SbCN12* in long days, therefore, the influence of Ma_2 alleles on *SbCN8/12/15* expression was investigated [21,27]. When plants were grown in long days, expression of *SbCN12* was ~10 fold higher in 80M compared to 100M consistent with earlier flowering in 80M (Fig 2).

Fig 2. Expression of the *S. bicolor FT*-like genes *SbCN8*, *SbCN12*, and *SbCN15* in long days at the expected peak of expression.
Expression of *SbCN* genes are all elevated in 80M, which is consistent with earlier flowering in that genotype. All expression values are normalized to *SbUBC* and are the mean of 3 biological replicates. Fold change was calculated as $2^{-[Ct(100M)-Ct(80M)]}$.

Previous studies showed that SbGHD7 represses *SbEHD1* expression and that alleles of *SbGHD7* differentially affect *SbCN8* expression (>SbCN12) [26]. Analysis of *SbEHD1* and *SbGHD7* expression in 100M and 80M showed that Ma_2 alleles modify the expression of these genes only to a small extent (S1 Fig).

The timing of the two daily peaks of *SbPRR37* and *SbCO* expression in sorghum is regulated by the circadian clock [21,26]. Therefore, it was possible that *Ma2* modifies *SbPRR37/SbCO* expression by altering clock gene expression. However, expression of the clock genes *TOC1* and *LHY* was similar in 100M and 80M (S1 Fig). Taken together, these results show that Ma_2 is an activator of *SbPRR37* and *SbCO* expression in long days. Prior studies showed that co-expression of *SbPRR37* and *SbCO* in long days inhibits expression of *SbCN12* and floral initiation [27]. Later flowering in sorghum genotypes that are Ma_1Ma_2 vs. Ma_1ma_2 in long days is consistent with lower *SbCN12* expression in Ma_1Ma_2 genotypes.

Genetic analysis of Ma2 and Ma4

An F_2 population derived from a cross of 100M (Ma_2) and 80M (ma_2) was generated to map the Ma_2 locus. Because 100M and 80M are nearly isogenic lines that differ at Ma_2, only Ma_2 alleles were expected to affect flowering time in this population [28]. The F_2 population (n = ~1100) segregated for flowering time in a 3:1 ratio as expected. The parental lines and F_2 individuals were genotyped by Digital Genotyping (DG) which identifies single nucleotide polymorphism (SNP) markers in thousands of sequenced sites that distinguish the parents of a

population [37]. The near isogenic nature of the parental lines resulted in a very sparse genetic map that lacked coverage of large regions of the sorghum genome including all of the long arm of SBI02. In retrospect, no *Ma₂* QTL for flowering time was identified using this genetic map because the gene is located on the long arm of SBI02 (see below).

To overcome the limitations associated with the 80M/100M population, a second mapping population was created to identify the genetic locus associated with *Ma₂*. An F_2 population (n = 215) that would segregate for *Ma2* and *Ma4* was constructed by crossing Hegari (*Ma₁**Ma₂**Ma₃**ma₄**Ma₅ma₆*) and 80M (*Ma₁**ma₂**Ma₃**Ma₄**Ma₅ma₆*) [30,50]. The population was grown in a greenhouse under long day conditions and phenotyped for days to flowering. QTL for flowering time were identified on SBI02 and SBI10 (Fig 3). Recessive alleles of *Ma2* and *Ma4* result in earlier flowering when plants are grown in long days. The Hegari haplotype across the QTL on SBI10 was associated with early flowering therefore this QTL corresponds to *Ma₄* (S3 Fig). The 80M haplotype across the QTL on SBI02 was associated with early flowering therefore the QTL on SBI02 corresponds to *Ma₂*.

Fig 3. Quantitative trail locus (QTL) map of flowering time in the Hegari/80M F_2 population.
Two QTL were identified for variation in flowering time in the F_2 population derived from Hegari (*Ma₁Ma₂Ma₃ma₄*) and 80M (*Ma₁ma₂Ma₃Ma₄*). This population was expected to segregate for *Ma₂* and *Ma₄*. Each recessive *Ma* allele causes earlier flowering. The QTL on LG10 corresponds to *Ma₄* because F_2 individuals carrying the Hegari allele contributed to accelerated flowering. F_2 individuals carrying the 80M allele at the QTL on LG02 flowered earlier, so this QTL corresponds to *Ma₂*.

Epistatic interactions between *Ma₂* and *Ma₄*

Previous studies indicated an epistatic interaction exists between *Ma₂* and *Ma₄* [28]. Therefore, Multiple QTL Mapping (MQM) analysis [51] was employed, using data from the

Hegari/80M F_2 population, to identify additional flowering time QTL and interactions amongst the QTL as previously described [52]. MQM analysis identified the QTL for flowering time on SBI02 and SBI10 and an additional QTL on SBI09. Additionally, an epistatic interaction was identified between *Ma₂* and *Ma₄* (pLOD = 42). Interaction plots showed that in a dominant *Ma₄* background, a dominant allele at *Ma₂* delays flowering, while in a recessive *Ma₄* background, *Ma₂* has a minimal impact on flowering time (Fig 4). The interaction between *Ma₂* and *Ma₄* identified by MQM analysis is consistent previous observations that in a recessive *ma₄* background flowering is early regardless of allelic variation in *Ma₂* [28].

Fig 4. Interaction plots for the Ma2 QTL and the Ma4 QTL
There is a known interaction between *Ma₂* (represented by marker c2_68327634) and *Ma₄* (represented by marker c10_3607821). This interaction was identified by multiple QTL mapping (MQM). Dominant alleles of the *Ma* genes delay flowering. In a recessive ma4 background (AA at c2_68327634), the effect of *Ma₂* on days to flowering is reduced. A represents the 80M allele and B represents the Hegari allele at each QTL. Reciprocal plots are shown.

Ma2 candidate gene identification

The Hegari/80M F_2 population located *Ma₂* on SBI02 between 67.3 Mbp to 69.1 Mbp (Fig 5). To further delimit the *Ma₂* locus, six lines from the Hegari/80M population that were heterozygous across the *Ma₂* QTL but fixed across the *Ma₄* locus (*Ma₄Ma₄*) were selfed to create heterogeneous inbred families (HIFs) (n=1000 F_3 plants) [53]. Analysis of these HIFs narrowed the region encoding *Ma₂* to ~600 kb (67.72 Mb-68.33 Mb) (Fig 5). Genotypes that were still heterozygous across the delimited locus were selfed and 100 F_4 plants were evaluated for differences in flowering time. This process narrowed the *Ma₂* locus to a region spanning ~500 kb containing 76 genes (67.72Mb-68.22Mb) (Fig 5, S2 Table).

Fig 5. Fine-mapping of the Ma2 QTL.
The *Ma_2* QTL spans from 67.3 Mpb to 69.1 Mbp (light blue bar). Five F_2 individuals that were heterozygous across the *Ma_2* QTL were self-pollinated to generate heterogeneous inbred families (HIFs) totaling 1000 F_3 individuals. Genotype and phenotype analysis of these HIFs narrowed the QTL region to ~600 kb (darker blue bar). Two additional rounds of fine-mapping narrowed the QTL region to ~500 kb (vertical dashed lines). This region contained 76 genes. The genotypes of relevant HIFs and the parents are shown to the left and their corresponding days to flowering are shown to the right. Blue regions correspond to the 80M genotype and red regions correspond to the Hegari genotype. Purple regions are heterozygous.

