UNIVERSITY OF CALIFORNIA SAN DIEGO

CRISPR-Cas9 mediated gene editing of promoter cis-elements in core Arabidopsis clock genes

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Andy Tran

Committee in charge:

Professor Jose L. Pruneda-Paz, Chair
Professor Martin F. Yanofsky, Co-chair
Professor Julian I. Schroeder

2021

The thesis of Andy Tran is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Jose Pruneda-Paz for being my mentor and guiding me throughout the end of my undergraduate studies and the entirety of my graduate studies. His guidance and patience despite all my questions has helped me immensely to grow. His trust in me to do independent tasks has helped me learn to self-guide myself and think like a scientist.

I would like to thank Dr. Jose Pruneda-Paz, Dr. Martin Yanofsky and Dr. Julian Schroeder for agreeing to be my committee members and supporting me throughout the graduate program.

I would like to acknowledge all the members of the Pruneda-Paz lab for giving me guidance whenever I got stuck doing a task and answering my general questions about anything. They helped me learn that I should not be afraid of asking questions and that a lab environment is an amazing place for collaboration.

I would like to acknowledge everyone in the Yanofsky lab for accepting me into my first laboratory experience and teaching me the basics to build my foundation as a scientist.

Finally, I would like to acknowledge my family and my friends for believing in me and always pushing me to greater heights. Without them, I would not be where I am today.

ABSTRACT OF THE THESIS

CRISPR-Cas9 mediated gene editing of promoter cis-elements in core Arabidopsis clock genes

by

Andy Tran

Master of Science in Biology

University of California San Diego, 2021

Professor Jose-Pruneda Paz, Chair
Professor Martin Yanofsky, Co-Chair

The circadian clock in plants coordinates internal biological processes to exhibit daily oscillations that are in synchrony with environmental rhythms. The roles of core clock genes *CCA1, LHY*, and *TOC1* that regulate the molecular mechanism have already been well established in *Arabidopsis thaliana* and the mechanism by which TOC1 regulates CCA1 and LHY appears to be mediated by transcription factors of the TCP family. However, the role of cis-elements in the promoters of these clock genes

has not been fully addressed. In this thesis, we utilized CRISPR/Cas9 technology to create mutations within the *CCA1*, *LHY* and *TOC1* promoter regions. Through this, we hope to better understand the role of cis-elements in the promoters of core clock genes and how they influence the clock feedback regulatory mechanism. We propose a procedure to efficiently identify genome editing when Cas9 is expressed in early gametogenesis stages. Using this procedure, we generated indel mutations in the *CCA1* and *LHY* promoters and found that one of them resulted in a short hypocotyl phenotype. However, the visual observation of mutant plants suggested that off-target mutations may be present, thus backcrossing is required before further analysis. As such, our results suggest that cis-elements in the promoters of core clock genes may play an important role in modulating the clock function and therefore, clock-controlled processes.

Introduction

Throughout the entirety of a day and night cycle, organisms such as animals, plants, and even cyanobacteria have adapted to constantly alter metabolic processes depending on the time of day (Bell-Pedersen et al. 2005). The changing of metabolic processes within the organism is possible due to the existence of an endogenous mechanism that many organisms possess, the circadian clock (Michael et al. 2003). From this internal clock, organisms can predict environmental changes in a diurnal rhythm, a repeating 24-hour cycle, to better coordinate their physiology to maximize their survivability (Michael et al. 2003). The circadian rhythm follows the internal circadian clock, allowing organisms to maintain self-sustained 24h oscillations of clock-controlled processes. This is largely achieved at the transcriptional level which results in genes being expressed rhythmically throughout the day and night cycle (Pruneda-Paz and Kay 2010).

Plants are a key organism in utilizing the circadian rhythm mechanism to enhance their survival as they are sessile in an environment that is periodically changing (Žádníková et al. 2015). The circadian clock is vital for the plant's growth as the internal clock has been found to affect the physiological responses in a multitude of different stages of the plant's development with some of them being stress response, growth, flowering time, and metabolism (Nohales and Kay 2016). For example, in *Oenothera biennis* (Evening Primrose), its flowers are opened or closed at specific times throughout the 24-hour day and night cycle to optimize the reproductive time where the plant can accept the pollen by opening the flower in the morning and then protect the incoming pollen by closing the flower at night (Van Doorn and Meeteren

2003). Another example is in the model organism, *Arabidopsis thaliania,* where its leaves are observed to be lowered in the daytime but point upwards during the night, most likely to optimize light absorption at different points of the day (Edwards and Millar 2007). Overall, using plants to study the circadian clock is advantageous because plants adapt to changes in light, temperature, and nutritional availability in its existing environment by regulating the plant's reproductive capabilities, physiological responses, and developmental growth to specific times of day (Covington, et al. 2008).

The basis of a circadian clock is built upon three different modules, the signal input from the environment, the oscillator mechanism, and the output pathways (Webb 2003). The signal input would be the environmental cues such as light or temperature which helps control the internal clock by keeping the clock and the environment synchronized (Nohales and Kay 2016). The oscillator mechanism is built upon a rhythmic cycling of gene expression that can stay self-sufficient through a feedback loop where an environmental cue causes a gene to produce a protein that would activate another gene that would then repress the initial gene, creating a repeating cycle where different genes will be expressed at different times of day (Webb 2003). This allows certain output signals to be expressed in a rhythmic manner such as leaf movement due to the oscillating expression of genes (Webb 2003).

This study focuses on the plant, *Arabidopsis thaliana,* and its circadian clock which utilizes a transcriptional feedback loop with three core clock genes, CIRCADIAN CLOCK ASSOCIATED 1 (CCA1), LATE ELONGATED HYPOCOTYL (LHY), and TIMING OF CAB EXPRESSION (TOC1) (Alabadi et al. 2001). These three core genes are expressed in most plant tissues such as leaves, roots, and stems to maintain the

functionality of the circadian clock throughout the plant (Alabadi et al. 2001). The circadian clock in *Arabidopsis* works around a 24-hour cycle and is regulated by the usage of the clock genes stated above that activate each other in succession to create an oscillating expression of the genes throughout the day where one gene can peak during the day and be repressed at night, or vice versa (Albrecht et al, 2009).

Particularly, the clock in *Arabidopsis* revolves around a reciprocal regulation starting with CCA1 and LHY, partially redundant morning expressed transcription factors that are a part of the MYB gene family and bind to the promoter region of the gene encoding the evening component, TOC1, and represses its expression (Ambawat et al. 2013). LHY and CCA1 represses TOC1 during the day through binding interactions with a motif within the promoter region, the evening element, or EE (AAAATATCT) (Alabadi et al. 2001). This motif was initially discovered in the nucleotides -734 to -687 of the *TOC1* promoter and its mutation resulted in a significant reduction in the clock rhythmicity (Alabadi et al. 2001). However, this study only analyzed the function of the EE in the context of a relatively short promoter construct, thus it is not known if the inactivation of the evening alone could result in the loss of rhythmicity of the clock as the upstream components in the promoter region could also play a role in regulating the clock function (Alabadi et al. 2001). As the day progresses, the CCA1 and LHY proteins will begin to degrade and reach their lowest levels, allowing TOC1 and other genes in the PRR (Pseudo response regulator) family that such as PRR9, PRR7, PRR3 to increase in expression respectively as their repression is alleviated (Ono et al. 2019). TOC1 is a negative regulator of CCA1 and LHY that keeps their expression levels reduced until the next morning when TOC1 protein is degraded,

allowing the circuit to begin again (Pruneda-Paz and Kay 2010). However, both an increase or decrease of TOC1 expression results in a decrease of CCA1 and LHY expression which indicates that there is a more complex mechanism behind the regulation of CCA1 and LHY (Alabadi et al. 2001, Makino et al. 2002).

One such interaction of CCA1 by TOC1 was previously discovered to occur due to the protein-protein interaction between TOC1 and a class I TCP transcription factor, CCA1 HIKING EXPEDITION (CHE), which allows for the targeting of a TCP-binding domain within the *CCA1* promoter where CHE can repress CCA1 expression (Pruneda Paz et al. 2009). However, CHE has not been found to bind to the *LHY* promoter so it is hypothesized that a similar mechanism is utilized by TOC1 with one of its binding partners to regulate LHY. With the discovery of the involvement of class I TCP transcription factors in the regulation of the clock, the question arises whether class II TCP transcription factors also play a role in clock regulation as they are closely related and the promoter regions of *CCA1* and *LHY* contain binding sites that can be recognized by class II TCP transcription factors.

The introduction of CRISPR/Cas9 technology has enabled greater ease of use and efficiency for editing in precise locations of the genome (Ran et al. 2013). For CRISPR/Cas9 genome editing to occur, two main components, Cas9 and gRNA, are required to be introduced into the organism to create a double stranded break in a targeted region (Ran et al. 2013). This eventually results in one of two DNA repair mechanisms to occur, non-homologous end joining (NHEJ) or template based homology directed repair, potentially producing mutations at the site of the cut (Ran et al. 2013).

To build upon existing studies, we will focus on functionally characterizing the cis-regulatory DNA motifs in the *CCA1*, *LHY*, and *TOC1* promoters by mutating the putative binding sites of various transcriptional regulators of the circadian clock at different locations to observe the effects on the circadian clock. Through the utilization of CRISPR/Cas9 technology, the mutations created from NHEJ will target the evening element in the *TOC1* promoter, the TCP binding site in the *CCA1* promoter, and the TCP-II binding sites in both *CCA1* and *LHY* promoters. Furthermore, the circadian clock and developmental phenotypes of the newly generated CRISPR/Cas9 mutants will be compared with the previous clock mutants created to validate previous findings and confirm the role of core clock components. Through the process, a streamlined procedure to efficiently genome edit is proposed to optimize the screening of CRISPR/Cas9 mutants when Cas9 protein is expressed at early gametogenesis stages.

Results

Streamlined procedure for identifying CRISPR/Cas9 mutants

To induce mutagenesis in the *CCA1*, *LHY*, and *TOC1* gene promoters through CRISPR/Cas9, constructs were generated by using P3J1, a binary vector where Cas9 is driven by an egg cell and early embryo *DD45* promoter. Single or double gRNAs can be expressed with this vector where one gRNA is driven by the *U6-26* promoter and the second gRNA is driven by the *U6-29* promoter.

To identify mutagenesis in the target regions of the clock genes in *Arabidopsis* through CRISPR/Cas9, we first identified T1 plants carrying the gene editing construct and genotyped each line individually. However, no evidence of genome editing was observed after testing 70 individual T1 plants carrying a single gRNA construct targeting the TBS-II site in the *CCA1* promoter. Similar results were also obtained after testing 48 T1 individuals carrying two gRNA constructs. We hypothesized that due to the *DD45* promoter activity at the early embryo stages, the analysis of T2 plants would provide better results. As such, seeds from individual T1 plants were collected and the progeny of the respective T1 individual were separated into pools. Instead of analyzing individual T2 plants at first, we extracted genomic DNA from 8 plants within a T2 pool for PCR and gel electrophoresis analysis. From this strategy of analyzing T2 plants, we were able to identify evidence of mutagenesis in 6/40 pools in CCA1pro_TBSII (CRISPR #13) and 10/10 pools in LHYpro_ΔTBS-II (CRISPR #15) (Table 3). DNA from the pools of T2s were then sent for Sanger sequencing and those that resulted in potential mutants were deconvoluted to identify the individual T2 plant that underwent mutagenesis. The T3

generation from the individual T2 plant that contained the mutant gene was followed to isolate Cas9 free plants by selecting for hygromycin sensitive plants and identifying plants that showed homozygosity for the intended mutation. With the success of identifying plants with mutagenesis through this methodology, we designed a general streamlined procedure to follow for the remaining transgenic lines to efficiently analyze a large amount of T2 plants when the Cas9 protein is expressed at early gametogenesis phases (Figure 1).

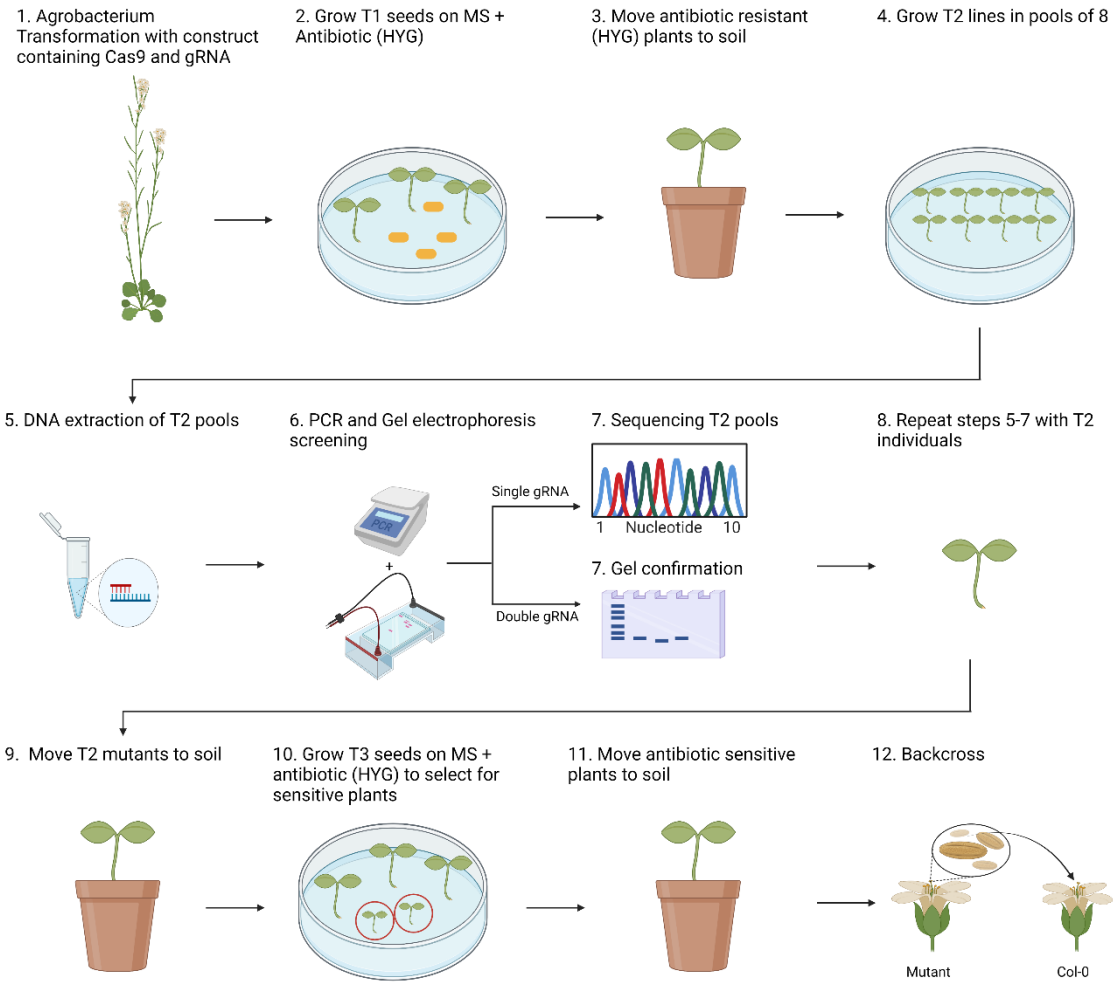


Figure 1: Procedure to identify CRISPR/Cas9 mutants. Outline of *Arabidopsis* transformation and screening strategy. Constructs containing the specific promoter-driven Cas9 nuclease and gRNAs (single or dual) targeting genes of interest were transformed into wild-type *Arabidopsis* Col-0 plants. T1 transgenic plants were selected with HYG and T2 seeds were collected to be grown in pools of 8 for DNA extraction and sequencing analysis. Sequencing analysis is repeated for all individuals in T2 pools with potential mutagenesis. T3 seeds from verified mutants are selected for hygromycin sensitivity and transferred to soil to be backcrossed with wild-type *Arabidopsis* Col-0 for further experiments. Figure was created from images that were obtained and edited from the Biorender database.

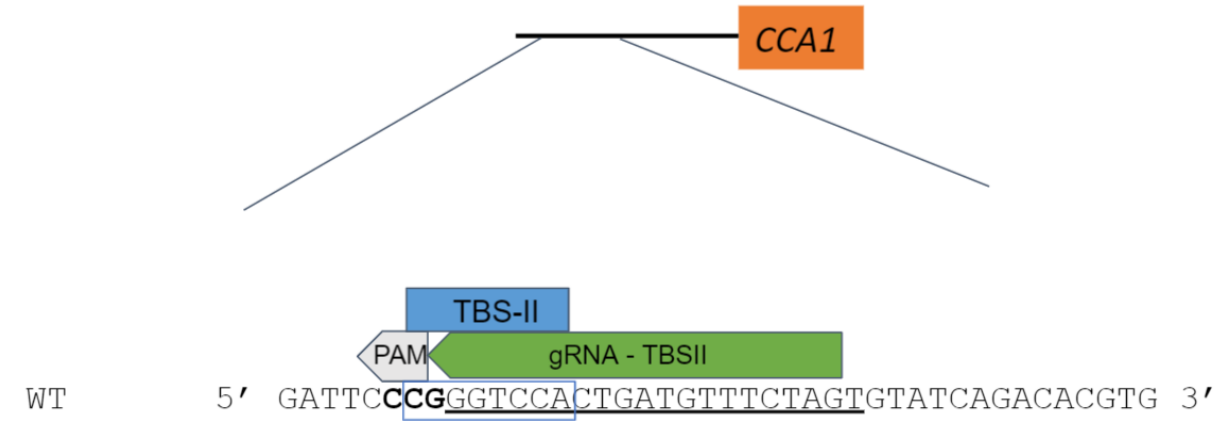
Targeted mutagenesis in promoter regions of *CCA1*, *LHY*, and *TOC1* genes

To target the TCP-II binding site in the *CCA1* promoter, the transgenic line CCA1pro_TBSII (CRISPR #13) was generated that utilizes a single gRNA that overlaps with the TCP-II binding site where the PAM site would allow Cas9 to cut within the binding site (Figure 2A). 5 out of 28 of the T2 pools screened showed heterozygosity for mutagenesis at the target region in the TCP-II binding site in the *CCA1* promoter and after analysis of the individual plants that contained the mutations, 5 different indel mutations were identified (Figure 2B and 2C). Furthermore, one CRISPR #13 T2 pool showed a deletion of all bases in the promoter sequence downstream of the cut site (Figure 2C).

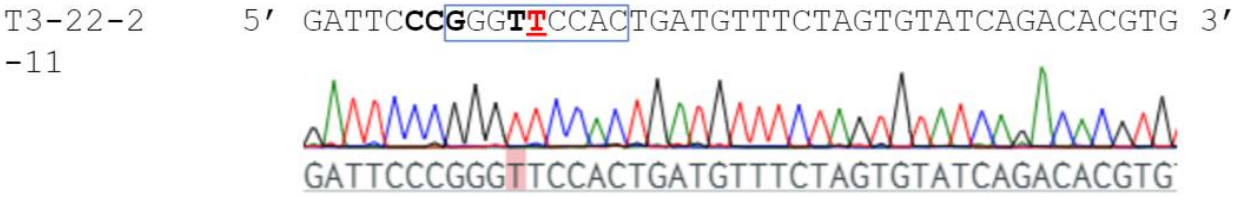
In the *LHY* promoter, an *Arabidopsis* transgenic line, LHYpro_ΔTBS-II (CRISPR #15), was generated utilizing a dual gRNA CRISPR/Cas9 system that flanks the TCP-II binding site (Figure 3A). To confirm if the deletion was successful, PCR amplification was done with primers spanning a 260 bp region that contains the TCP-II binding site. From the screening and Sanger sequencing of the T2 pools, 1 of the 10 T2 pools were found to be heterozygous for a 31 bp deletion between (find number) in the *LHY* promoter region (Figure 3A). From this deletion, the entirety of the TCP-II binding site in the *LHY* promoter is removed (Figure 3A). Through agarose gel electrophoresis, a clear difference in band length can be seen between the wild-type Col-0 band that was at around 260 bp whereas the all bands for the transgenic CRISPR #15 lines T4-1-6-1 and T4-1-6-5 samples were at around 230bp (Figure 3B).

A dual gRNA system was again used in an attempt to delete the evening element within the TOC1 promoter in the transgenic line TOC1pro_ΔEE (CRISPR #17). If successful, a ~50 bp deletion would occur that includes the entirety of the evening element (Figure 4). After screening through 115 T2 pools, none of them resulted in the deletion of the evening element.

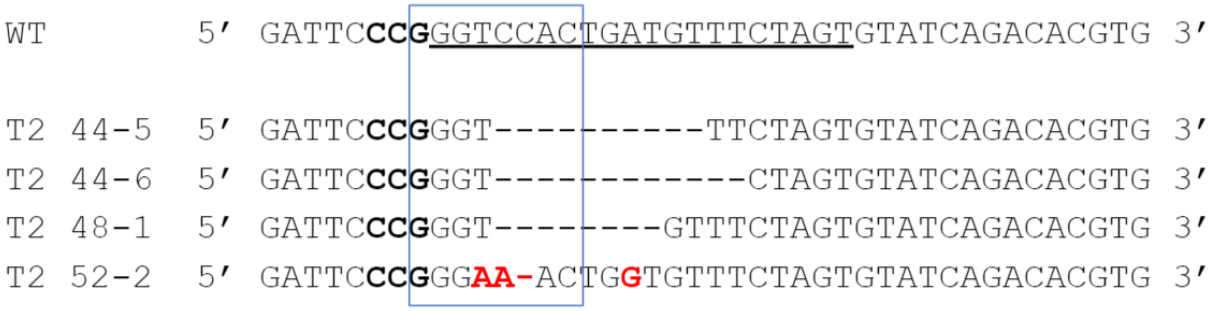
A.



B.



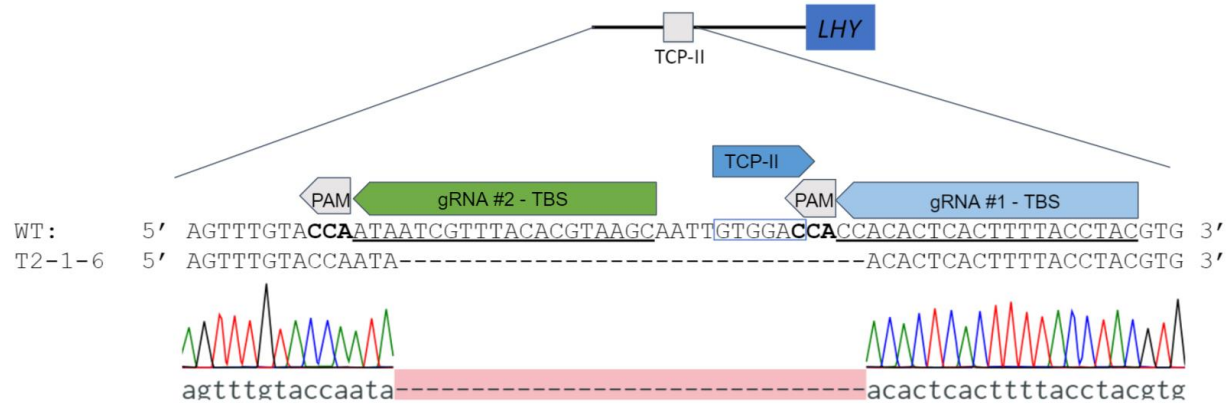
C.



D.

Figure 2: Identification of TBS-II indel mutations in the *CCA1* promoter. (A) Diagram of the wild type *CCA1* promoter sequence showing the single guideRNA binding site (underlined), PAM sequence (bolded), and the targeted TBS-II binding site (boxed in blue). (B) Sequencing of the CR #13 T3-22-2-11 mutant which contained a single T base insertion within the TBS-II site in homozygosity. The chromatogram shows the result of sequencing the selected mutant line. (C) Additional CR#13 T2 individuals T2-44-5, T2-44-6, T2-48-1, and T2-52-2 were identified to have indels in the TBS-II binding site. (D) An unexpected deletion was identified in T2-9-3 where the rest of the *CCA1* promoter region was deleted downstream of cut site.

A.



B.

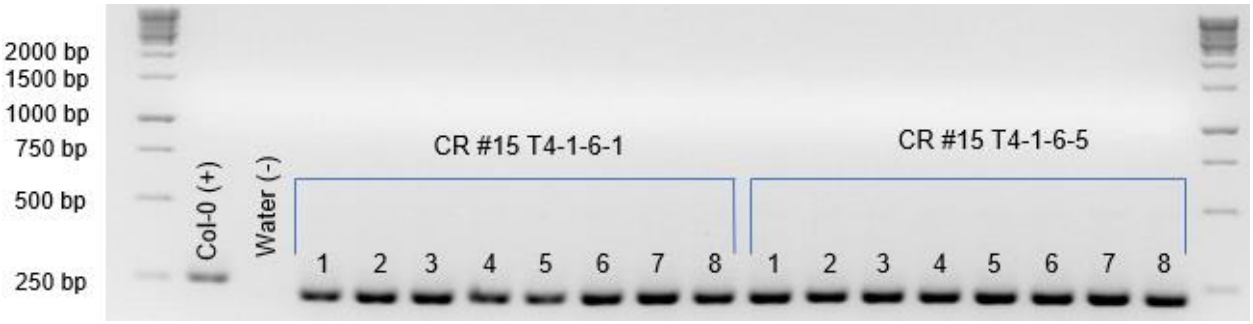


Figure 3: Identification of TBS-II deletion in the *LHY* promoter. (A) Diagram of wild type *LHY* promoter sequence showing the PAM sequences (bolded), TBS-II site (boxed in blue), and the two gRNA binding sites (underlined) used for the CRISPR-Cas9 reaction. Chromatogram from sequencing reaction is shown as well. (B) Agarose gel electrophoresis picture of PCR products of DNA extracted from leaf samples of CRISPR #15 for a 31 bp specific-sized deletion mutation in the *LHY* promoter region. PCR products were separated on a 2% agarose gel and each lane is a separate individual plant. A 1 kb ladder was used for reference. Wild type col-0 *Arabidopsis* ecotype sample was used as a positive control (lane 1) whereas PCR water was used as a negative control (lane 2).

A.

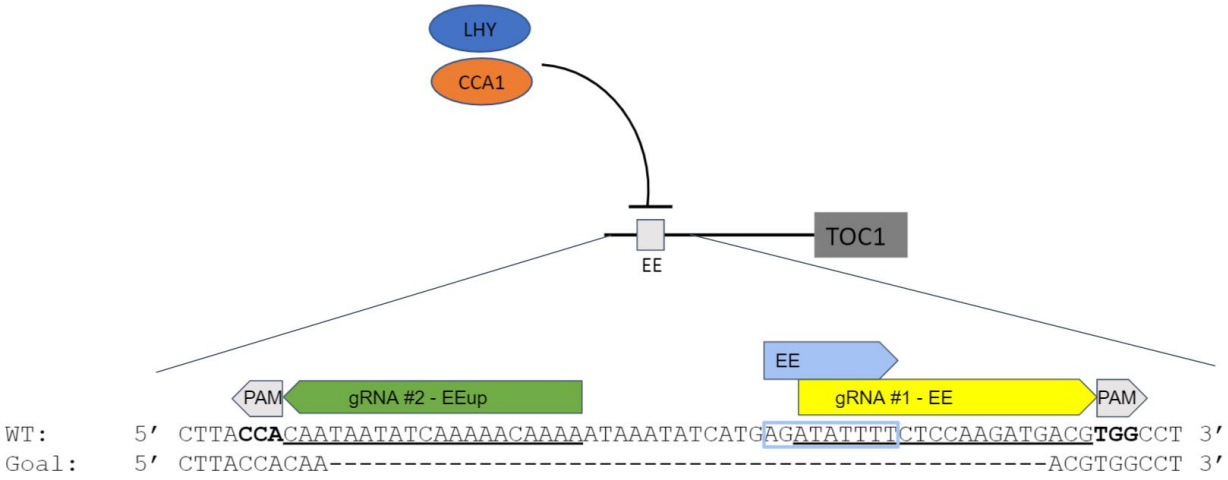


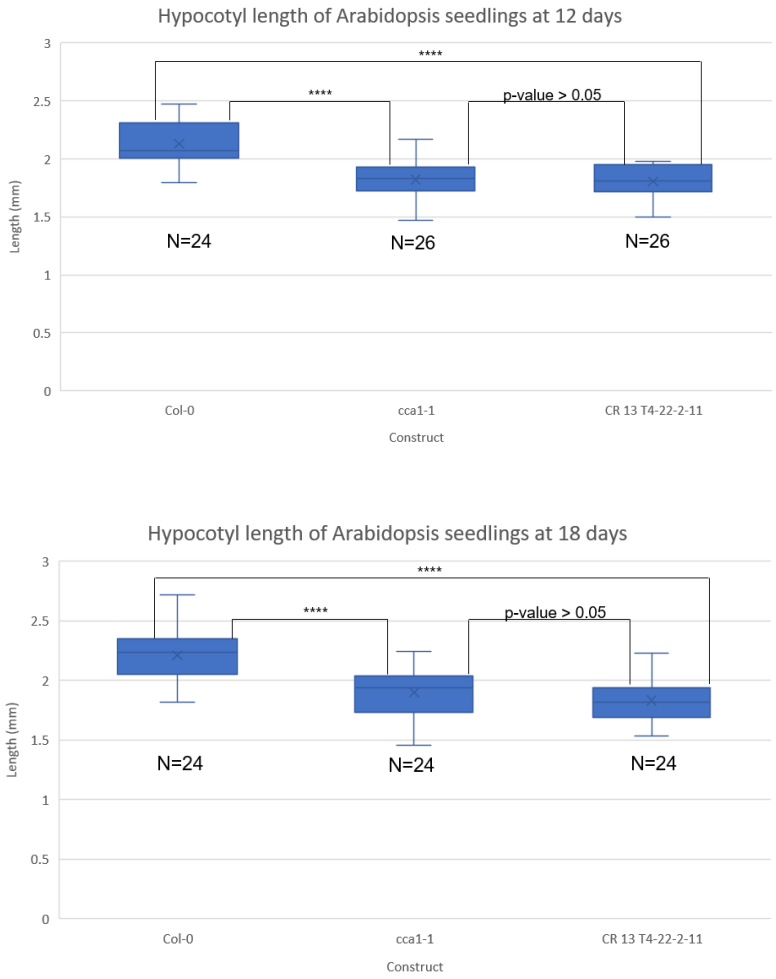
Figure 4: Expected Evening Element deletion in the *TOC1* promoter. (A) Diagram of the wild type *TOC1* promoter sequence showing the guideRNA binding sites (underlined), PAM sequences (bolded), and the targeted Evening Element (EE) site (boxed in blue). Below the wild type of sequence is the expected deletion mutation to occur if Cas9 can successfully cut the target sites.

Hypocotyl analysis in TBS-II of *CCA1* promoter mutants

To confirm if a mutation in the TBS-II site in the promoter of *CCA1* could affect plant growth and clock function, hypocotyl length analysis was done on the progeny of the CRISPR #13 T3-22-2-11 line described previously, wild-type Col-0, and the null *CCA1* mutant *cca1-1*. The hypocotyl length was measured in 12-day old seedlings (Figure 4B). Mean hypocotyl lengths were 1.80 ± 0.14 mm for CRISPR #13 T4-22-2-11 (n=24), 2.12 ± 0.18 mm for Col-0 (n=26), and 1.82 ± 0.18 mm for CCA1-1 (n=26) (Figure 4A). Statistical T-test analysis showed a significant difference between Col-0 and *cca1-1* and Col-0 and no significant difference between *cca1-1* and CRISPR #13 T4-22-2-11 (Figure 4A).

To confirm these results, hypocotyl lengths were measured again in 18-day old seedlings. Similar to the previous experiment, we observed that the mean hypocotyl lengths were 1.83 ± 0.183 for CRISPR #13 T4-22-2-11, 2.21 ± 0.20 for Col-0, and 1.90 ± 0.21 for *cca1-1* (Figure 4A). The statistical analysis of these results confirmed that there was a significant difference between Col-0 with *cca1-1* and Col-0 with CRISPR #13 T4-22-2-11 whereas there was no significant difference between *cca1-1* and CRISPR #13 T4-22-2-11 (Figure 4A).

A



B

Col-0

cca1-1

T4-22-2-11

Figure 5: Hypocotyl length analysis of single base insertion mutation in TBS-II of *CCA1* promoter: (A) Data shown in a boxplot for the hypocotyl lengths of *Arabidopsis* plants at 12d and 18d for wild-type Col-0 (n=24 for 12d and 18d), null mutant *cca1-1* (n=26 for 12d, n=24 for 18d), and insertion mutant CR 13 T4-22-2-11 (n=26 for 12d and n=24 for 18d) where the y-axis shows length in mm while the x-axis represents the construct measured. T-test analysis in both 12d and 18d data sets resulted in a significant p-value below 0.0001 when comparing Col-0 with *cca1-1* and Col-0 with CR 13 T4-22-2-11. T-test analysis in both 12d and 18d data sets of *cca1-1* to CR 13 T3-22-2-11 resulted in a non-significant p-value greater than 0.05. (B) Pictures of individual plants for Col-0, *cca1-1*, and CR 13 T3-22-2-11 at 12 days old.

Phenotype observations in the transformed *Arabidopsis* Col-0 plants

From visual observation in the T4 generation of CRISPR #15 that resulted in the deletion of the TCP-II binding site in the *LHY* promoter, there were three different phenotypes from the progeny of the same T3 individual plant after being moved to soil at the same age of three weeks (Figure 5A and Figure 5B). In two out of seven individual T4 plants, the leaves in the rosette were small and stacked on top of each other. The stage of growth of the two plants was also similar to that of a more mature plant as inflorescence emerged with flower buds already being visible. Another two out of the seven individual T4 plants also had early flowering as the inflorescence has already began growing. The last three out of the seven individual T4 plants were more similar to that of wildtype as they did not have inflorescence emerging and the leaves in the rosette for these four plants were also significantly larger than the other plants.

A **B**

.

Figure 6: Phenotype variation in T4 generation mutants with a TBS-II deletion in the *LHY* promoter (A) Picture of 4 plants in the CR#15 T4-1-6-5 line. Individual plants are numbered from 1-4 and 3 different types of phenotypes can be seen. Plant #1 has an inflorescence and a small rosette with stacked leaves. Plants #2 and #3 have an early flowering phenotype as they already have an inflorescence. Plant #4 does not have an inflorescence and the leaves in the rosette are significantly larger than the other plants. (B) Picture of 3 plants in the CR #15 T4-1-6-5 line with individual plants numbered from 5-7. Plants #5 and #6 do not have an inflorescence and have large leaves in the rosette. Plant #7 has an inflorescence and a small rosette with stacked leaves.

Discussion

Here we have utilized CRISPR-Cas9 technology to induce precise mutations at target regions of the *CCA1*, *LHY*, and *TOC1* clock genes. CRISPR-Cas9 has rapidly been growing as an effective method to incorporate genome engineering in a laboratory setting due to its ease of use and precision while only requiring two major components, the Cas9 nuclease that can create a double stranded break at a DNA site, and a gRNA strand that can lead the Cas9 to a target region. Even though CRISPR-Cas9 is simple to use because of the ability to modify the gRNA to target a gene for the Cas9, there are many limitations that could follow.

Previous studies have explored the importance of the promoter driving expression of the Cas9 nuclease which are transformed into *Arabidopsis* in the efficiency of genome editing (Wang et al. 2015). From our testing, we obtained no successful mutations in the T1 generation but were able to find evidence of mutagenesis in multiple constructs that can be inherited to further generations in the T2 generation when using the egg-cell specific promoter DD45, showing that the timing of the expression of Cas9 potentially plays an important role in efficiency of genome editing. When we initially attempted to identify genome edits in individuals from the T1 generation utilizing the egg-cell specific promoter, we were unable to find any mutations. We hypothesize that this is due to the low expression of Cas9 with the egg-cell specific promoter DD45 in the first transgenic generation because agrobacterium mediated-transformation would normally occur too late after egg cell development in the wild-type plant that is transformed. To overcome this issue, we analyzed T2 pools instead of T1 individuals because the new generation will undergo gametogenesis and

thus allow for high Cas9 expression at the egg cell state. In fact, we found the expected edits for CCA1pro_TBSII (CRISPR #13 T2-22-2, T2-44-5, T2-44-6, T2-48-1, and T2-52-2) and LHYpro_ΔTBSII (CRISPR #15 T2-1-6) in T2 transgenic plants. As such, we established a streamlined procedure to efficiently identify genome edits when the Cas9 protein is expressed at early gametogenesis stages.

Previous research has also observed that the efficiency of CRISPR/Cas9 could be influenced by the GC content as constructs with greater than 50% GC content had a higher editing efficiency than constructs with less than 50% GC (Ren et al. 2016 and Ren et al. 2019). This could be a possible reasoning behind the difficulty of finding evidence of mutagenesis in our CRISPR #17 construct despite analyzing 115 of T2 pools because one of the gRNAs used only had a 15% GC content (Table 3). In contrast, the gRNAs in the CRISPR #13 and CRISPR #15 constructs had 35-45% GC content and we were able to identify successful genome edits (Table 3). As such, the GC content of gRNAs used should be taken into consideration when designing further constructs to target regions of interest. However, this poses an issue as the promoter regions in *Arabidopsis* have a higher AT content compared to GC content (The Arabidopsis Genome Initiative 2000).

We utilized the CRISPR/Cas9 technology to create mutations in the TBS-II elements within the *CCA1* and *LHY* promoters that enable TOC1 regulation. Recent experiments in the Pruneda-Paz lab have identified multiple TCP-II transcription factors which are able to bind to the *CCA1* and *LHY* promoters, but their role in the clock feedback loop remains unknown. To eliminate the possibility of redundancy of these transcription factors, the removal of the TBS-II site where these TCP-II transcription

factors bind will allow for analysis of the clock function without any TCP-II binding interactions in the *CCA1* and *LHY* promoter.

The CRISPR #13 T4-22-2-11 line where a single T base was inserted within the TBS-II site in the *CCA*1 promoter was compared against the wild-type Col-0 and *cca1-1*, which has a null CCA1 mutation, for hypocotyl length analysis. Our results showed that the hypocotyl lengths of *cca1-1* and CRISPR #13 T4-22-2-11 were similar with a p value >0.05, which is unexpected as a single base insertion shows a similar phenotype to that of a null CCA1 mutation. As such, additional CRISPR #13 lines were generated and 4 different indels in the TBS-II site in the *CCA1* promoter were identified. Once screened for homozygosity for the indel mutations, the new CRISPR#13 lines will undergo the same hypocotyl analysis to verify the shorter hypocotyl phenotype.

Even after identification of a mutation at the desired location through CRISPR/Cas9 genome editing, off-target mutations could happen occasionally as there is still a possibility that the gRNA and Cas9 targeting resulted in another gene being edited that can cause additional phenotypic changes than expected (Zhang et al. 2018). This can be observed in the phenotypic analysis in our CRISPR #15 that showed three different phenotypes in the same progeny of an individual plant containing the deletion of the TCP-II binding site of the *LHY* promoter. Genotype analysis on the *LHY* promoter of these plants that had different phenotypes showed that they all had the same deletion mutation in the *LHY* promoter, but it is difficult to confirm off-target mutagenesis without observing the entirety of the genome. To counteract this, backcrossing the *Arabidopsis* plant containing the desired mutation to the wild-type Col-0 *Arabidopsis* in multiple

generations could prove beneficial to remove any extraneous mutations and create a generation that contains only the desired mutation before further analysis.

The next step for this project will be to compare circadian clock function and plant growth/development phenotypes of the CRISPR/Cas9 mutants to wild type and mutants of previous research. To do this, backcrossing of the identified mutants will be done to remove off-target mutations and then the mutants can be crossed with *Arabidopsis* plants where the core clock genes are fused with a luciferase reporter. Fusing clock genes with a luciferase reporter can allow for the measurement of gene expression and the circadian clock function. These comparisons will give insight into the degree that these cis elements impact the circadian clock function. With the development of a streamlined procedure for identification of genome editing with CRISPR/Cas9, we can induce mutagenesis to other cis-elements in the promoter region or in the coding regions of the core clock genes to further understand the clock regulatory mechanism.

Materials and Methods

Vector Construction

The plasmid pGreenII_0179 was used as the backbone for the construction of binary vector P3J1 where plant codon optimized Cas9 and the nuclear localization signals are driven by the egg cell and early embryo *DD45* promoter. Hygromycin resistance is driven by the *CaMV 35S* promoter. In this binary vector, single gRNAs are driven by the *U6-26* promoter whereas for double gRNA constructs, the second gRNA is driven by the *U6-29* promoter. To generate constructs for CRISPR/Cas9, single or double gRNAs (Table 1) are PCR amplified and cloned into the *BamH*I digested P3J1 binary vector through DNA assembly using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs). Each construct is then transformed into competent *E.coli* Top10 cells to screen for the plasmid containing the desired gRNA construct by colony PCR.

Agrobacterium-mediated *Arabidopsis* transformation

Transgenic *Arabidopsis thaliana* (ecotype col-0) plants were created through Agrobacterium-mediated *Arabidopsis* transformation using the common floral dip method (Zhang et al. 2006). The CRISPR-Cas9 constructs previously described were transformed into AGL0 Agrobacterium through electroporation and then incubated in 1 mL of LB broth for 1-1.5 hours in a 30°C shaker. The Agrobacterium is then placed on LB, Kanamycin (50 mg/L), and Rifampicin (25 mg/L) plates to be incubated at 30°C for 2 days. Agrobacterium colonies are selected and inoculated in 400 mL of LB broth with the appropriate antibiotics until reaching a stationary phase in growth. The cells are then collected through centrifugation and are resuspended using 400 ml of 5% sucrose before adding 0.02% Silwet. Wild-type *Arabidopsis* Col-0 plants with inflorescence are inverted and dipped into the Agrobacterium suspension solution for 30 seconds before letting it drain. After dipping, the *Arabidopsis* plants are then placed in plastic bags and laid on their sides for one day. After one day, the plants are placed back upright, and the plastic bags are removed to allow the seeds to grow normally.

T1 seeds from transformed *Arabidopsis thaliana* plants were surface sterilized for 4 hours with chlorine gas and plated on Murashige and Skoog (MS) medium with 1% sucrose and hygromycin (20 mg/L) antibiotic. Plates were stratified for 3 days at 4°C and moved into an incubator where plants were grown at 22°C with a 12-hour light dark cycle. After 14 days, hygromycin resistant seedlings were transplanted to soil. T2 seeds were sterilized and plated on the same medium as the T1 seeds. After verifying mutagenesis, T3 seeds were selected for hygromycin sensitivity to ensure the plants are Cas9 free to prevent off-target mutations.

Genomic DNA extraction and PCR analysis

Arabidopsis leaf tissue was collected in pools of 8 in the T2 generation. Genomic DNA of *Arabidopsis* was extracted through a chloroform DNA extraction protocol (adapted from Edwards et al, 1991). To identify *Arabidopsis* plants with mutagenesis, promoter regions of the clock gene for the respective CRISPR lines were amplified with specific primers that flank the target transcription factor binding site (Table 2). Amplification was done through Touchdown PCR using Choice Taq polymerase with the following protocol: 95°C for 3 min, (95°C for 30 s, 70-50°C for 30 s, 72°C for 1min/kb) for 40 cycles with touchdown -0.5°C in each cycle, (95°C for 30 s, 52°C for 30s, 72°C for 1 min/kb) for 30 cycles, 72°C for 5 min, and 4°C to hold. The PCR product was run on a 2% agarose gel to verify presence of DNA and identify fragment deletions. After gel analysis, DNA from PCR product is precipitated with an ethanol precipitation method and sent for Sanger sequencing. T2 pools with mutations are deconvoluted and DNA extraction is repeated to identify individual plants that contain the mutation.

Hypocotyl Length Measurement

To measure the length of the hypocotyls of *Arabidopsis* plants, all seeds were grown in the same conditions. Seeds were surface sterilized with chlorine gas and then were planted on an MS medium with 1% sucrose and stratified for 3 days at 4°C. After stratification, plates were placed into an incubator to be grown at 22°C with a 12-hour light dark cycle. At the age of 12 days, seedlings are moved onto a plastic screen protector with the hypocotyl exposed and then scanned to download an image to a computer. Images were analyzed using ImageJ.

Table 1: List of CRISPR-Cas9 constructs targeting promoter cis-elements

Construct	Gene region	Target region	gRNA #1 sequence (5'→3')	gRNA #2 sequence (5'→3')
CCA1pro_TBSII (CRISPR #13)	CCA1 promoter	TBS-II	GGTCCACTGATGTTTCTAGT	N/A
LHYpro_ΔTBS-II (CRISPR #15)	LHY promoter	TBS-II	CCACACTCACTTTTACCTAC	ATAATCGTTTACACGTAAGC
TOC1pro_ΔEE (CRISPR #17)	TOC1 promoter	Evening Element	ATATTTTCTCCAAGATGACG	CAATAATATCAAAAACAAAA

Table 2: List of CRISPR-Cas9 Construct PCR Primers

Construct	Forward Primer (5'→3')	Reverse Primer (5'→3')
CCA1pro_TBSII (CRISPR #13)	CCACGTCCTTCCTTCAATC	CCAGATTTCGTTGGATAAG
LHYpro_ΔTBS-II (CRISPR #15)	GAATAGGCGTAAAAGTGAGG	GTAGTGGCTGTCTACTGC
TOC1pro_ΔEE (CRISPR #17)	CAAGGAGATGACGTGGAC	GGTTTCTGATGGTTTGGTC

Table 3: CRISPR/Cas9 T2 pool genome editing efficiency

Construct	Promoter driving Cas9	T2 Pool PCR Count	T2 pool Percent efficiency
CCA1pro_TBSII (CRISPR #13)	DD45	28	21.4% (6/28)
LHYpro_ΔTBS-II (CRISPR #15)	DD45	10	100% (10/10)
TOC1pro_ΔEE (CRISPR #17)	DD45	115	0% (0/115)

References

Alabadi, D., Oyama, T., Yanovsky, M. J., Harmon, F. G., Mas, P, & Kay, S. A. (2001). Reciprocal Regulation between TOC1 and LHY/CCA1 Within the Arabidopsis Circadian Clock. *Science, 293 (5531),* 880. https://doi.org/10.1126/science.1061320

Albrecht U., Ripperger J.A. (2009) Clock Genes. In: Binder M.D., Hirokawa N., Windhorst U. (eds) Encyclopedia of Neuroscience. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_1080

Ambawat, S., Sharma, P., Yadav, N. R., & Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and molecular biology of plants: an international journal of functional plant biology*, *19*(3), 307–321. https://doi.org/10.1007/s12298-013-0179-1

Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., & Zoran, M. J. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics*, *6*(7), 544–556. https://doi.org/10.1038/nrg1633

Covington, M. F., Maloof, J. N., Straume, M., Kay, S. A., & Harmer, S. L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome biology*, *9*(8), R130. https://doi.org/10.1186/gb-2008-9-8-r130

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic acids research*, *19*(6), 1349. https://doi.org/10.1093/nar/19.6.1349

Edwards, K. D., & Millar, A. J. (2007). Analysis of circadian leaf movement rhythms in Arabidopsis thaliana. *Methods in molecular biology (Clifton, N.J.)*, *362*, 103–113. https://doi.org/10.1007/978-1-59745-257-1_7

Gelvin S. B. (2003). Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiology and molecular biology reviews: MMBR*, *67*(1), 16–37. https://doi.org/10.1128/MMBR.67.1.16-37.2003

Green, R. M., & Tobin, E. M. (1999). Loss of the circadian clock-associated protein 1 in Arabidopsis results in altered clock-regulated gene expression. *Proceedings of*

the National Academy of Sciences of the United States of America, *96*(7), 4176–4179. https://doi.org/10.1073/pnas.96.7.4176

Makino, S., Matsushika, A., Kojima, M., Yamashino, T., & Mizuno, T. (2002). The APRR1/TOC1 quintet implicated in circadian rhythms of Arabidopsis thaliana: I. Characterization with APRR1-overexpressing plants. *Plant & cell physiology*, *43*(1), 58–69. https://doi.org/10.1093/pcp/pcf005

Michael, T. P., Salomé, P. A., Yu, H. J., Spencer, T. R., Sharp, E. L., McPeek, M. A., Alonso, J. M., Ecker, J. R., & McClung, C. R. (2003). Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science (New York, N.Y.)*, *302*(5647), 1049–1053. https://doi.org/10.1126/science.1082971

Nohales, M. A., & Kay, S. A. (2016). Molecular mechanisms at the core of the plant circadian oscillator. *Nature structural & molecular biology*, *23*(12), 1061–1069. https://doi.org/10.1038/nsmb.3327

Ono, A., Sato, A., Fujimoto, K. J., Matsuo, H., Yanai, T., Kinoshita, T., & Nakamichi, N. (2019). 3,4-Dibromo-7-Azaindole Modulates Arabidopsis Circadian Clock by Inhibiting Casein Kinase 1 Activity. *Plant & cell physiology*, *60*(11), 2360–2368. https://doi.org/10.1093/pcp/pcz183

Pruneda-Paz, J. L., & Kay, S. A. (2010). An expanding universe of circadian networks in higher plants. Trends in Plant Science, 15(5), 259–265. doi:10.1016/j.tplants.2010.03.003

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Ren, C., Liu, X., Zhang, Z., Wang, Y., Duan, W., Li, S., & Liang, Z. (2016). CRISPR/Cas9-mediated efficient targeted mutagenesis in Chardonnay (Vitis vinifera L.). *Scientific reports*, *6*, 32289. https://doi.org/10.1038/srep32289

Ren, F., Ren, C., Zhang, Z., Duan, W., Lecourieux, D., Li, S., & Liang, Z. (2019). Efficiency Optimization of CRISPR/Cas9-Mediated Targeted Mutagenesis in Grape. *Frontiers in plant science*, *10*, 612. https://doi.org/10.3389/fpls.2019.00612

The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. https://doi.org/10.1038/35048692

Van Doorn, W., & Van Meeteren, U. (2003). Flower opening and closure: A review. *Journal of Experimental Botany, 54*(389), 1801-1812. https://doi.org/10.1093/jxb/erg213

Wang, