Enhancer-Promoter Interaction of SELF PRUNING 5G Shapes Photoperiod Adaptation

Shuaibin Zhang, Zhicheng Jiao, Lei Liu, Ketao Wang, Deyi Zhong, Shengben Li, Tingting Zhao, and Xia Cui

Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China
Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China
Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China
Northeast Agricultural University, Harbin 150030, China

ORCID IDs: 0000-0001-9565-6499 (D.Z.); 0000-0002-1415-0865 (X.C.)

Tomato (Solanum lycopersicum) is a major vegetable fruit grown and consumed worldwide. Modern cultivated tomatoes are derived from their wild relative, Solanum pimpinellifolium, a short-day plant that originated from the Andean region of South America. The molecular underpinnings of the regional adaptation and expansion of domesticated tomato remain largely unclear. In this study, we examined flowering time in wild and cultivated tomatoes under both long-day and short-day conditions. Using quantitative trait locus mapping in a recombinant inbred line population, we identified SELF PRUNING 5G (SP5G) as a major locus influencing daylength adaptation in tomato. Genetic diversity analysis revealed that the genomic region harboring SP5G shows signatures of a domestication sweep. We found that a 52-bp sequence within the 3’ untranslated region of SP5G is essential for photoperiodic response in cultivated tomatoes, uncovering a regulatory mechanism that could potentially be used to manipulate flowering time in tomato through novel biotechnological approaches.

Flowering is the transition from vegetative growth to reproductive growth and depends on internal signals and external cues (Bäurle and Dean, 2006; Andrés and Coupland, 2012). Plants evolved distinct photoperiodic responses to adapt to their local environments. For long-day (LD) plants, prolonged daytime induces flowering, while short-day (SD) plants flower earlier under SD conditions. The ability of plants to respond to photoperiod requires the detection of daylength. In the LD plant Arabidopsis (Arabidopsis thaliana), CONSTANS (CO) is a major regulator of photoperiodic flowering (Samach et al., 2000; Suárez-López et al., 2001; Imaizumi and Kay, 2006; Song et al., 2015). The florigen component FLOWERING LOCUS T (FT), a CENTRORADIALIS/TERMINAL FLOWER 1/SELF-PRUNING (CETS) family member, is the final output of the photoperiodic response downstream of CO (Samach et al., 2000; Wigge et al., 2005). CO is regulated at both the transcriptional (Fornara et al., 2009) and post-transcriptional (Valverde et al., 2004; Jang et al., 2008; Liu et al., 2008) levels during the light period, allowing increased accumulation of CO protein during the daytime to induce FT, which, in turn, promotes flowering (Samach et al., 2000). In the SD plant rice (Oryza sativa), the CO homolog HEADING DATE 1 (HD1) represses the FT homolog HD3A under LD conditions but activates it under SD (Izawa et al., 2002; Jang et al., 2008; Liu et al., 2008) levels during the light period, allowing increased accumulation of CO protein during the daytime to induce FT, which, in turn, promotes flowering (Samach et al., 2000). In the SD plant rice (Oryza sativa), the CO homolog HEADING DATE 1 (HD1) represses the FT homolog HD3A under LD conditions but activates it under SD (Izawa et al., 2002; Hayama et al., 2003), which ultimately results in flowering. In addition, a rice-specific regulatory pathway, GRAIN NUMBER, PLANT HEIGHT, AND HEADING DATE 7 (GHD7)-EARLY HEADING DATE1 (EHD1)-HD3A/RISE FLOWERING LOCUS T1, also regulates daylength responses in rice (Song et al., 2015).

Over the course of domestication, the photoperiodic sensitivity of many crops has been modified to suit...
broad environmental conditions and different latitudes. The variances of several photoperiodic pathway genes, including GHD7, HD1, HD3A, and EHD1 in rice (Fujino and Sekiguchi, 2005; Xue et al., 2008; Takahashi et al., 2009; Fujino et al., 2013; Matsubara et al., 2014) and ZmCCT (a CCT domain-containing protein) in maize (Zea mays; Yang et al., 2013), resulted in the earlier flowering of these SD crops under LD conditions. As another example, the reduced photoperiodic sensitivity in wheat (Triticum aestivum) and barley (Hordeum vulgare) caused by mutation of the PHOTOPERIOD1 gene contributed greatly to the Green Revolution (Turner et al., 2005; Beales et al., 2007; Wilhelm et al., 2009). Thus, decreased daylength sensitivity plays a vital role in crop adaptation and spread.