The low rate of recombination across the *Ma_2* locus led us to utilize whole genome sequencing in conjunction with fine mapping to identify a candidate gene for *Ma_2*. Since 100M and 80M are near isogenic lines that have very few sequence differences along the long arm of SBI02 where the *Ma_2* QTL is located, whole genome sequences (WGS) of 100M and 80M were generated in collaboration with JGI (sequences available at www.phytozome.jgi.doe.gov). The genome sequences were scanned for polymorphisms within the 500 kb locus spanning *Ma_2*. Only one T → A single nucleotide polymorphism (SNP) located in Sobic.002G302700 was identified that distinguished 100M and 80M within the region spanning the *Ma_2* locus. The T → A mutation causes a Lys141* change in the third exon, resulting a truncated protein. A 500 bp DNA sequence spanning the T to A polymorphism in Sobic.002G302700 was sequenced from 80M and 100M to confirm the SNP identified by comparison of the whole genome sequences (Table 1). The T → A point mutation was present in 80M (*ma_2*) whereas 100M (*Ma_2*) encoded a functional version of Sobic.002G302700 that encodes a full length protein. Since this mutation was the only sequence variant between 100M and 80M in the fine-mapped locus, Sobic.002G302700 was identified as the best candidate gene for *Ma_2*.

Sobic.002G302700 is annotated as a SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) and MYND (Myeloid-Nervy-DEAF1) (SMYD) domain-containing protein. SMYD domain family proteins in humans have been found to methylate histone lysines and non-histone

346 targets and have roles in regulating chromatin state, transcription, signal transduction, and cell
347 cycling [54,55]. The SET domain in SMYD-containing proteins is composed of two sub-
348 domains that are divided by the MYND zinc-finger domain. The SET domain includes conserved
349 sequences involved in methyltransferase activity including nine cysteine residues that are present
350 in the protein encoded by Sobic.002G303700 (Fig 6) [56]. The MYND domain is involved in
351 binding DNA and is enriched in cysteine and histidine residues [57]. Protein sequence alignment
352 of Sobic.002G302700 homologs revealed that the SYMD protein candidate for Ma2 is highly
353 conserved across flowering plants (Fig 6).
354

355 **Fig 6. Alignment of Sobic.002G302700 with its closest homologs in several plant species.**
356 Sobic.002G302700 is highly conserved across plant species. It is annotated as a Set and MYND
357 (SMYD) protein. SMYD proteins have lysine methyltransferase activity. The MYND region is
358 highlighted in red. The nine conserved Cys residues typical of SMYD proteins are indicated by
359 asterisks.
360

361 To learn more about *Ma2* regulation, the expression of Sobic.002G302700 in 100M and
362 80M was characterized during a 48h L:D/L:L cycle. Ma_2 showed a small increase in expression
363 from morning to evening and somewhat higher expression in 100M compared to 80M during the
364 evening (S1 Fig).
365

366 Distribution of Ma_2 alleles in the sorghum germplasm

367 Recessive ma_2 was originally found in the Milo background and used to construct Double
368 Dwarf Yellow Milo (Ma_1***ma*****$_2$**$ma_3Ma_4Ma_5ma_6$) [28]. Double Dwarf Yellow Milo was crossed to
369 Early White Milo (ma_1***Ma*****$_2$**$Ma_3Ma_4Ma_5ma_6$) and the progeny selected to create 100M, 80M and
370 the other Milo maturity standards [1,28,58]. Several of the Milo maturity standards were
371 recorded as recessive Ma_2 (80M, 60M, SM80, SM60, 44M, 38M) and others as Ma_2 dominant

372 (100M, 90M, SM100, SM90, 52/58M). In order to confirm the *Ma_2* genotype of the maturity
373 standards, the 500 bp sequence spanning the Lys141* mutation in Sobic.002G302700 was
374 obtained from most of these genotypes (Table 1). Kalo was also identified as carrying a recessive
375 allele of *Ma_2*. Kalo was derived from a cross of Dwarf Yellow Milo (*ma_2*), Pink Kafir (*Ma_2*),
376 and CI432 (*Ma_2*), therefore it was concluded that DYM is the likely source of recessive *ma_2* [28].
377 Sequence analysis showed that the genotypes previously identified as *ma_2* including Kalo, 80M,
378 SM80, 60M, 44M, 38M, and 58M carry the recessive mutation in Sobic.002G302700 identified
379 in 80M. 100M, SM100, and Hegari that were identified as *Ma_2*, did not contain the mutated
380 version of Sobic.002G302700 (Table 1). Additionally, sequences of *Ma_2* from 52 sorghum
381 genotypes with publicly available genome sequences were compared [40]. Sobic.002G302700
382 was predicted to encode functional proteins in all except one of these sorghum genotypes. A
383 possible second recessive *Ma_2* allele was found in IS3614-2 corresponding to an M83T missense
384 mutation that was predicted to be deleterious by PROVEAN [59].

385

386 **Table 1. Sequence variants of Sobic.002G203700 and their predicted effect on protein**
387 **function**

Genotype	Historical *Ma_2* allele	Sequence variant	Effect on protein function
100M	Ma_2	-	-
SM100	Ma_2	-	-
SM90	Ma_2	-	-
Hegari	Ma_2	-	-
80M	ma_2	L141*	Deleterious
SM80	ma_2	L141*	Deleterious
60M	ma_2	L141*	Deleterious
44M	ma_2	L141*	Deleterious
38M	ma_2	L141*	Deleterious
58M	ma_2	L141*	Deleterious
Kalo	ma_2	L141*	Deleterious
IS3614-2	-	M83T	Deleterious

*Sequenced by Sanger sequencing

388
389

Discussion

Sorghum is a facultative short day plant. In photoperiod sensitive sorghum genotypes, following the vegetative juvenile phase, day length has the greatest impact on flowering time under normal growing conditions. The development of early flowering grain sorghum adapted to temperate regions of the US was based on the selection of mutations in numerous genes that reduced photoperiod sensitivity. Genetic analysis of the loci and genes containing these mutations beginning in the 1940's [50,58] identified six *Maturity* loci (Ma_1-Ma_6) that resulted in earlier flowering time when plants were grown in long days. Recessive alleles at each of the six *Ma* loci reduces photoperiod sensitivity [30,31,58]. Molecular identification of the genes corresponding to Ma_1, Ma_3, Ma_5 and Ma_6 and other genes in the sorghum flowering time pathway (i.e., *SbCO, SbEHD1, SbCN8/12*) and an understanding of their regulation by photoperiod and the circadian clock led to the model of the flowering time pathway shown in Figure 7 [60]. The current study showed that Ma2 represses flowering in long days by increasing the expression of SbPRR37 (Ma_1) and *SbCO*. The study also located QTL for Ma_2 and Ma_4, confirmed an epistatic interaction between Ma_2 and Ma_4, and identified a candidate gene for Ma_2.

Fig 7. A model of the flowering time regulatory pathway in *S. bicolor*.
Ma_2 and Ma_4 work codependently to enhance the expression of *SbPRR37* and *SbCO*. In LD, SbPRR37 and SbCO in turn repress the expression of the *SbCN* genes, especially *SbCN12*, to repress the floral transition.

The recessive ma_2 allele characterized in this study arose in a highly photoperiod sensitive Milo genotype that was introduced into the US in the late 1800's and then selected for early flowering to enhance grain production. Quinby and Karper [29] created near isogenic Milo

maturity genotypes with allelic variation at specific *Ma* loci to facilitate genetic and physiological analysis of flowering time regulation. In the current study, we utilized two of these maturity genotypes, 100M ($Ma_1\mathbf{Ma_2}Ma_3Ma_4Ma_5ma_6$) and 80M ($Ma_1\mathbf{ma_2}Ma_3Ma_4Ma_5ma_6$), to characterize how allelic variation in Ma_2 affects the expression of genes in the sorghum photoperiod regulated flowering time pathway (Fig 7). This analysis showed that mutation of ma_2 (80M) significantly reduced the amplitude of the morning and evening peaks of *SbPRR37* and *SbCO* expression compared to 100M (Ma_2) without altering the timing of their expression. In addition, the expression of *SbCN12* (*FT*-like) increased 8-fold in leaves of 80M compared to 100M consistent with earlier flowering in 80M. In contrast, expression of clock genes (*TOC1, LHY*) and other genes (i.e., GHD7, EHD1) in the photoperiod regulated flowering time pathway were modified to only a small extent by allelic variation in Ma_2. Based on these results, we tentatively place Ma_2 in the flowering time pathway downstream of day length sensing phytochromes and circadian clock regulation and identify Ma_2 as a factor that enhances *SbPRR37* and *SbCO* expression (Fig 7).

