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

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

☉ These authors contributed equally to this work.
* jmullet@tamu.edu

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. Eight sorghum genotypes previously identified as recessive for Ma_2 contained the mutated version of Sobic.002G302700 present in 80M (ma_2) and one additional putative recessive ma_2 allele was identified in diverse sorghum accessions.

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 contrast, grain sorghum was adapted for production in temperate regions by selecting genotypes that have reduced photoperiod sensitivity resulting in earlier flowering and reduced risk of

exposure to drought, heat, or cold temperatures during the reproductive phase. A range of flowering times are found among forage and sweet sorghums [6]. Sweet sorghum genotypes with longer vegetative growth duration have larger stems that have greater potential for sucrose accumulation [6–8].

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

Many of the genes in the *Arabidopsis* flowering time pathway are found in grass species such as *Oryza sativa* (rice) [10], maize [19], and sorghum [3] however, the regulation of flowering time in these grasses has diverged from Arabidopsis in several important ways. Most genotypes of rice and sorghum are facultative short-day (SD) plants. In rice, the expression of the *FT-like* gene *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]. Both of these Milo genotypes were later found to encode the same *Ghd7* allele (*ghd7-1*) containing a stop codon [26]. Genetic analysis determined that mutations in three independently segregating *Maturity* (*Ma*) loci (*Ma$_1$, Ma$_2$, Ma$_3$*) were responsible for variation in flowering times in the Milo genotypes. A cross between Early White Milo (*ma$_1$Ma$_2$Ma$_3$*) and Dwarf Yellow Milo (*Ma$_1$ma$_2$ma$_3$*) was used to construct a set of Milo maturity standards (i.e., 100M, SM100, 80M, SM80), 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$_6$* segregating in other sorghum populations [31]. Subsequent research showed that all of the Milos are dominant for *Ma$_5$* and recessive for *ma$_6$* (*ghd7-1*)[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 potential activators of *SbCN8* and *SbCN12* expression, sources of florigen in sorghum.

The genes corresponding to four of the six *Maturity* loci have been identified. *Ma₁*, the locus with the greatest influence on flowering time photoperiod sensitivity, encodes *SbPRR37*, a pseudo-response regulator that inhibits flowering in LD [21]. *Ma₃* encodes phytochome B (phyB) [36], *Ma₅* encodes phytochrome C (phyC) [23], and *Ma₆* encodes Ghd7 a repressor of flowering in long days [26]. The genes corresponding to *Ma₂* and *Ma₄* have not been identified but recessive alleles at either locus results in early flowering in long days in sorghum lines that are photoperiod sensitive and have *Ma₁* genotypes [28]. Prior studies also noted that genotypes recessive for *Ma₂* flower later in genotypes that are photoperiod insensitive and recessive for *Ma₁* and *Ma₆* [28].

In this study, the impact of *Ma₂* alleles on the expression of genes in the sorghum flowering time pathway was characterized. A QTL corresponding to *Ma₂* was mapped and a candidate gene for *Ma₂* identified by fine mapping and genome sequencing. The results show that *Ma₂* enhances *SbPRR37* (*Ma₁*) and *SbCO* expression consistent with the impact of *Ma₂* alleles on flowering time in genotypes that vary in *Ma₁* alleles.

Methods

Plant growing conditions and populations

Seeds for all genotypes used in this study were obtained from the Sorghum Breeding Lab at Texas A&M University in College Station, TX. 100M (*Ma₁**Ma₂**Ma₃Ma₄Ma₅ma₆*) and 80M (*Ma₁**ma₂**Ma₃Ma₄Ma₅ma₆*) are sorghum maturity standards with defined maturity/flowering genotypes [1]. The maturity genotypes were selected from a cross between Early White Milo (*ma₁Ma₂Ma₃Ma₄Ma₅ma₆*) and Dwarf Yellow Milo (*Ma₁ma₂ma₃Ma₄Ma₅ma₆*). 100M and 80M are nearly isogenic and differ at *Ma2*.

The cross of 100M and 80M was carried out by the Sorghum Breeding Lab at Texas A&M University in College Station, TX. F₁ plants were grown in the field in Puerto Rico and self-pollinated to generate the F₂ population used in this study. The 100M/80M F₂ 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. F₁ plants were confirmed and self-pollinated to generate the F₂ population used in this study. The Hegari/80M F₂ 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 Metro-Mix 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 Light/Dark (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 heterogeneous inbred family (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 reads were mapped back to the sorghum reference genome (v1.0,

Phytozome v6). Genetic maps were created using MapMaker 3.0B with the Kosambi function [38]. QTL were mapped using WinQTLCartographer (v2.5.010) using composite interval mapping with a 1.0 cM walk speed and forward and backward model selection [39]. The threshold was set using 1000 permutations and α = 0.05. Upon release of v3.1 of the sorghum reference genome, the QTL coordinates were updated [40].