Cultivated tomato was domesticated from the wild species Solanum pimpinellifolium, which originated in the Andean region of South America near the equator (Jenkins, 1948). Tomato was taken to Europe in the 16th century and then spread worldwide. The wild ancestors of tomato are considered SD plants, while most modern cultivars are considered day neutral plants due to the loss of or weakened photoperiodic sensitivity, rendering them adaptive to northern latitudes. In tomato, homologs of the Arabidopsis CETS family member in tomato, acts as a flowering repressor in daylength responses (Cao et al., 2016; Pnueli et al., 1998), play important roles in controlling flowering time. However, none of them were proven to affect daylength sensitivity, suggesting that the photoperiod adaptation genes in tomato are different from those in Arabidopsis and rice. Recent studies showed that the SP5G gene, another CETS family member in tomato, acts as a flowering repressor in daylength responses (Cao et al., 2016; Soyk et al., 2017). Higher SP5G expression in Solanum galapagense causes later flowering under LD, whereas the weakened daytime induction of its transcripts in cultivated tomato results in early flowering under LD (Soyk et al., 2017). However, the detailed molecular mechanism that regulates SP5G expression in response to daylength in wild and cultivated tomatoes remains unknown.

In this study, we identified and characterized a major flowering quantitative trait locus (QTL), Q-fo1_01, from a tomato recombinant inbred line (RIL) population. Our analysis revealed that SP5G is the gene underlying this QTL for adaptation to daylength and that the genomic region around SP5G was selected during domestication. We found that the 3’ untranslated region (UTR) of SP5G acts as an enhancer to activate its expression. Moreover, a 52-bp sequence within the SP5G 3’ UTR is essential for the formation of an enhancer-promoter loop that regulates SP5G expression in tomato. The absence of the 52-bp sequence in many big-fruit cultivars corresponds to reduced SP5G expression, resulting in daylength insensitivity in these cultivars. Our results demonstrate that the cis-regulatory variation at the enhancer region of the SP5G 3’ UTR confers reduced photoperiodic response in cultivated tomatoes.

RESULTS

The Flowering Time of Cultivated Tomato Is Less Sensitive to Daylength

Unlike the tomato wild relative S. pimpinellifolium, which is an SD plant, many tomato cultivars are day neutral (Jenkins, 1948; Soyk et al., 2017). To better understand the daylength response of the wild progenitor and cultivated tomatoes, we measured flowering time by counting the number of leaves before the first inflorescence under LD and SD conditions in 31 big-fruit cultivars, 26 wild accessions (S. pimpinellifolium), and 11 cherry tomatoes, which are considered intermediate between the wild ancestor and cultivars (Blanca et al., 2012). Under SD conditions, most of the wild species flowered earlier than the big-fruit cultivars (Fig. 1A). The flowering time of many big-fruit tomatoes was earlier than that of wild species under LD conditions, although nearly all accessions exhibited delayed flowering as daylength increased (Fig. 1A). We defined the relative flowering time as the difference in leaf number between LD and SD. The average relative flowering time of wild species was much greater than that of the big-fruit cultivars (Fig. 1B), indicating that the wild species are more sensitive to daylength than the cultivated tomatoes.

SP5G Is a Major Locus Contributing to Daylength Adaptation in Tomato

To explore the genetic basis of the difference in photoperiodic sensitivity between cultivated tomatoes and their wild relatives, we generated a stable RIL population of 219 lines after 10 selfing generations by crossing S. lycopersicum var cerasiforme (LA1310; CC) with S. lycopersicum cv Moneymaker (MM; Fig. 2A), which are sensitive and insensitive to daylength, respectively (Fig. 2B). To identify the QTLs associated with flowering time, we used whole-genome sequencing to detect single-nucleotide polymorphisms (SNPs) in the two parents and all RIL lines. On average, each independent line was sequenced with approximately 1.7-fold coverage, and about 970,729 SNPs were identified. Under LD conditions, there was one peak identified above the P value cutoff of 10−10 that was not present in the SD results (Fig. 2C; Supplemental Fig. S1). This peak region was located on the long arm of chromosome 5, spanning 420 kb and containing 51 genes (Fig. 2D).

To verify that this locus contributes to the daylength sensitivity of CC, we generated near-isogenic lines (NILs; Fig. 2E) in which the Q-fo1_01 locus of the MM genotype was introgressed into CC by backcrossing six times. A total of 88 markers along 12 chromosomes were used to evaluate the background of the NILs, and both NIL-fo1_01CC and NIL-fo1_01MM had the CC genotype for nearly 95% (83 of 88) of the tested markers (Supplemental Table S1). Between NIL-fo1_01CC and NIL-fo1_01MM, the region of difference around the
Q-fo1_01 locus was only about 50 kb (Fig. 2D; Supplemental Fig. S2). When evaluating the flowering time of the NILs under LD and SD conditions, we found that NILs with different genotypes at Q-fo1_01 had a similar flowering time under SD. Under LD conditions, however, NIL-fo1_01 CC plants flowered much later than NIL-fo1_01 MM (Fig. 2F).