The differential increase in *SbCN12* expression in 80M (vs. 100M) is consistent with differential inhibition of *SbCN12* expression in long days by the concerted action of SbPRR37 and SbCO which has been previously shown to inhibit *SbCN12* expression [27]. Prior studies showed that 100M (Ma_2) flowers later than 80M (ma_2) in long days [28]. The impact of Ma_2 alleles on the expression of *SbPRR37* and *SbCO* is consistent with the effect of these alleles on flowering time in long days. Genetic studies showed that floral repression mediated by SbPRR37 requires SbCO as a co-repressor [27]. Therefore, enhanced expression of both *SbPRR37* (Ma_1) and *SbCO* by Ma_2 in Ma_1Ma_2 genotypes in long days is consistent with delayed flowering under these conditions relative to genotypes such as 80M that are Ma_1ma_2. Molecular genetic studies

438 also showed that SbCO is an activator of *SbCN12* expression and flowering in ma_1 genetic
439 backgrounds [27]. This is consistent with the observation that ma_1Ma_2 genotypes flower earlier
440 than ma_1ma_2 genotypes when grown in long days [28].
441

442 Interactions between Ma_2 and Ma_4

443 Multiple QTL (MQM) analysis of results from a population derived from Hegari/80M
444 identified an interaction between Ma_2 and Ma_4 as well as one additional flowering QTL on
445 SBI09. Flowering time QTL on SBI09 have been identified in other mapping populations, but
446 the gene(s) involved have not been identified [33,34]. The interaction between Ma_2 and Ma_4
447 confirmed previous observations that recessive ma_4 causes accelerated flowering in long days in
448 Ma_1Ma_2 genotypes [28]. Interestingly, the influence of Ma_2 and Ma_4 alleles on flowering time is
449 affected by temperature [28,61]. The influence of temperature on flowering time pathway gene
450 expression in 80M and 100M in the current study was minimized by growing plants at constant
451 30C. However, analysis of the temperature dependence of Ma2 and Ma4 on flowering time may
452 help elucidate interactions between photoperiod and flowering time that have been previously
453 documented [28,62]. Positional cloning of Ma_4 is underway to better understand the molecular
454 basis of Ma_2 and Ma_4 interaction and their impact on flowering time.
455

456 Identification of a candidate gene for Ma_2

457 A mapping population derived from Hegari/80M that segregated for Ma_2 and Ma_4
458 enabled localization of the corresponding flowering time QTL in the sorghum genome (SBI02,
459 Ma_2; SBI10, Ma_4). The Ma_2 QTL on SBI02 was fine-mapped using heterozygous inbred families

(HIFs) from Hegari/80M. Several rounds of fine-mapping delimited the QTL to a ~500kb region containing 76 genes. Low recombination rates in this region of SBI02 made it difficult to delimit the QTL further using break point analysis therefore comparison of genome sequences from 80M and 100M was used to help identify a candidate gene for *Ma2*. The recessive ma_2 allele present in 80M arose in a Milo genotype similar to 100M [28] and genetic analysis of 100M and 80M showed that these near isogenic genotypes lacked DNA markers on the long arm of SBI02 where Ma_2 is located. Indeed, a scan of the whole genome sequences of 100M and 80M identified only a single T to A mutation in the 500 kb region spanning the fine-mapped Ma_2 locus. This mutation caused a Lys141* change in the third exon of Sobic.002G302700 resulting in protein truncation. Based on this information Sobic.002G302700 was tentatively identified as the best candidate gene for Ma_2.

Sobic.002G302700 encodes a SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) and MYND (Myeloid-Nervy-DEAF1) (SMYD) domain containing protein. In humans, SMYD proteins act as lysine methyltransferases, and the SET domain is critical to this activity. Therefore, Ma_2 could be altering the expression of *SbPRR37* and *SbCO* by modifying histones associated with these genes. The identification of this SMYD family protein's involvement in flowering in sorghum as well as the identification of highly conserved homologs in other plant species suggests that *Ma2* may correspond to a novel regulator of sorghum flowering. While a role for SYMD-proteins (lysine methyltransferases) as regulators of flowering time has not been previously reported, genes encoding histone lysine demethylases (i.e., JMJ30/32) have been found to regulate temperature modulated flowering time in Arabidopsis [63].

J.R. Quinby [50] identified only one recessive allele of *Ma_2* among the sorghum genotypes used in the Texas sorghum breeding program. The maturity standard lines including 80M that are recessive for *ma_2* and the genotype Kalo were reported to be derived from the same recessive *ma_2* Milo genotype [28]. To confirm this, *Ma_2* alleles in the relevant maturity standards and Kalo were sequenced confirming that all of these *ma_2* genotypes carried the same mutation identified in 80M (Table 1). Among the 52 sorghum genotypes with available whole genome sequences, only 80M carried the mutation in Ma2 [40]. One possible additional allele of *ma_2* was identified in IS36214-2, which contained a M83T missense mutation that was predicted to be deleterious to protein function by PROVEAN [59].

In conclusion, we have shown that *Ma_2* represses flowering in long days by promoting the expression of the floral repressor *SbPRR37* and *SbCO*, a gene that acts as a co-repressor in long days (Fig 7). Sobic.002G302700 was identified as the best candidate for the sorghum *Maturity* locus *Ma_2* although further validation such as targeted mutation of Sobic.002G302700 in a *Ma_1Ma_2* sorghum genotype or complementation of *Ma_1ma_2* genotypes will be required to confirm this gene assignment. The identification of this gene and its interaction with *Ma_4* begins to elucidate a new element of the photoperiod flowering regulation pathway in sorghum.

Acknowledgements

This research was supported by the Perry Adkisson Chair in Agricultural Biology and by the Agriculture and Food Research Initiative Competitive Grant 2016-67013-24617 from the USDA National Institute of Food and Agriculture. The authors would like to thank Dr. Daryl Morishige for his assistance in performing the Hegari/80M cross as well as Robin Poncik for her assistance in recording flowering dates.

References

1. Quinby JR. The Genetic Control of Flowering and Growth in Sorghum. Adv Agron. 1974;25:125–62.
2. Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, et al. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun. 2013;4:337–42.
3. Mullet J, Morishige D, McCormick R, Truong S, Hilley J, McKinley B, et al. Energy sorghum--a genetic model for the design of C4 grass bioenergy crops. J Exp Bot. 2014;65(13):3479–89.
4. Boyles RE, Brenton ZW, Kresovich S. Genetic and Genomic Resources of Sorghum to Connect Genotype with Phenotype in Contrasting Environments. Plant J. 2018;19–39.
5. Olson SN, Ritter K, Rooney W, Kemanian A, McCarl BA, Zhang Y, et al. High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. Biofuels, Bioprod Biorefining. 2012;6(3):246–56.
6. Burks PS, Felderhoff TJ, Viator HP, Rooney WL. The influence of hybrid maturity and planting date on sweet sorghum productivity during a harvest season. Agron J. 2013;105(1):263–7.
7. Teetor VH, Duclos D V., Wittenberg ET, Young KM, Chawhuaymak J, Riley MR, et al. Effects of planting date on sugar and ethanol yield of sweet sorghum grown in Arizona. Ind Crops Prod. 2011;34(2):1293–300.
8. Rooney WL, Blumenthal J, Bean B, Mullet JE. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioprod Biorefining. 2007;1:147–57.
9. Song YH, Ito S, Imaizumi T. Flowering time regulation: Photoperiod- and temperature-sensing in leaves. Trends Plant Sci. Elsevier Ltd; 2013;18(10):575–83.
10. Tsuji H, Taoka KI, Shimamoto K. Florigen in rice: Complex gene network for florigen transcription, florigen activation complex, and multiple functions. Curr Opin Plant Biol. Elsevier Ltd; 2013;16:228–35.
11. Sanchez SE, Kay SA. The plant circadian clock: from a simple timekeeper to a complex developmental manager. Cold Spring Harb Perspect Biol. 2016;a027748.
12. Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, et al. Long-Distance Signaling in Floral Induction of Arabidopsis. Science (80-). 2007;316:1030–3.
13. Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, et al. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science (80-). 2005;309:1052–6.
14. Schaffer R, Ramsay N, Samach A, Corden S, Putterill J, Carré IA, et al. The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell. 1998;93(7):1219–29.
15. Millar AJ, Carre IA, Strayer CA, Chua NH, Kay SA. Circadian clock mutants in Arabidopsis identified by luciferase imaging. Science (80-). 1995;267(5201):1161 LP-1163.
16. Park DH, Somers DE, Kim YS, Choy YH, Lim HK, Soh MS, et al. Control of circadian rhythms and photoperiodic flowering by the Arabidopsis GIGANTEA gene. Science (80-). 1999;285(5433):1579–82.
17. Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G. CONSTANS mediates between the circadian clock and the control of flowering in

Arabidopsis. Nature. 2001;410(6832):1116.
18. Turnbull C. Long-distance regulation of flowering time. J Exp Bot. 2011;62(13):4399–413.
19. Dong Z, Danilevskaya O, Abadie T, Messina C, Coles N, Cooper M. A gene regulatory network model for Floral transition of the shoot apex in maize and its dynamic modeling. PLoS One. 2012;7(8).
20. Tamaki S, Matsuo S, Wong HL, Yokoi S, Shimamoto K. Hd3a Protein Is a Mobile Flowering Signal in Rice. Science (80-). 2007;316:1033–6.
21. Murphy RL, Klein RR, Morishige DT, Brady JA, Rooney WL, Miller FR, et al. Coincident light and clock regulation of controls photoperiodic flowering in sorghum. Proc Natl Acad Sci U S A. 2011;37:1–6.
22. Wolabu TW, Zhang F, Niu L, Kalve S, Bhatnagar-Mathur P, Muszynski MG, et al. Three FLOWERING LOCUS T-like genes function as potential florigens and mediate photoperiod response in sorghum. New Phytol. 2016;210:946–59.
23. Yang S, Murphy RL, Morishige DT, Klein PE, Rooney WL, Mullet JE. Sorghum Phytochrome B Inhibits Flowering in Long Days by Activating Expression of SbPRR37 and SbGHD7, Repressors of SbEHD1, SbCN8 and SbCN12. PLoS One. 2014;9(8):e105352.
24. Cho L-H, Yoon J, Pasriga R, An G. Homodimerization of Ehd1 is required to induce flowering in rice. Plant Physiol. 2016;170:2159–71.
25. Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet. 2008;40:761.
26. Murphy RL, Morishige DT, Brady J a., Rooney WL, Yang S, Klein PE, et al. Ghd7 (Ma6) Represses Sorghum Flowering in Long Days: Alleles Enhance Biomass Accumulation and Grain Production. Plant Genome. 2014;7(2):1–10.
27. Yang S, Weers BD, Morishige DT, Mullet JE. CONSTANS is a photoperiod regulated activator of flowering in sorghum. BMC Plant Biol. 2014;14:148.
28. Quinby JR. The Maturity Genes of Sorghum. Adv Agron. 1967;19:267–305.
29. Karper RE, Quinby JR. The history and evolution of milo in the United States. Agron J. 1946;38(5):441–53.
30. Quinby JR. Fourth maturity gene locus in sorghum. Crop Sci. 1966;(6):516–8.
31. Rooney WL, Aydin S. Genetic Control of a Photoperiod-Sensitive Response in Sorghum bicolor (L.) Moench. Crop Sci. 1999;39:397–400.
32. Hart GE, Schertz KF, Peng Y, Syed NH. Genetic mapping of Sorghum bicolor (L.) Moench QTLs that control variation in tillering and other morphological characters. Theor Appl Genet. 2001;103(8):1232–42.
33. Higgins RH, Thurber CS, Assaranurak I, Brown PJ. Multiparental mapping of plant height and flowering time QTL in partially isogenic sorghum families. G3. 2014;4:1593–602.
34. Zhao J, Mantilla Perez MB, Hu J, Salas Fernandez MG. Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum. Plant Genome. 2016;9(2):1–14.
35. Lin Y, Schertz KF, Patemon AH. Comparative Analysis of QTLs Affecting Plant Height and Maturity Across the Poaceae,. 1995. 1995;141:391–411.
36. Childs KL, Miller FR, Cordonnier-Pratt MM, Pratt LH, Morgan PW, Mullet JE. The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B. Plant Physiol. 1997;113:611–9.
37. Morishige DT, Klein PE, Hilley JL, Sahraeian SME, Sharma A, Mullet JE. Digital

genotyping of sorghum - a diverse plant species with a large repeat-rich genome. BMC Genomics. 2013;14:1–19.

38. Lander Eric S., Green Philip, Abrahamson Jeff, Barlow Aaron, Daly Mark J., Lincoln Stephen E., et al. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics. 1987;1(2):174–81.
39. Wang S, Basten CJ, Zeng ZB. Windows QTL Cartographer 2.5. Raleigh, NC: Department of Statistics, North Carolina State University; 2012.
40. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, et al. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. Plant J. 2018;93(2):338–54.
41. Manichaikul A, Moon JY, Sen Ś, Yandell BS, Broman KW. A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics. 2009;181(3):1077–86.
42. Broman KW, Wu H, Sen Ś, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003;19(7):889–90.
43. Casto AL, McKinley BA, Yu KMJ, Rooney WL, Mullet JE. Sorghum stem aerenchyma formation is regulated by SbNAC_D during internode development. Plant Direct. 2018;2(11):e00085.
44. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-ΔΔCt Method. Methods. 2001;25(4):402–8.
45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. England; 2004;32(5):1792–7.
46. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–91.
47. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.
48. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870–4.
49. McCormick RF, Truong SK, Mullet JE. RIG: Recalibration and Interrelation of Genomic Sequence Data with the GATK. G3 (Bethesda). 2015;5(4):655–65.
50. Quinby JR, Karper R. The inheritance of three genes that influence time of floral initiation and maturity date in Milo. J Am Soc Agron. 1945;(901):916–36.
51. Arends D, Prins P, Jansen RC, Broman KW. R/qtl: high-throughput multiple QTL mapping. Bioinformatics. 2010;26(23):2990–2.
52. Truong SK, McCormick RF, Rooney WL, Mullet JE. Harnessing genetic variation in leaf angle to increase productivity of sorghum bicolor. Genetics. 2015;201(3):1229–38.
53. Tuinstra MR, Ejeta G, Goldsbrough PB. Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. Theor Appl Genet. 1997;95(5–6):1005–11.
54. Xu S, Zhong C, Zhang T, Ding J. Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins. J Mol Cell Biol. 2011;3(5):293–300.
55. Spellmon N, Holcomb J, Trescott L, Sirinupong N, Yang Z. Structure and function of SET and MYND domain-containing proteins. Int J Mol Sci. 2015;16(1):1406–28.

644 56. Min J, Zhang X, Cheng X, Grewal SIS, Xu RM. Structure of the SET domain histone
645 lysine methyltransferase Clr4. Nat Struct Biol. 2002;9(11):828–32.
646 57. Gross CT, McGinnis W. DEAF-1, a novel protein that binds an essential region in a
647 Deformed response element. EMBO J. 1996;15(8):1961–70.
648 58. Quinby JR, Karper RE. Effect of Different Alleles on the Growth of Sorghum Hybrids.
649 1948;255–9.
650 59. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of
651 Amino Acid Substitutions and Indels. PLoS One. 2012;7(10):e46688.
652 60. Mullet JE. High-Biomass C4 Grasses − Filling the Yield Gap. Plant Sci. Elsevier;
653 2017;261:10–7.
654 61. Major DJ, Rood SB, Miller FR. Temperature and Photoperiod Effects Mediated by the
655 Sorghum Maturity Genes. Crop Sci. 1990;30(2):305–10.
656 62. Tarumoto I. Thermo-sensitivity and photoperiod sensitivity genes controlling heading
657 time and flower bud initiation in Sorghum, Sorghum bicolor Moench. Japan Agric Res Q.
658 2011;45(1):69–76.
659 63. Gan ES, Xu Y, Wong JY, Geraldine Goh J, Sun B, Wee WY, et al. Jumonji demethylases
660 moderate precocious flowering at elevated temperature via regulation of FLC in
661 Arabidopsis. Nat Commun. 2014;5:1–13.
662

Supporting information

S1 Fig. Circadian expression of Sobic.002G302700 in 100M and 80M
The expression of Sobic.002G302700 does not cycle diurnally in 100M (solid black line) or 80M (dashed red line). There was no difference in expression between 100M and 80M in the first day. Expression was slightly elevated in 100M compared to 80M during the night and through the following morning.

S2 Fig. Circadian expression of *SbTOC1, SbLHY, SbGhd7,* and *SbEhd1*
There were no consistent differences in expression of (A) *SbTOC1*, (B) *SbLHY*, (C) *SbGhd7*, and (D) *SbEhd1* between 100M (solid black line) and 80M (dashed red line).

S3 Fig. Genotype x phenotype plots for the QTL on SBI02 and SBI10
Recessive alleles of *Maturity* genes contribute to earlier flowering. 80M (AA) is recessive for ma_2, while Hegari (BB) is dominant. Individuals genotyped AA for the QTL on SBI02 (represented by marker c2_68327634) flowered ~100 d earlier than those genotyped BB. 80M is dominant for Ma_4, and individuals genotyped AA at the QTL on SBI10 (represented by marker c10_3607821) flowered ~100 d earlier than those genotyped BB.

S1 Table. Ma_2 (Sobic.002G302700) sequencing and qPCR primers

S2 Table. Genes in the fine-mapped *Ma2* QTL region

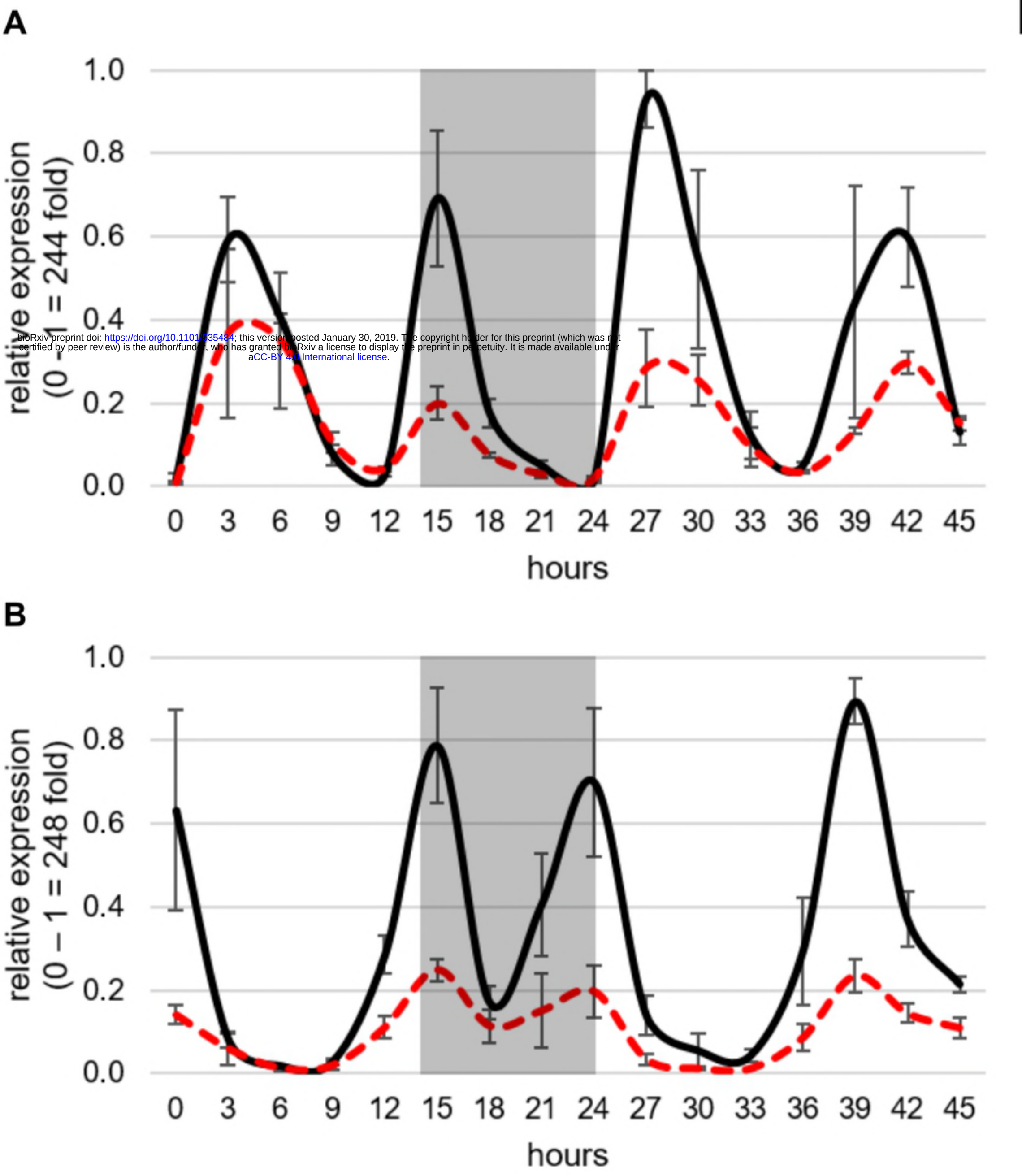


Figure 1 `

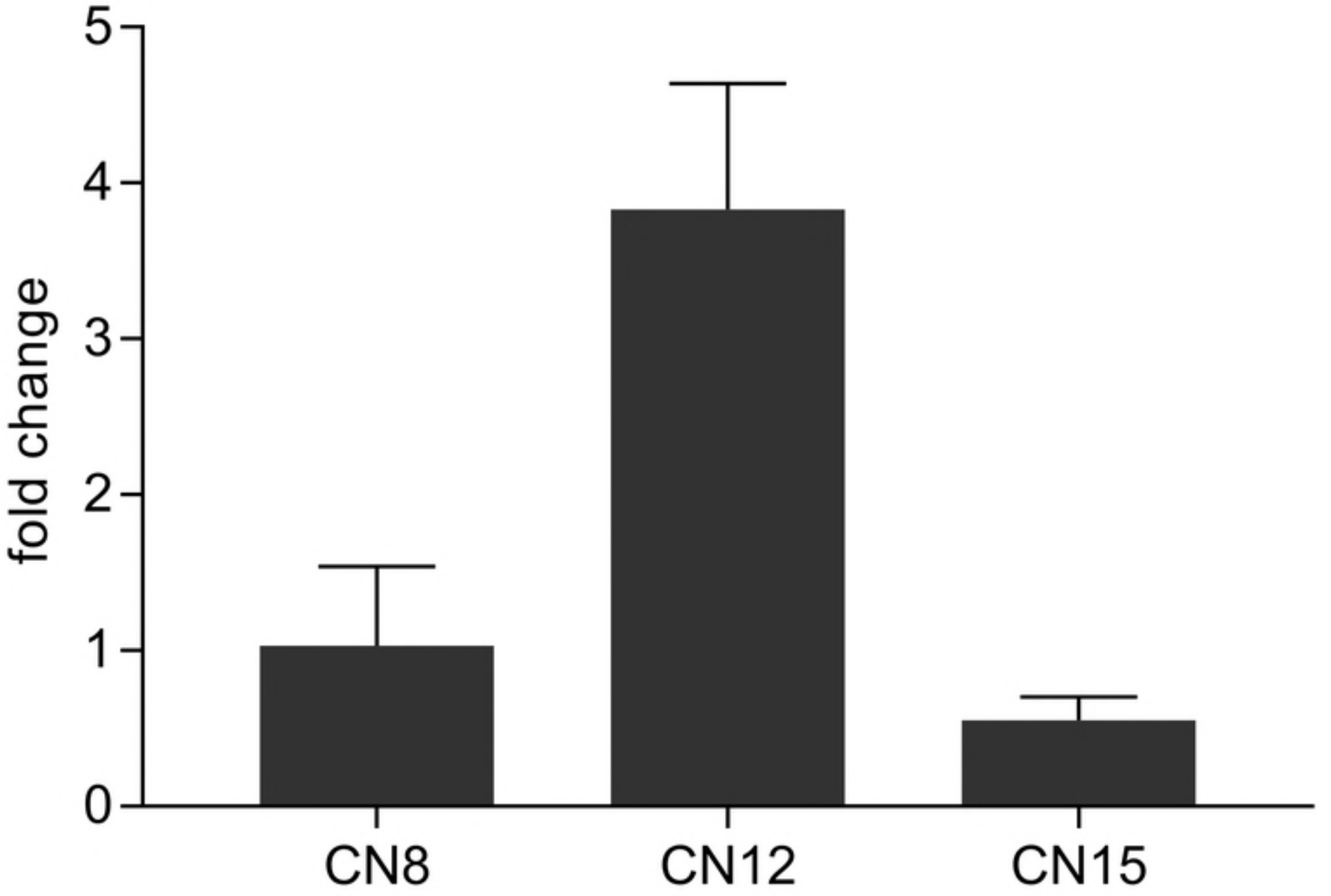


Figure 2

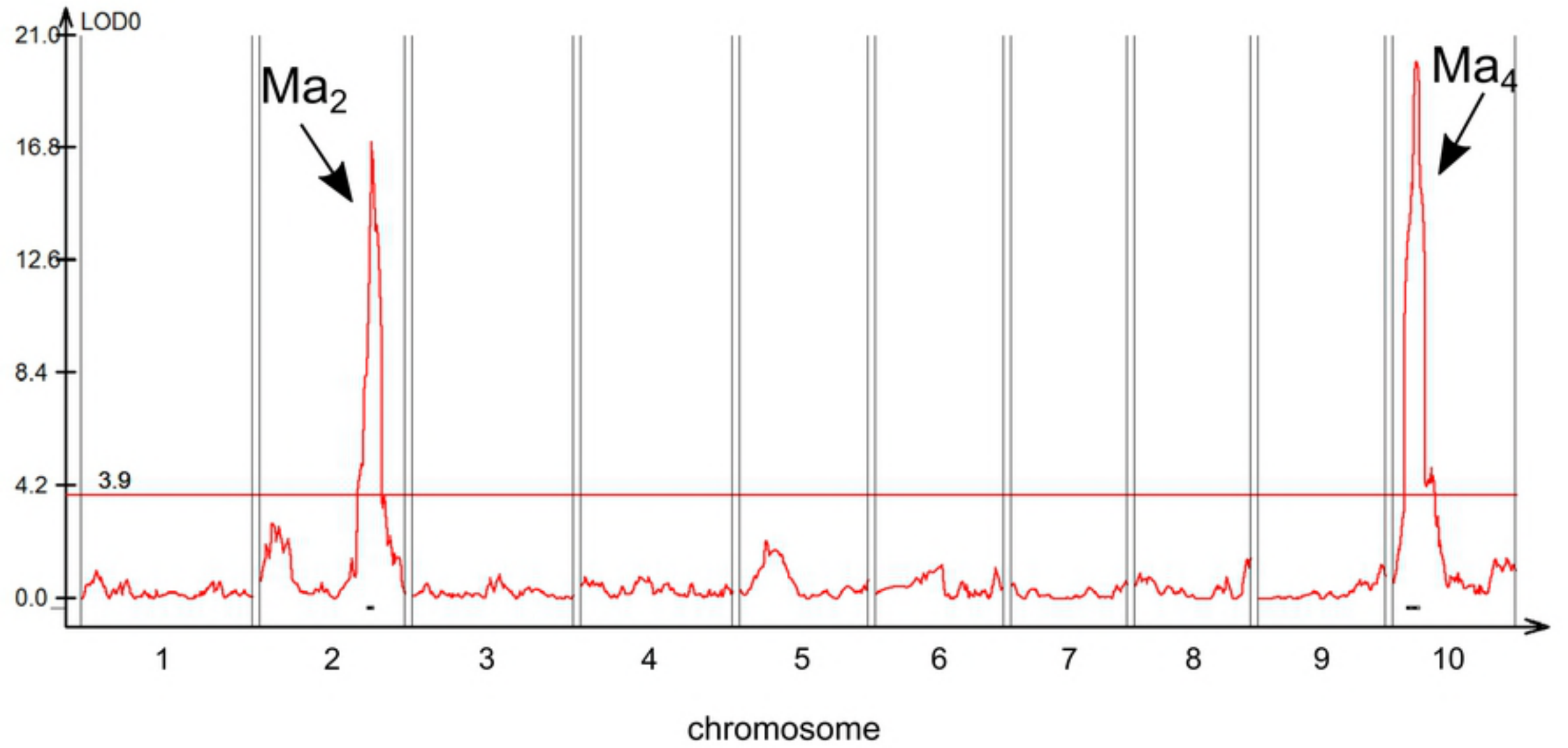
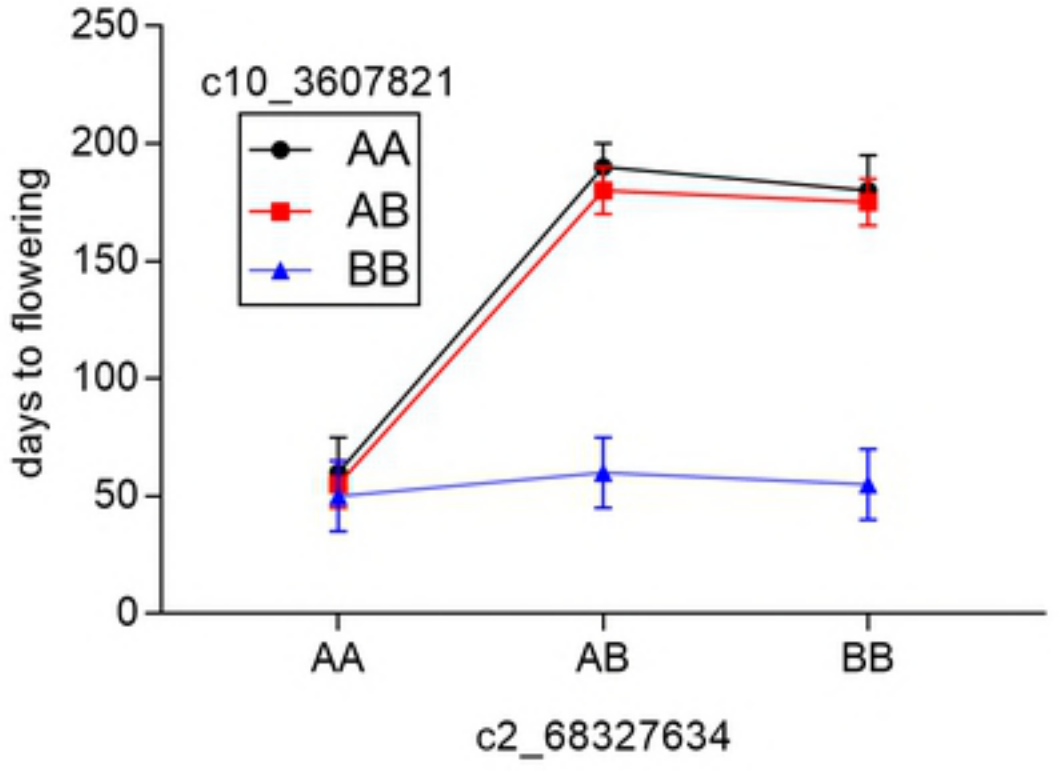


Figure 3


c10_3607821
AA
AB
BB
days to flowering
250
200
150
100
50
0
AA
AB
BB
c2_68327634

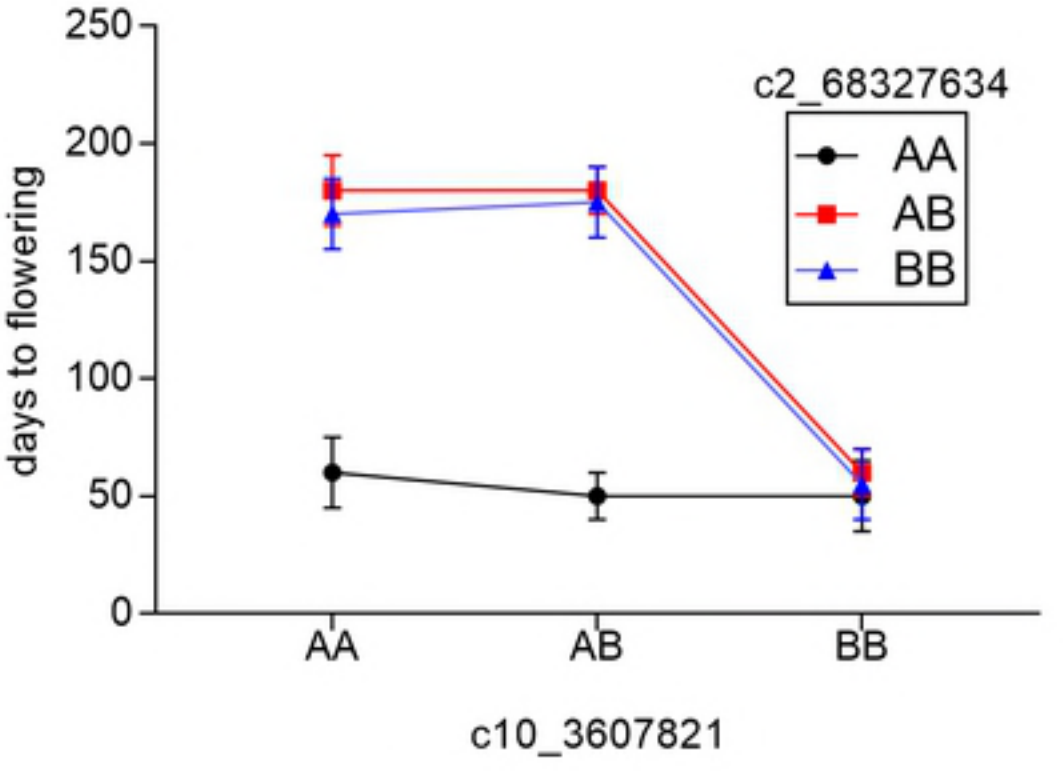

c2_68327634
AA
AB
BB
days to flowering
250
200
150
100
50
0
AA
AB
BB
c10_3607821

Figure 4

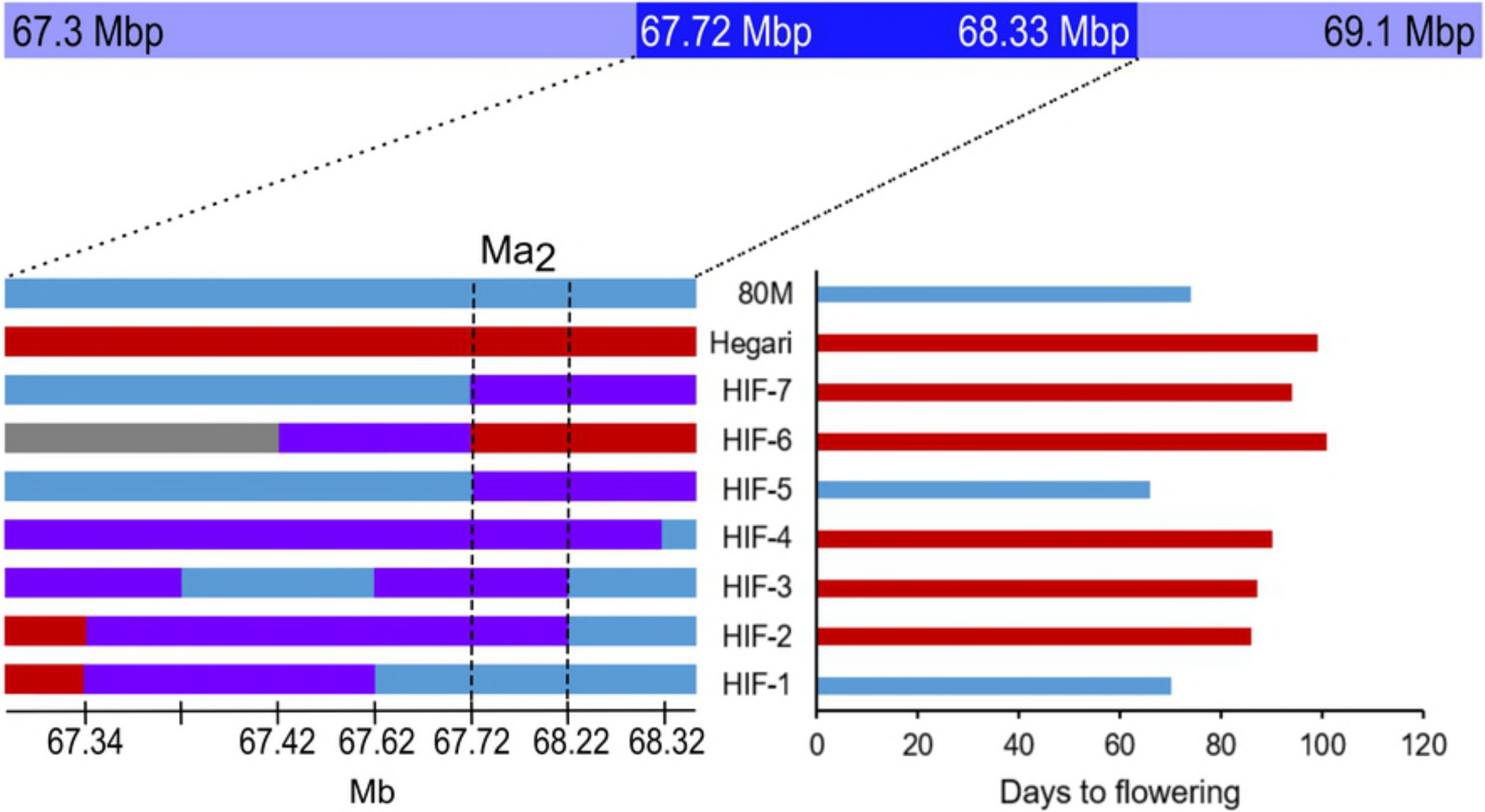

67.3 Mbp
67.72 Mbp
68.33 Mbp
69.1 Mbp
Ma2
80M
Hegari
HIF-7
HIF-6
HIF-5
HIF-4
HIF-3
HIF-2
HIF-1
67.34
67.42
67.62
67.72
68.22
68.32
Mb
0
20
40
60
80
100
120
Days to flowering

Figure 5

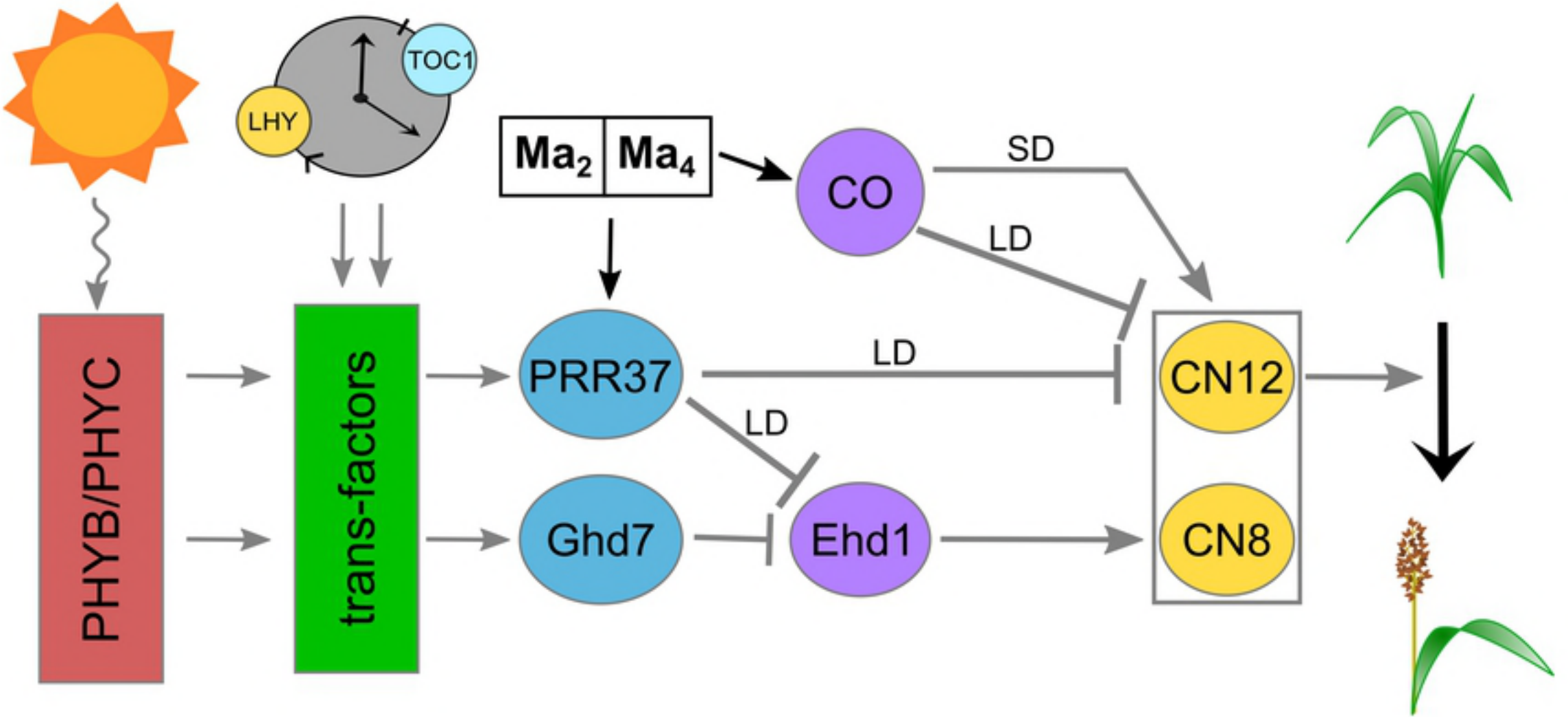

TOC1
LHY
Ma2
Ma4
CO
SD
LD
PHYB/PHYC
trans-factors
PRR37
LD
LD
CN12
Ghd7
Ehd1
CN8

Figure 7

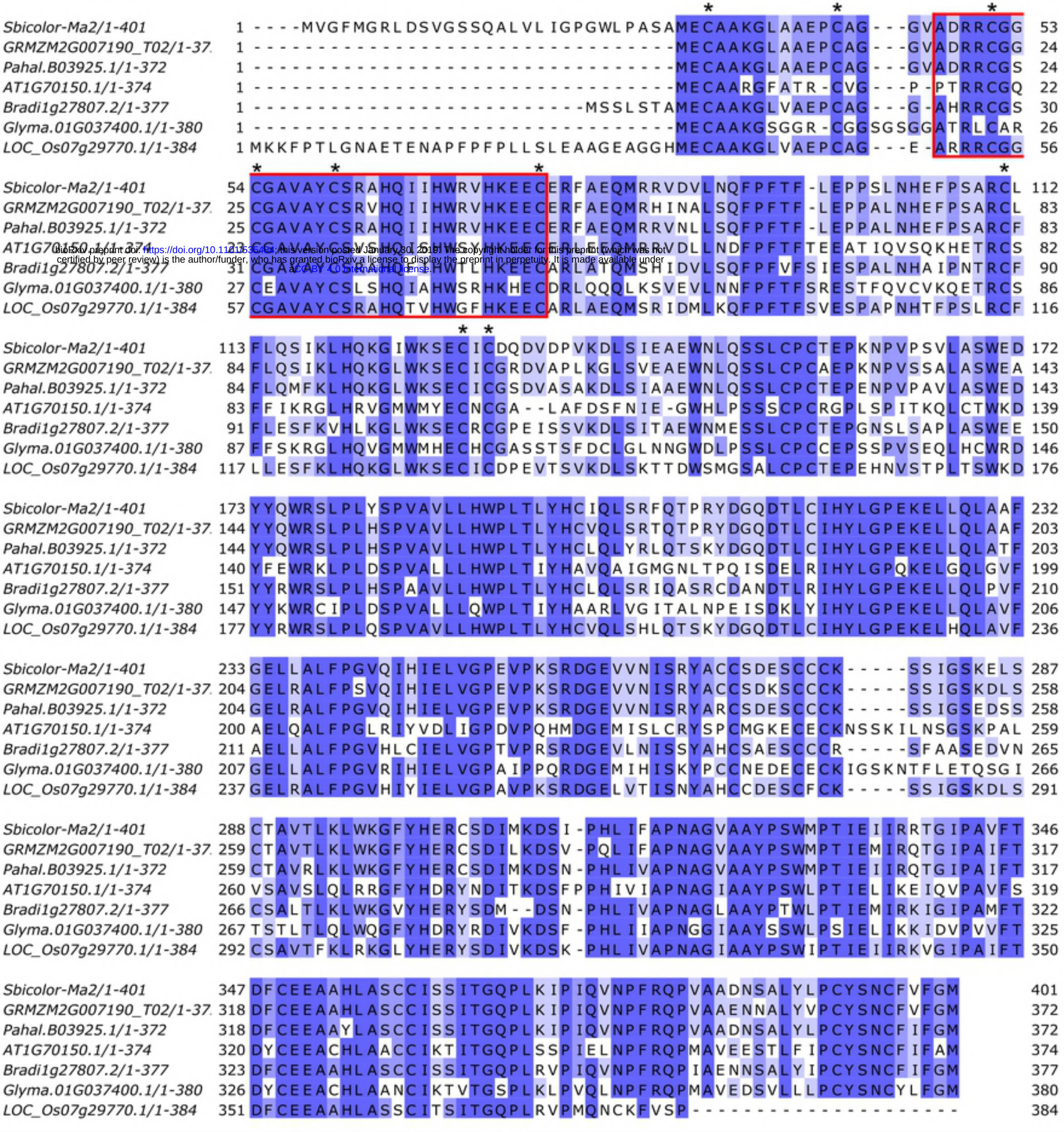


Figure 6