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

Fine mapping of the *Ma₂* QTL

All fine mapping populations for the *Ma₂* QTL were derived from F_2 individuals from the Hegari/80M population. The genetic distance spanning the *Ma₂* locus is 2 cM corresponding to a physical distance of ~1.8 Mbp, so 1000 progeny would be required to obtain 20 recombinants within the *Ma₂* QTL region. Six individuals that were heterozygous across the *Ma₂* QTL were self-pollinated to generate six HIFs totaling 1000 F_3 individuals. These individuals were grown out in the greenhouse, and flowering time was recorded. They 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 14h/10h L/D for the first 24 h and constant light for the second 24 h at 30°C. Plants were entrained for 3 d under these growth chamber conditions before beginning tissue collection. Leaf tissue was collected and pooled from 3 plants every 3 h for 48 h. The first sample was taken at lights-on on the first day of sample collection. 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 *Ma₂* 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.

Ma₂ phylogenetic analysis

Protein sequences of the closest homologs of Ma₂ were identified using BLAST analysis. Protein sequences were aligned using MUSCLE [45] and visualized using Jalview [46].

Table 1. Sequence variants of Sobic.002G203700 and their predicted effect on protein 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
Kalo	ma_2	L141*	Deleterious
IS3614-2	-	M83T	Deleterious

*Sequenced by Sanger sequencing

https://doi.org/10.1371/journal.pone.0212154.t001

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 ($Ma_1\boldsymbol{ma_2}Ma_3Ma_4Ma_5ma_6$) was previously reported to cause 80M to flower earlier than100M ($Ma_1\boldsymbol{Ma_2}Ma_3Ma_4Ma_5ma_6$) 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) in plants grown in LD [21] (Fig 1B). Analysis of *SbCO* expression in 100M and 80M showed that both peaks of *SbCO* expression were reduced in 80M compared to 100M (Fig 1B).

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* is expressed at lower levels and shows minimal response to day length [21,26]. SbPRR37 and SbCO are co-repressors of the expression of *SbCN8* and *SbCN12* in long
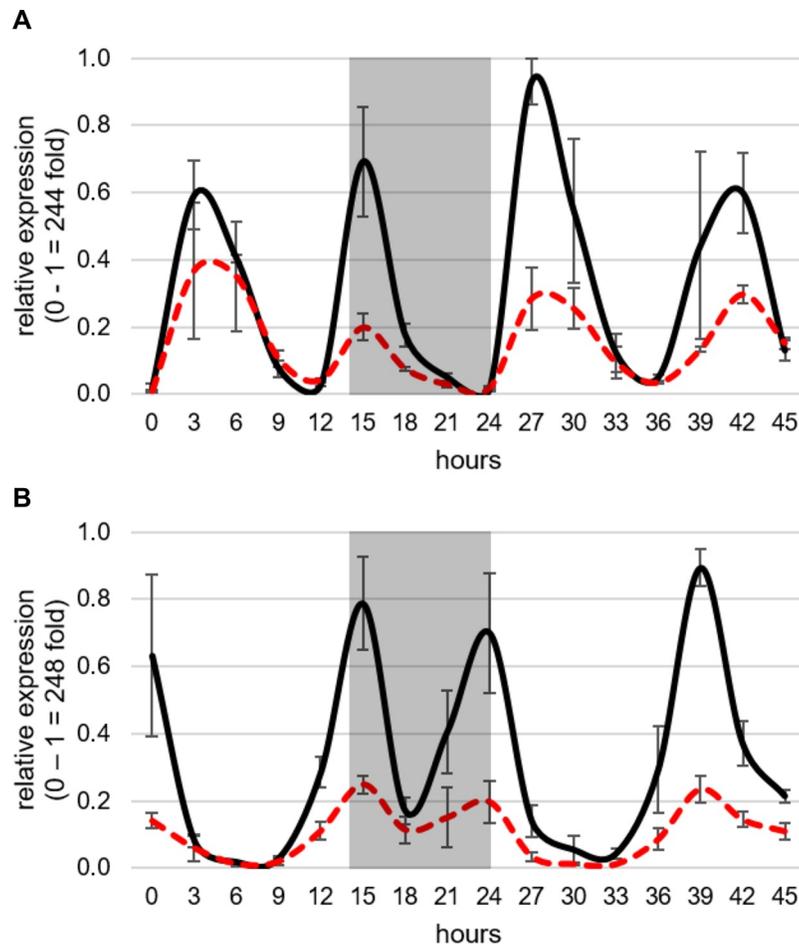
Fig 1. Circadian expression of genes regulating flowering in *S. bicolor* in 100M and 80M under long days. (A) Expression of *SbPRR37* in 100M (solid black lines) and 80M (dashed red lines). The expression peaks of *SbPRR37* are reduced in 80M. This is consistent with earlier flowering in 80M because *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.

https://doi.org/10.1371/journal.pone.0212154.g001

days, therefore, the influence of *Ma₂* alleles on *SbCN8/12/15* expression was investigated [21,27]. When plants were grown in long days, expression of *SbCN12* was ~5 fold higher in 80M compared to 100M consistent with earlier flowering in 80M (Fig 2).

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 *Ma₂* alleles have a limited influence on the expression of these genes (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 Ma₂ 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 *Ma₁Ma₂* vs. *Ma₁ma₂* in long days is consistent with lower *SbCN12* expression in *Ma₁Ma₂* genotypes.
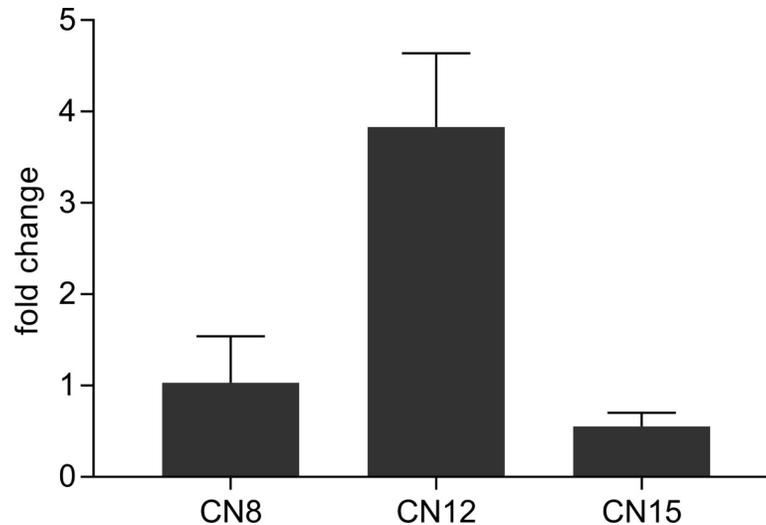
Fig 2. Expression of the *S. bicolor FT*-like genes *SbCN8*, *SbCN12*, and *SbCN15* in long days at the expected peak of expression. Expression of *SbCN12* is elevated in 80M, which is consistent with earlier flowering in that genotype. All expression values are normalized to *SbUBC* and are the mean of 3 biological replicates. Fold change was calculated as $2^{-[Ct(100M)-Ct(80M)]}$.

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 *Ma$_2$* 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 lack of DNA markers associated with the 80M/100M population, a second mapping population was created to identify the genetic locus associated with *Ma$_2$*. An F_2 population (n = 215) that would segregate for *Ma2* and *Ma4* was constructed by crossing Hegari (*Ma$_1$**Ma$_2$***Ma$_3$**ma$_4$**Ma$_5$ma$_6$*) and 80M (*Ma$_1$**ma$_2$***Ma$_3$**Ma$_4$***Ma$_5$ma$_6$*) [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 *Ma$_4$* (S2 Fig). The 80M haplotype across the QTL on SBI02 was associated with early flowering therefore the QTL on SBI02 was assigned to *Ma$_2$*.

Epistatic interactions between *Ma$_2$* and *Ma$_4$*

Previous studies indicated an epistatic interaction exists between *Ma$_2$* and *Ma$_4$* [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
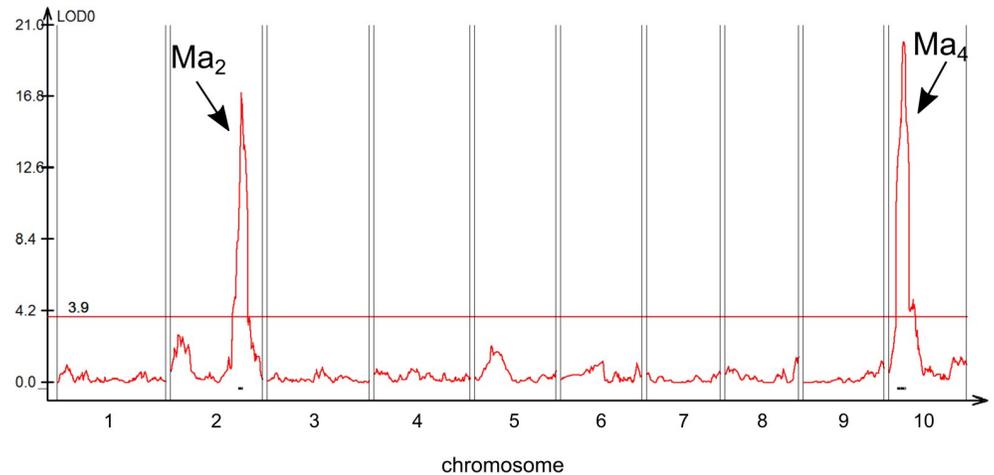
Fig 3. Quantitative trail locus (QTL) map of flowering time in the Hegari/80M F$_2$ population. Two QTL were identified for variation in flowering time in the F$_2$ population derived from Hegari ($Ma_1Ma_2Ma_3ma_4$) and 80M ($Ma_1ma_2Ma_3Ma_4$). This population was expected to segregate for Ma_2 and Ma_4. Each recessive *Ma* allele causes earlier flowering. The QTL on LG10 corresponds to Ma_4 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 Ma_2.

SBI10 and an additional QTL on SBI09. Additionally, an epistatic interaction was identified between Ma_2 and Ma_4 (pLOD = 42). Interaction plots showed that in a dominant Ma_4 background, a dominant allele at Ma_2 delays flowering, while in a recessive Ma_4 background, Ma_2 has a minimal impact on flowering time (Fig 4). The interaction between Ma_2 and Ma_4 identified by MQM analysis is consistent previous observations that in a recessive ma_4 background flowering is early regardless of allelic variation in Ma_2 [28].

Ma2 candidate gene identification

The Hegari/80M F$_2$ population located Ma_2 on SBI02 between 67.3 Mbp to 69.1 Mbp (Fig 5). To further delimit the Ma_2 locus, six lines from the Hegari/80M population that were heterozygous across the Ma_2 QTL but fixed across the Ma_4 locus (Ma_4Ma_4) were selfed to create heterogeneous inbred families (HIFs) (n = 1000 F$_3$ plants) [53]. Analysis of these HIFs narrowed the region encoding Ma_2 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
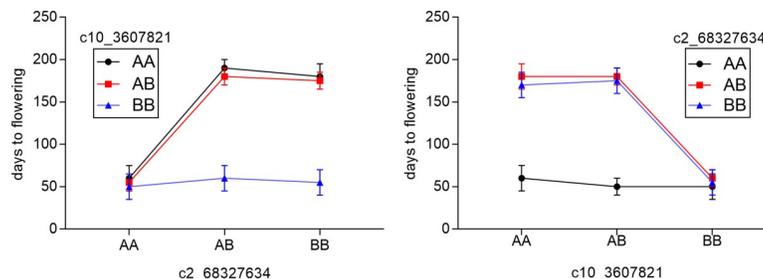

Fig 4. Interaction plots for the Ma2 QTL and the Ma4 QTL. There is a known interaction between Ma_2 (represented by marker c2_68327634) and Ma_4 (represented by marker c10_3607821). This interaction was identified by multiple QTL mapping (MQM). Dominant alleles of the *Ma* genes delay flowering. In a recessive ma4 background (AA at c2_68327634), the effect of Ma_2 on days to flowering is reduced. A represents the 80M allele and B represents the Hegari allele at each QTL. Reciprocal plots are shown.

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).

The low rate of recombination across the *Ma₂* locus led us to utilize whole genome sequencing in conjunction with fine mapping to identify a candidate gene for *Ma₂*. Since 100M and 80M are near isogenic lines that have very few sequence differences along the long arm of SBI02 where the *Ma₂* QTL is located, whole genome sequences (WGS) of 100M and 80M were generated in collaboration with JGI (sequences available at https://phytozome.jgi.doe.gov). The genome sequences were scanned for polymorphisms within the 500 kb locus spanning *Ma₂*. Only one T → A single nucleotide polymorphism (SNP) located in Sobic.002G302700 was identified that distinguished 100M and 80M within the region spanning the *Ma₂* 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 (*ma₂*) whereas 100M (*Ma₂*) 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 *Ma₂*.

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 targets and have roles in regulating chromatin state, transcription, signal transduction, and cell cycling [54,55]. The SET domain in SMYD-containing proteins is composed of two sub-domains that are divided by the MYND zinc-finger domain. The SET domain includes conserved sequences involved in methyltransferase activity including nine cysteine residues that are present in the protein encoded by Sobic.002G303700 (Fig 6) [56]. The MYND domain
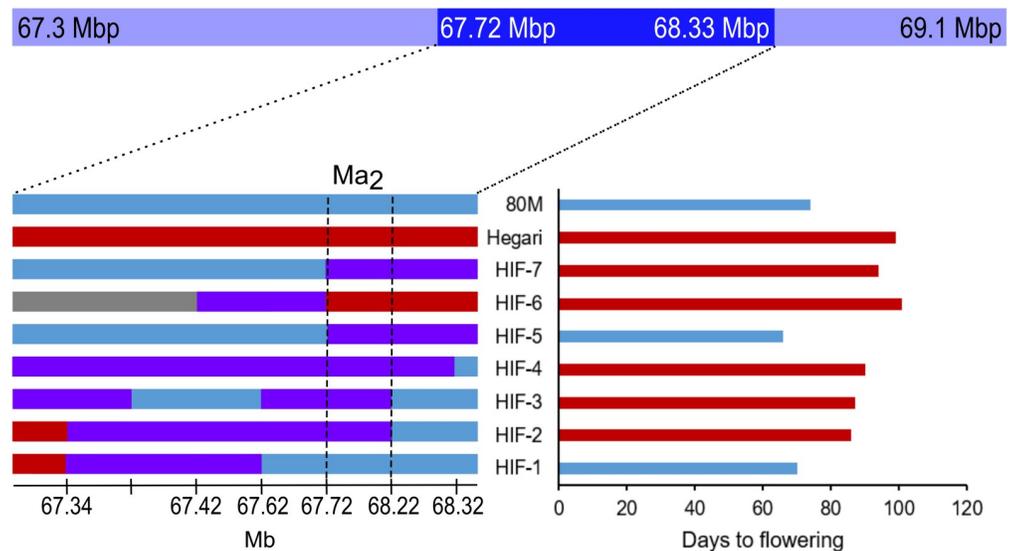

Fig 5. Fine-mapping of the Ma2 QTL. The *Ma₂* QTL spans from 67.3 Mpb to 69.1 Mbp (light blue bar). Five F₂ individuals that were heterozygous across the *Ma₂* QTL were self-pollinated to generate heterogeneous inbred families (HIFs) totaling 1000 F₃ individuals. Genotype and phenotype analysis of these 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.
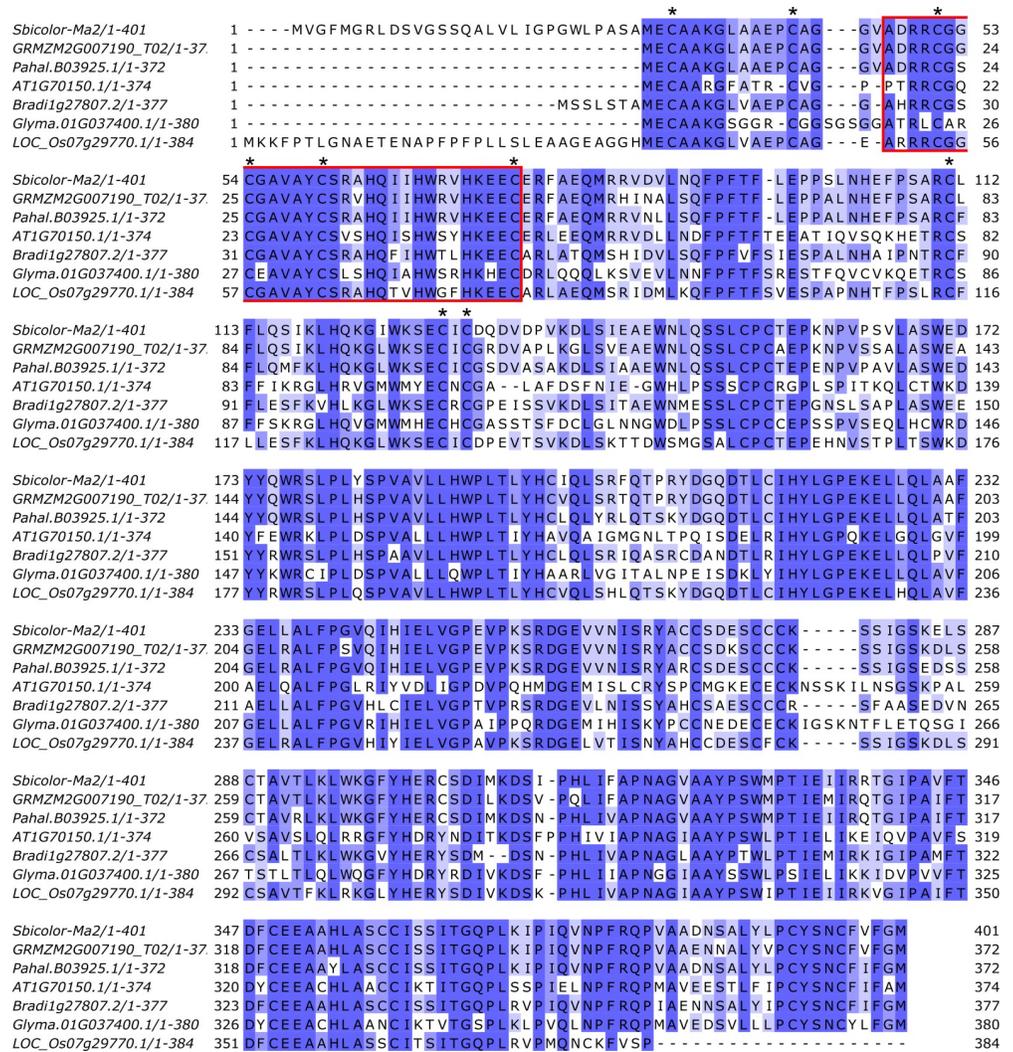
https://doi.org/10.1371/journal.pone.0212154.g005

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

https://doi.org/10.1371/journal.pone.0212154.g006

is involved in binding DNA and is enriched in cysteine and histidine residues [57]. Protein sequence alignment of Sobic.002G302700 homologs revealed that the SYMD protein candidate for Ma2 is highly conserved across flowering plants (Fig 6).

To learn more about *Ma2* regulation, the expression of Sobic.002G302700 in 100M and 80M was characterized during a 48h L:D/L:L cycle. Ma_2 showed a small increase in expression from morning to evening and somewhat higher expression in 100M compared to 80M during the evening (S3 Fig).

Distribution of Ma₂ alleles in the sorghum germplasm

Recessive *ma₂* was originally found in the Milo background and used to construct Double Dwarf Yellow Milo ($Ma_1\boldsymbol{ma_2}ma_3Ma_4Ma_5ma_6$) [28]. Double Dwarf Yellow Milo was crossed to Early White Milo ($ma_1\boldsymbol{Ma_2}Ma_3Ma_4Ma_5ma_6$) and the progeny selected to create 100M, 80M and the other Milo maturity standards [1,28,58]. Several of the Milo maturity standards were

recorded as recessive *Ma₂* (80M, 60M, SM80, SM60, 44M, 38M) and others as *Ma₂* dominant (100M, 90M, SM100, SM90, 52M). In order to confirm the *Ma₂* genotype of the maturity standards, the 500 bp sequence spanning the Lys141* mutation in Sobic.002G302700 was obtained from most of these genotypes (Table 1). Kalo was also identified as carrying a recessive allele of *Ma₂*. Kalo was derived from a cross of Dwarf Yellow Milo (*ma₂*), Pink Kafir (*Ma₂*), and CI432 (*Ma₂*), therefore it was concluded that DYM is the likely source of recessive *ma₂* [28]. Sequence analysis showed that the genotypes previously identified as *ma₂* including Kalo, 80M, SM80, 60M, 44M, and 38M carry the recessive mutation in Sobic.002G302700 identified in 80M. 100M, SM100, and Hegari that were identified as *Ma₂*, did not contain the mutated version of Sobic.002G302700 (Table 1). Additionally, sequences of *Ma₂* from 52 sorghum genotypes with publicly available genome sequences were compared [40]. Sobic.002G302700 was predicted to encode functional proteins in all except one of these sorghum genotypes. A possible second recessive *Ma₂* allele was found in IS3614-2 corresponding to an M83T missense mutation that was predicted to be deleterious by PROVEAN [59].

Discussion

In photoperiod sensitive sorghum genotypes, following the vegetative juvenile phase, day length has the greatest impact on flowering time under normal growing conditions. Molecular identification of the genes corresponding to *Ma₁*, *Ma₃*, *Ma₅* and *Ma₆* 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 Fig 7 [60]. The current study showed that Ma2 represses flowering in long days by increasing the expression of SbPRR37 (*Ma₁*) and *SbCO*. The study also located QTL for *Ma₂* and *Ma₄*, confirmed an epistatic interaction between *Ma₂* and *Ma₄*, and identified a candidate gene for *Ma₂*.

In the current study, two near isogenic Milo maturity genotypes, 100M (*Ma₂*) and 80M (*ma₂*), were used to characterize how allelic variation in *Ma₂* affects the expression of genes in the sorghum photoperiod regulated flowering time pathway. This analysis showed that mutation of *Ma₂* significantly reduced the amplitude of the morning and evening peaks of *SbPRR37* and *SbCO* expression without altering the timing of their expression. In parallel, the expression of *SbCN12* (*FT*-like) increased 8-fold in leaves of 80M compared to 100M, consistent with
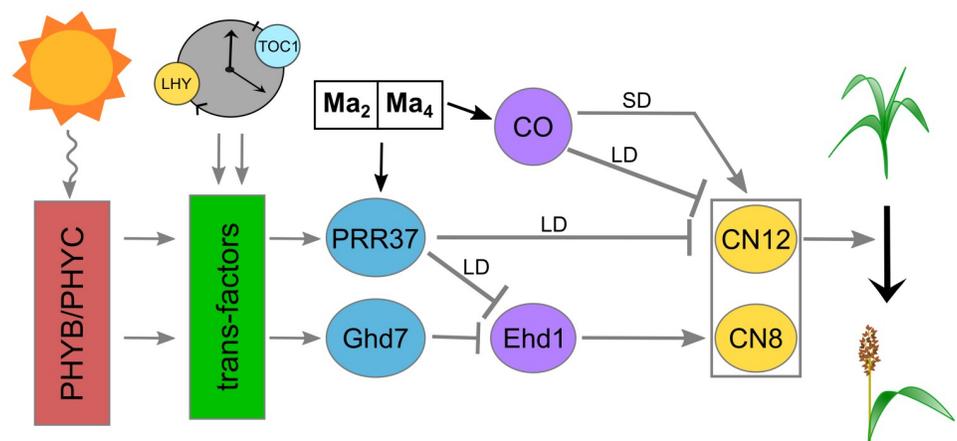

Fig 7. A model of the flowering time regulatory pathway in *S. bicolor*. *Ma₂* and *Ma₄* work codependently to enhance the expression of *SbPRR37* and *SbCO*. In LD, SbPRR37 and SbCO in turn repress the expression of the *SbCN* genes, especially *SbCN12*, to repress the floral transition.

https://doi.org/10.1371/journal.pone.0212154.g007

prior studies showing that 80M (*ma₂*) flowers earlier than 100M (*Ma₂*) in long days [28]. 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 *Ma₂*. Based on these results, we tentatively place *Ma₂* in the flowering time pathway downstream of the light sensing phytochromes and circadian clock and identify *Ma₂* as a factor that enhances *SbPRR37* and *SbCO* expression (Fig 7).

The differential increase in *SbCN12* expression in 80M (vs. 100M) is consistent with inhibition of *SbCN12* expression in long days by the concerted action of SbPRR37 and SbCO [27]. Genetic studies showed that floral repression mediated by SbPRR37 requires SbCO as a co-repressor [27]. Therefore, enhanced expression of both *SbPRR37* (*Ma₁*) and *SbCO* by *Ma₂* in *Ma₁Ma₂* genotypes in long days is consistent with delayed flowering under these conditions relative to genotypes such as 80M that are *Ma₁ma₂*. Molecular genetic studies also showed that SbCO is an activator of *SbCN12* expression and flowering in *ma₁* genetic backgrounds [27]. This is consistent with the observation that *ma₁Ma₂* genotypes flower earlier than *ma₁ma₂* genotypes when grown in long days [28].

Interactions between *Ma₂* and *Ma₄*

Multiple QTL (MQM) analysis of results from the population derived from Hegari/80M identified an interaction between *Ma₂* and *Ma₄* as well as one additional flowering QTL on SBI09. Flowering time QTL on SBI09 have been identified in other mapping populations, but the gene(s) involved have not been identified [33,34]. The interaction between *Ma₂* and *Ma₄* confirmed previous observations that recessive *ma₄* causes accelerated flowering in long days in *Ma₁Ma₂* genotypes [28]. Interestingly, the influence of *Ma₂* and *Ma₄* alleles on flowering time is affected by temperature [28,61]. The influence of temperature on flowering time pathway gene expression in 80M and 100M in the current study was minimized by growing plants at constant 30C. Further analysis of the temperature dependence of Ma2 and Ma4 on flowering time may help elucidate interactions between photoperiod and flowering time that have been previously documented [28,62]. Positional cloning of *Ma₄* is underway to better understand the molecular basis of *Ma₂* and *Ma₄* interaction and their impact on flowering time.

Identification of a candidate gene for *Ma₂*

A mapping population derived from Hegari/80M that segregated for *Ma₂* and *Ma₄* enabled localization of the corresponding flowering time QTL in the sorghum genome (SBI02, *Ma₂*; SBI10, *Ma₄*). The *Ma₂* QTL on SBI02 was fine-mapped using heterozygous inbred families (HIFs) from Hegari/80M. Identification of a candidate gene for *Ma₂* was subsequently aided by comparison of genome sequences from the closely related 80M and 100M genotypes [28] 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 *Ma₂* locus that 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 *Ma₂*.

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, Ma₂ 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 long day floral co-repressors *SbPRR37* and *SbCO* (Fig 7). Sobic.002G302700 was identified as the best candidate for the sorghum *Maturity* locus Ma_2. 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 help elucidate an additional module of the photoperiod flowering regulation pathway in sorghum.

Supporting information

S1 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).
(TIF)

S2 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.
(TIF)

S3 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.
(TIF)

S1 Table. *Ma$_2$* (Sobic.002G302700) sequencing and qPCR primers.
(DOCX)

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

Acknowledgments

Morishige for his assistance in constructing the Hegari/80M cross as well as Robin Poncik for her assistance in recording flowering dates.

Author Contributions

Conceptualization: John E. Mullet.

Data curation: Anna L. Casto, Ashley J. Mattison.

Formal analysis: Anna L. Casto, Ashley J. Mattison, Sara N. Olson, Manish Thakran.

Funding acquisition: John E. Mullet.

Investigation: Anna L. Casto.

Methodology: Anna L. Casto, Ashley J. Mattison, Sara N. Olson.

Resources: William L. Rooney.

Supervision: John E. Mullet.

Writing – original draft: Anna L. Casto, Ashley J. Mattison, John E. Mullet.

Writing – review & editing: Anna L. Casto, John E. Mullet.

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. https://doi.org/10.1093/jxb/eru229 PMID: 24958898

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. https://doi.org/10.1111/tpj.14113 PMID: 30260043

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. https://doi.org/10.1016/j.tplants.2013.05.003 PMID: 23790253

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. https://doi.org/10.1016/j.pbi.2013.01.005 PMID: 23453779

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. https://doi.org/10.1101/cshperspect.a027748 PMID: 27663772

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. https://doi.org/10.1126/science.1141752 PMID: 17446353

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. https://doi.org/10.1126/science.1115983 PMID: 16099979

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. PMID: 9657154

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. PMID: 10477524

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. https://doi.org/10.1038/35074138 PMID: 11323677

18. Turnbull C. Long-distance regulation of flowering time. J Exp Bot. 2011; 62(13):4399–413. https://doi.org/10.1093/jxb/err191 PMID: 21778182

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. https://doi.org/10.1126/science.1141753 PMID: 17446351

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. https://doi.org/10.1111/nph.13834 PMID: 26765652

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. https://doi.org/10.1371/journal.pone.0105352 PMID: 25122453

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. https://doi.org/10.1104/pp.15.01723 PMID: 26864016

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. https://doi.org/10.1038/ng.143 PMID: 18454147

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. https://doi.org/10.1186/1471-2229-14-148 PMID: 24884377

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. https://doi.org/10.1534/g3.114.013318 PMID: 25237111

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. PMID: 9046599

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. https://doi.org/10.1186/1471-2164-14-1 PMID: 23323973

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. PMID: 3692487

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. https://doi.org/10.1111/tpj.13781 PMID: 29161754

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. https://doi.org/10.1534/genetics.108.094565 PMID: 19104078

42. Broman KW, Wu H, Sen Ś, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 2003; 19(7):889–90. PMID: 12724300

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. https://doi.org/10.1006/meth.2001.1262 PMID: 11846609

45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. England; 2004; 32(5):1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

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. https://doi.org/10.1093/bioinformatics/btp033 PMID: 19151095

47. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4(4):406–25. https://doi.org/10.1093/oxfordjournals.molbev.a040454 PMID: 3447015

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. https://doi.org/10.1093/molbev/msw054 PMID: 27004904

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. https://doi.org/10.1093/bioinformatics/btq565 PMID: 20966004

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. https://doi.org/10.1534/genetics.115.178608 PMID: 26323882

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. https://doi.org/10.1093/jmcb/mjr015 PMID: 21724641

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. https://doi.org/10.3390/ijms16011406 PMID: 25580534

56. Min J, Zhang X, Cheng X, Grewal SIS, Xu RM. Structure of the SET domain histone lysine methyltransferase Clr4. Nat Struct Biol. 2002; 9(11):828–32. https://doi.org/10.1038/nsb860 PMID: 12389037

57. Gross CT, McGinnis W. DEAF-1, a novel protein that binds an essential region in a Deformed response element. EMBO J. 1996; 15(8):1961–70. PMID: 8617243

58. Quinby JR, Karper RE. Effect of Different Alleles on the Growth of Sorghum Hybrids. 1948;255–9.

59. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS One. 2012; 7(10):e46688. https://doi.org/10.1371/journal.pone.0046688 PMID: 23056405

60. Mullet JE. High-Biomass C4 Grasses − Filling the Yield Gap. Plant Sci. Elsevier; 2017; 261:10–7. https://doi.org/10.1016/j.plantsci.2017.05.003 PMID: 28554689

61. Major DJ, Rood SB, Miller FR. Temperature and Photoperiod Effects Mediated by the Sorghum Maturity Genes. Crop Sci. 1990; 30(2):305–10.

62. Tarumoto I. Thermo-sensitivity and photoperiod sensitivity genes controlling heading time and flower bud initiation in Sorghum, Sorghum bicolor Moench. Japan Agric Res Q. 2011; 45(1):69–76.

63. Gan ES, Xu Y, Wong JY, Geraldine Goh J, Sun B, Wee WY, et al. Jumonji demethylases moderate precocious flowering at elevated temperature via regulation of FLC in Arabidopsis. Nat Commun. 2014; 5:1–13.