One of the four predicted genes in this 50-kb region was Solyc05g053850 (Fig. 2D), which encodes the FT homolog SP5G and was reported previously to regulate the daylength response in tomato (Soyk et al., 2017). To directly test whether this gene was responsible for the flowering phenotypes observed in NIL-fo1_01 CC and MM, we mutated SP5G in NIL-fo1_01 CC and MM by CRISPR/Cas9-mediated genome editing. The four null-allele sp5g mutants obtained in NIL-fo1_01 CC, named sp5g-cr1/2/3/4, had a 1-bp insertion, a 2-bp insertion, a 1-bp deletion, and an 81-bp deletion, respectively; the mutant in the MM background had a 4-bp deletion (Fig. 2G). The sp5g-cr mutants in the NIL-fo1_01 CC and MM backgrounds exhibited earlier flowering under LD conditions than their respective controls (Fig. 2H). These results confirmed that SP5G is the gene in the Q-fo1_01 locus that regulates daylength sensitivity in tomato.

SP5G Was Artificially Selected during Tomato Domestication

Modern cultivated tomatoes became day-neutral plants for adaptation to broad or diverse regions of cultivation. Consistently, many of the S. pimpinellifolium species were found to be more sensitive to LD in terms of flowering time compared with the big-fruit cultivars. To assess whether the altered daylength response conferred an adaptive advantage and was selected during tomato breeding, we used the published genomic sequences of 166 big-fruit S. lycopersicum varieties, 112 cherry tomatoes, and 53 S. pimpinellifolium accessions (Lin et al., 2014) to scan for signatures of selection surrounding the SP5G gene. Notably, we found that the region around SP5G had reduced nucleotide diversity (\( \pi \)) in cherry and big-fruit tomatoes compared with wild species (Fig. 3A). The ratio of nucleotide diversity for wild to cherry tomatoes (\( \frac{\pi_{\text{wild}}}{\pi_{\text{cherry}}} \)) in this region was above 3, which was defined previously as the cutoff for a domestication sweep (Lin et al., 2014). However, the ratio for cherry to big-fruit tomatoes was below 6.9, the cutoff for designating an improvement sweep (Lin et al., 2014). These findings suggest that SP5G was positively selected during domestication but not in tomato improvement (Fig. 3B).

SP5G Expression Regulates Daylength Adaptation

To further investigate the role of SP5G in daylength responses, its diurnal expression pattern was analyzed in the NILs exposed to different daylengths. Under LD conditions, SP5G was highly expressed at 4 h after dawn (Zeitgeber time 4 [ZT4]) in NIL-fo1_01 CC and then decreased; at the end of the day, SP5G transcripts increased again and then dropped rapidly after dusk, as reported previously (Cao et al., 2016; Soyk et al., 2017; Fig. 4A). Notably, the expression peak of SP5G at ZT4 was reduced dramatically in NIL-fo1_01 MM, resulting in a relatively low and stable expression from ZT0 to ZT8 in NIL-fo1_01 MM (Fig. 4A). After dawn (Zeitgeber time 4 [ZT4]) in NIL-fo1_01 CC and then decreased; at the end of the day, SP5G transcripts increased again and then dropped rapidly after dusk, as reported previously (Cao et al., 2016; Soyk et al., 2017; Fig. 4A). Notably, the expression peak of SP5G at ZT4 was reduced dramatically in NIL-fo1_01 MM, resulting in a relatively low and stable expression from ZT0 to ZT8 in NIL-fo1_01 MM (Fig. 4A). Under SD conditions, SP5G showed extremely low expression in both NIL-fo1_01 CC and NIL-fo1_01 MM (Fig. 4A). Given that higher SP5G expression levels result in more delayed flowering time under LD conditions in tomato (Soyk et al., 2017), our findings suggest that the differences in SP5G expression under LD conditions in these two NILs may underlie their different photoperiodic responses.

To identify the sequence variations responsible for the altered SP5G expression, we compared the SP5G genomic sequences of MM and CC along with 3 kb of the upstream promoter region. Some polymorphisms
Figure 2. SP5G is responsible for photoperiod sensitivity in tomatoes. A, Schematic representation summarizing the construction of the RILs. MM indicates the big-fruit cv S. lycopersicum cv Moneymaker, and CC indicates the cherry tomato S. lycopersicum var cerasiforme (LA1310). The red and yellow bars indicate the chromosome regions with the MM and CC genotypes, respectively. B, Flowering time of the parental lines of the RILs under LD and SD conditions. Bars and lines indicate the means and sd of six individual plants. **, P < 0.01 (Student’s t test). C, Manhattan plots for flowering time under LD conditions. The black dashed line indicates the cutoff P value (1E-10). The red arrow indicates the main locus Q-fo1_01. D, Peak region of the Q-fo1_01 locus using the P value of 1E-10 (top) and the genomic region of difference between NIL-Q-fo1_01MM and NIL-Q-fo1_01CC (bottom). E, Appearance of the NILs of Q-fo1_01 under LD conditions. The white arrows indicate flowers. Bar = 10 cm. F, Flowering time of the NILs under LD and SD conditions. Bars and lines indicate means and sd of six individual plants. **, P < 0.01 (Student’s t test). G, SP5G mutations generated by CRISPR/Cas9 using two single-guide RNAs. Blue lines indicate the target sites of the guide RNAs. The nucleotides underlined in black indicate the protospacer-adjacent motif (PAM). The sequences of the four sp5g mutant alleles (sp5g-cr1/2/3/4) of NIL-Q-fo1_01MM and one allele (sp5g-cr5) of MM are shown. H, Flowering time of the sp5g mutant alleles compared with their controls under LD conditions. Bars and lines indicate means and sd of more than six individual plants. **, P < 0.01 (Student’s t test).
SP5G Expression Regulates Daylength Adaptation

To investigate how these polymorphisms affect SP5G expression, we first tested the stability of the transcripts from NIL-fo1_01\textsuperscript{MM} and NIL-fo1_01\textsuperscript{CC}. We treated seedlings of the NILs with cordycepin C, a transcriptional inhibitor that blocks mRNA synthesis (Barrett et al., 2012; Matoulkova et al., 2012), and quantified SP5G transcripts at different time points after treatment. The degradation rates of SP5G mRNA were similar between NIL-fo1_01\textsuperscript{MM} and NIL-fo1_01\textsuperscript{CC} (Supplemental Fig. S4), indicating that these variants do not influence the stability of SP5G mRNA.

Next, we investigated whether the promoter SNPs or the 52-bp InDel affect SP5G transcription using a dual luciferase reporter assay. The SP5G 3' UTR with (SP5G UTR-L) or without (SP5G UTR-S) the 52-bp sequence was fused with a firefly luciferase sequence driven by the native SP5G promoter from TS30M or TS30C. Renilla luciferase driven by the cauliflower mosaic virus 35S promoter within the same construct was used as an internal control (Fig. 5A). The constructs were expressed transiently in tobacco (Nicotiana tabacum) leaves, and luciferase activity was detected to calculate the ratio of firefly luciferase to Renilla luciferase activity as an indicator of the transcriptional efficiency of SP5G. Regardless of the promoter used (TS30M or TS30C), the ratio of firefly luciferase to Renilla luciferase activity was much higher for constructs containing the full-length 3’ UTR compared with constructs lacking the 52-bp sequence (Fig. 5, B and C). This finding indicated that the loss of the 52-bp sequence in the SP5G 3’ UTR greatly reduced its expression. The relative signal of firefly luciferase driven by the TS30C promoter was slightly higher than that driven by the TS30M promoter, but the difference was not significant when using the full-length 3’ UTR, suggesting that the polymorphisms in the SP5G promoter have little effect on its expression (Fig. 5C). The analysis indicated that the 3’ UTR is an enhancer that regulates SP5G transcripts and that the 52-bp sequence is essential for its enhancer activity. Notably, the SP5G
UTR did not affect the activity of the 35S minimal promoter, as there was no significant difference in relative firefly luciferase activity between constructs with the full-length or truncated 3′ UTR driven by the 35S minimal promoter (Supplemental Fig. S5). Thus, the enhancer activity of the 3′ UTR appears to be specific to the SP5G native promoter.

In plants, active enhancers are associated with specific histone modifications, such as H3K27 acetylation (Zhu et al., 2015; Oka et al., 2017). We examined the distribution of H3K27ac around the SP5G region by chromatin immunoprecipitation (ChIP) in TS30M and TS30C lines under LD conditions. Low levels of H3K27ac were observed at the transcription start site (TSS) and gene body of SP5G in TS30M and TS30C under LD conditions (Fig. 5D). However, there was a significant enrichment of H3K27ac at the SP5G 3′ UTR in TS30C compared with TS30M (Fig. 5D), supporting the existence of a functional enhancer at this region in TS30C. Moreover, H3K27ac abundance at the SP5G 3′ UTR in TS30C was reduced dramatically under SD conditions (Fig. 5D), indicating that the enhancer activity was dependent on daylength.

We then grouped the tomato accessions by the presence (noted as C) or absence (noted as M) of the 52-bp sequence within the SP5G 3′ UTR (Supplemental Fig. S6A). Association analysis revealed that the relative flowering time of these accessions was highly correlated with their SP5G genotype rather than their species (Supplemental Fig. S6, B and C). For example, TS433 is a wild species with the MM SP5G genotype, and this line was more insensitive to daylength than other wild lines (Supplemental Fig. S6B).

### The 52-bp Sequence in the SP5G 3′ UTR Is Essential for Promoter-Enhancer Interaction

In eukaryotes, mRNA is transcribed by RNA polymerase II (PolII), which can be recruited to an enhancer region to facilitate promoter-enhancer loop formation.
and subsequently enhance target gene transcription (Chen et al., 2017). To detect whether PolII is recruited to the SP5G 3′ UTR, we performed ChIP using antibodies against PolII in TS30M and TS30C lines at ZT4 under LD conditions. Our results showed that PolII accumulation in the TSS region of SP5G was similar between the two lines (Fig. 6A). However, PolII accumulation at the SP5G 3′ UTR around the 52-bp region was significantly greater in TS30C than in TS30M under LD conditions (Fig. 6A), indicating that the 52-bp sequence is vital for PolII enrichment at the SP5G 3′ UTR. Under SD conditions, PolII accumulation was low (Fig. 6A), consistent with the finding that SP5G was hardly transcribed under SD (Fig. 4A).

Numerous studies have shown that an enhancer can form a loop structure with the promoter of its target gene and activate its transcription (Levine et al., 2014; Mora et al., 2016; Weber et al., 2016). To check whether a loop structure is formed between the SP5G 3′ UTR and promoter, we performed a chromosome conformation capture (3C) assay, a technique that detects distal DNA interactions (Dekker et al., 2002), in TS30M and TS30C in TS30C at ZT4 under LD conditions. With the sequence containing the 52-bp InDel as the anchor point, we tested the cross-linking interactions between this region and various DpnII-released DNA fragments. A peak was detected between the anchor point and fragment 3 near the start codon (Fig. 6B), suggesting a physical interaction between the SP5G 3′ UTR and the promoter. The strength of the interaction was weakened significantly in TS30M compared with TS30C (Fig. 6B), indicating that the 52-bp sequence in the SP5G 3′ UTR is crucial for the promoter-enhancer interaction. Moreover, the interaction between the SP5G 3′ UTR and promoter was reduced dramatically under SD conditions, indicating that the regulatory role of the SP5G enhancer is dependent on daylength (Fig. 6B). Based on these observations, we propose the following model of enhancer-mediated photoperiodic sensitivity in tomato. In wild tomatoes, the SP5G 3′ UTR contains
a functional enhancer that forms a loop with the SP5G promoter and activates the expression of the gene under LD conditions. In tomato cultivars, the enhancer is disrupted by the loss of the 52-bp sequence, resulting in weakened SP5G expression under LD conditions and consequent photoperiodic insensitivity (Fig. 6C).

DISCUSSION

In contrast to SD wild tomato species, modern cultivars have reduced photoperiodic sensitivity and are day neutral. Consistent with a recent report that altered SP5G expression contributes to different daylength responses in tomato (Soyk et al., 2017), our QTL analysis identified SP5G as a major locus regulating daylength sensitivity in tomato that has been subject to selection during tomato domestication. We determined that a 52-bp sequence within the SP5G 3′UTR is essential for the enhanced expression of SP5G under LD conditions. Moreover, the SP5G 3′UTR is an enhancer capable of forming a loop structure with the SP5G promoter, as reported for other enhancers (Sakabe et al., 2012; Weber et al., 2016). The 52-bp deletion within the 3′UTR in many big-fruit cultivars weakened the enhancer-promoter physical interaction, thereby decreasing SP5G expression and causing daylength insensitivity. PolII accumulation at the promoter and gene body of SP5G was similar in TS30C and TS30M under LD conditions, which did not reflect the different SP5G expression levels observed between these two lines. Gene expression is the output of the whole transcription cycle, including transcription initiation, elongation, and termination (Hajheidari et al., 2013). PolII carries out transcription throughout the genome and undergoes regulated pauses and releases at enhancers to resist or promote termination, effectively a mechanism for gene transcription regulation (Porrua and Libri, 2015; Henriques et al., 2018). Thus, the loss of the 52-bp sequence and the consequent reduction of PolII occupancy at the SP5G 3′UTR may have led to improper transcription termination of SP5G and reduced its expression. In eukaryotes, posttranslational modifications of the C-terminal domain, the largest PolII subunit, are crucial for the regulation of multiple steps in transcription (Harlen and Churchman, 2017). In addition to the occupancy of PolII, detecting differences in the phosphorylation status of the C-terminal domain along the gene may help us better understand the transcription dynamics of SP5G.

Transcription factors (TFs) can be recruited to enhancers to facilitate promoter-enhancer interactions or mediate the interaction between PolII and DNA (Kim et al., 2010; Levine et al., 2014). Loss of the 52-bp sequence at the SP5G 3′UTR enhancer may have led to the reduced occupancy of regulatory TFs potentially important for promoter-enhancer loop formation or acting with PolII to activate SP5G expression. Our observation that the decreased SP5G expression level in TS30M was much more drastic than the decreased promoter-enhancer interaction suggests that the promoter-enhancer interaction is not linearly correlated with SP5G expression. One reason may be that more TFs are required for the transcription of SP5G rather than as mediators of the promoter-enhancer loop. In fact, motifs recognized by distinct TFs were found around the 52-bp sequence, some of which participate in flowering regulation in Arabidopsis (Supplemental Table S3). Thus, the future identification of regulatory proteins that bind to the SP5G 3′UTR will improve
our understanding of the photoperiodic regulation of flowering in tomato.

It should be noted, although SP5G expression was lower due to the absence of the 52-bp sequence in the 3′ UTR enhancer, SP5G was still induced in cultivated tomatoes under LD, which conferred a weak photoperiodic sensitivity. In agreement with the SP5G transcript levels, H3K27ac and PolII enrichment also were observed in TS30M under LD conditions. Additionally, the enhancer-promoter loop was weakened but not abolished in TS30M. These results suggest that the absence of the 52-bp sequence did not fully disrupt the enhancer activity. Consistent with this hypothesis, the SP5G knockout mutants generated by CRISPR/Cas9 in the MM background flowered earlier than MM under LD, indicating that SP5G still functions in cultivated tomatoes. We also observed that TS30H plants (heterozygous at SP5G) displayed intermediate daylength sensitivity between TS30M and TS30C under LD (Supplemental Fig. S7). These results suggest that SP5G regulates flowering time in a dose-dependent manner and that the cis-regulatory variants of SP5G cause subtle phenotypic changes by modifying the expression of SP5G. In fact, transcriptional changes causing gene dosage effects are often favored by breeders (Meyer and Purugganan, 2013). In maize, tb1 encodes a transcriptional regulator that represses branch elongation (Doebley et al., 1997). Its expression level in the wild teosinte (Zea spp.) allele is approximately half the level of the maize allele, and the difference is attributable to an upstream transposon insertion that acts as an enhancer (Clark et al., 2006; Studer et al., 2011; Zhou et al., 2011). The fto2 locus in tomato contributes up to 30% of the fruit weight increase observed in big-fruit cultivars versus wild tomatoes (Frary et al., 2000), and the increase may have been caused by nucleotide changes in the upstream promoter region leading to an altered expression profile (Nesbitt and Tanksley, 2002). Thus, cis-regulatory alleles provide beneficial variation for breeding in many crops. In a recent report, genome editing by CRISPR/Cas9 was used to modify the regulatory sequences of genes to obtain novel alleles affecting some important traits (Rodriguez-Leal et al., 2017). Finding novel regulatory elements of genes controlling important agronomic traits will facilitate the development of favorable alleles for breeding. In this context, our work characterizes a new target that could potentially be used to further modify tomato flowering time by manipulating the expression of SP5G through cis-regulatory variation.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

RILs were derived by single-seed descent from a cross between the cultivated tomato (Solanum lycopersicum ‘Moneymaker’) and the cherry tomato accession Solanum lycopersicum var. cerasiforme (LA3310). All plants were self-fertilized to the F10 generation, and a RIL population containing 219 lines was derived. For flowering time assessment, all tomato plants were grown in soil in a greenhouse. The flowering time was evaluated as the leaf number to the first inflorescence under LD conditions (Beijing, 40°N, from July to August, 14 h of light on average) and SD conditions (Beijing, 40°N, from February to March, 11.5 h of light on average). At least six individual plants of each accession were used for flowering time assessment.

Genome Sequencing, SNP Calling, and the RIL Association Study

The genomic DNA of MM, CC, and the RIL population was extracted following the standard CTAB method and prepared for sequencing on the Illumina HiSeq 2500 system with a 125-bp paired-end strategy. The sequencing depth was about 15-fold coverage for the two parental lines and about 1.7-fold coverage for the 219 RILs. The paired-end reads of MM, CC, and the RILs were mapped to the tomato reference genome (SL2.50 build; Tomato Genome Consortium, 2012) using Burrows-Wheeler Aligner version 0.7.10–789 with default parameters (Li and Durbin, 2009). SNP calling was performed on the alignment results using the Genome Analysis Toolkit (GATK) version 3.1.1 and Picard package version 1.119 (McKenna et al., 2010) with the following steps: (1) deleting the unmapped reads, (2) deleting the duplicate reads, (3) conducting the alignment using the IndelRealigner package in GATK, and (4) SNP calling for each sample using theUnifiedGenotyper package in GATK with a minimum base quality score of 20. To ensure the quality of SNP calling in MM and CC, SNPs were filtered further with the VariantFiltration package in GATK using parameters QD < 2.0 || FS > 60.0 | MQ < 40.0 | MQRankSum < −12.5 | ReadPosRankSum < −8.0, and SNPs with DP < 10 also were removed. The remaining SNPs between MM and CC were used as the index for SNP calling in the 219 RILs. The SNPs in the 219 RILs were filtered with the LowQual tag marked by the GATK UnifiedGenotyper package. To infer the missing genotype at an SNP site of an individual line, 20 SNPs flanking the target SNP in other lines of the population were compared with the individual line. If all of the lines having the same genotype for the flanking SNPs as the individual line had the same genotype at the target SNP, this SNP genotype was inferred/imputed to the individual line. The RIL association study was conducted using the imputed information for the SNPs of the 219 RILs by the compressed mixed linear model implemented in GAPIT (Lipka et al., 2012). The cutoff P value was set as 1E-10.

Identification of Selective Sweeps

To identify the selection region around SP5G, the SNPs near the SP5G locus in the tomato genome were obtained (corresponding to chromosome 9: 63.3–64.5 Mb, SL2.50). We measured the level of nucleotide diversity (π) using a 100-kb window with a step size of 10 kb in wild species (wild), cherry tomatoes (cherry), and big-fruit cultivars (big). The ratios of nucleotide diversity between wild and cherry (πwild/πcherry) and cherry and big (πcherry/πbig) also were calculated. According to the strategy described in a previous study (Lin et al., 2014), the top 5% of ratios were used as the cutoff for sweeps.

RNA Stability Assay

More than five 18-d-old tomato seedlings under LD conditions were collected at ZT4 in each 50-ml tube and incubated with 16 ml of incubation buffer (1 mM PIPES, pH 6.25, 1 mM sodium citrate, 1 mM KCl, and 15 mM sucrose) at 75 rpm for 30 min at room temperature. Cordycepin C was added to a final concentration of 0.5 mM. The seedlings were incubated on a shaker at 75 rpm and harvested at 0, 20, 40, 60, 80, 100, and 120 min after treatment with cordycepin C. The materials were frozen in liquid nitrogen and stored at −80°C until use.

RNA Extraction and Quantitative PCR

RNA was extracted using TRNzol Universal reagent (Tiangen; DP424). DNA contamination was removed using the TURBO DNA-free Kit (Ambion; AM1907). Reverse transcription was performed with Transcript II First-Strand CDNA Synthesis SuperMix (Transgen; AH301) using 2 μg of total RNA. Quantitative PCR (qPCR) was performed with the KAPA SYBR FAST Universal qPCR Kit (Kapa; KK4601) on a Bio-Rad CFX96 Real-Time PCR instrument.
using the following program: 3 min at 95°C followed by 40 cycles of 20 s at 95°C, 30 s at 60°C, and 20 s at 72°C. U6B3 (Solyc01g056400) was used as the internal control for reverse transcription-qPCR. The primers used for qPCR are listed in Supplemental Table S4.

**Dual Luciferase Reporter Assay**

For plasmid construction, the plus-35Rluc backbone vector was obtained from pZP211 (Hajdukiewicz et al., 1994). The nearly 1-kb SP5G 3′ UTR with the 52-bp sequence was amplified from LA1310 genomic DNA and integrated into plus-35Rluc using the Infusion HD cloning kit (Clontech; 639649). For the plasmids with the short version of the 3′ UTR lacking the 52-bp sequence, primers XP1347 and XP1348 were used with the plasmid containing the full-length 3′ UTR as template. The 4.17-kb promoter of SP5G was amplified using TS30M and TS30C genomic DNA as templates, then integrated into the constructs upstream of the luc sequence using the Infusion HD cloning kit (Clontech; 639649). The primers used for the constructs are listed in Supplemental Table S4. The plasmids were transformed into Agrobacterium tumefaciens EHA105 competent cells. A single colony was cultured in Luria-Bertani medium until the OD_{600} value reached 1. The A. tumefaciens cells were collected by centrifugation and suspended with 10 mM MgCl2 and 150 μM acetosyringone. The cell suspensions containing the luciferase construct and the p19 plasmid were mixed at equivalent volumes and infiltrated into tobacco (Nicotiana tabacum) leaves using a syringe. The leaves were harvested and ground in liquid nitrogen at 2°C after infiltration. The activities of firefly luciferase and Renilla luciferase were measured using the Dual-Luciferase Reporter Assay System (Promega; E1910) on a Promega GLomax 20/20 LUMINOMETER device following the manual instructions.

**ChIP and qPCR**

The ChIP assay was performed using 1 g of 18-d-old seedlings collected at ZT4 under LD conditions as described previously with minor modifications (Zhang et al., 2015). In brief, the seedlings were cross-linked with formaldehyde, and the nuclei were isolated. The chromatin was sheared using a Diagenode Bioruptor Plus instrument to obtain ~300-bp fragments. The anti-H3K-27ac (Merck; 07-360) and anti-RNA PolII (Abcam; ab817) antibodies were used for immunoprecipitation. The ChIP DNA was extracted after reverse cross-linking, qPCR was performed using the KAPA SYBR Fast Universal qPCR Kit (Kapa; KK4601) on a Bio-Rad CFX-96 Real-Time PCR instrument using the following program: 3 min at 95°C followed by 40 cycles of 20 s at 95°C, 30 s at 60°C, and 20 s at 72°C. The region around the transcription start of the three genes was amplified separately, double digested with endonucleases, and then ligated, creating a pool of all possible ligation products for the locus of interest, correcting for differences between the templates. To eliminate the differences in primer set efficiency, 5 μg of the control plasmid (CP268) was amplified as the internal control to correct for differences between the templates. To eliminate the differences in primer set efficiency, 5 μg of the control plasmid (CP268) was used as an internal control. The primers used for ChIP-qPCR are listed in Supplemental Table S4.

**3C Assay**

The 3C assay was performed as described previously (Hagège et al., 2007; Louwers et al., 2009). Eighteen-day-old seedlings collected at ZT4 under LD conditions were cross-linked using formaldehyde, and nuclei were isolated. The cross-linked chromatin was digested overnight using DpnII (New England Biolabs; R0543), followed by ligation with T4 DNA ligase (Thermo Fisher; E0103) in a large volume and reverse cross-linking overnight. Ligated DNA was extracted by phenol-chloroform and precipitated with 3 volumes of ethanol. A ligated fragment of EF1α (Solyc03g078400) was amplified as a control. The primers used for the ChIP-qPCR are listed in Supplemental Table S4.

**Accession Numbers**

Sequence data from this article can be found in Sequence Read Archive under accession numbers SRP903370 and SRR2391865. The names and accession numbers of all genes mentioned in this work can be found in Supplemental Table S4.

**Supplemental Data**

The following supplemental materials are available.

- **Supplemental Figure S1.** Manhattan plots for flowering time under SD conditions.
- **Supplemental Figure S2.** Genotyping results showed the difference of nearly 50 kb between the NILs.
- **Supplemental Figure S3.** SP5G sequence polymorphisms between MM and CC from 3 kb upstream of the start codon to the 3′ UTR.
- **Supplemental Figure S4.** The mRNA stability analysis of SP5G showed no difference between NIL-Q-fo1_01MM and NIL-Q-fo1_01CC.
- **Supplemental Figure S5.** The 3′ UTR of SP5G does not display enhancer activity when driven by the 35S minimal promoter.
- **Supplemental Figure S6.** SP5G controls the photoperiodic response in tomato.
- **Supplemental Figure S7.** Flowering time of TS30M, TS30C, and the heterozygous progeny of TS50 under LD and SD conditions.
- **Supplemental Table S1.** Markers used for comparing the background between the NILs.
- **Supplemental Table S2.** Polymorphisms within SP5G in 40 tomato accessions.
- **Supplemental Table S3.** TF-binding motifs around the 52-bp sequence in the SP5G 3′ UTR.
- **Supplemental Table S4.** List of primers.

**ACKNOWLEDGMENTS**

We thank Dr. Sanwen Huang and Dr. Jianchang Gao for donating seeds of the wild, cherry, and big-fruited tomatoes. We also thank Dr. Guodong Yang for providing pZP211 plasmids, Dr. Lin Guo for technical help on the 3C assay, and Dr. Haiyang Wang for advice on article improvement. Received September 17, 2018; accepted September 17, 2018; published October 10, 2018.

**LITERATURE CITED**

Andrés F., Coupland G. (2012) The genetic basis of flowering responses to seasonal cues. Nat Rev Genet 13: 627–639
Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslatable gene regions and other non-coding elements. Cell Mol Life Sci 69: 3633–3634
Baurle I, Dean C. (2006) The timing of developmental transitions in plants. Cell 125: 655–664
Beales J, Turner A, Griffiths S, Snape JW, Laurie DA (2007) A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theor Appl Genet 115: 721–733
Blanca J, Calizares J, Cordero L, Pascual I, Diez MJ, Nuez F (2012) Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. PLoS ONE 7: e48198
Cao K, Cui L, Zhou X, Ye L, Zou Z, Deng S (2016) Four tomato FLOWER-ING LOCUS T-like proteins act antagonistically to regulate floral initiation. Front Plant Sci 6: 1213
Chen H, Du G, Song X, Li L (2017) Non-coding transcripts from enhancers: new insights into enhancer activity and gene expression regulation. Genomics Proteomics Bioinformatics 15: 201–207
Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, (2008) Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet 40: 761–767
Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T, (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc Natl Acad Sci USA 110: 16969–16974
Zhang S, Zhou B, Kang Y, Cui X, Liu A, Deleiris A, Greenberg MV, Cui X, Qiu Q, Lu E, (2015) C-terminal domains of a histone demethylase interact with a pair of transcription factors and mediate specific chromatin association. Cell Discov 1: 15003
Zhou L, Zhang J, Yan J, Song R (2011) Two transposable element insertions are causative mutations for the major domestication gene teosinte branched 1 in modern maize. Cell Res 21: 1267–1270
Zhu B, Zhang W, Zhang T, Liu B, Jiang J (2015) Genome-wide prediction and validation of intergenic enhancers in Arabidopsis using open chromatin signatures. Plant Cell 27: 2415–2426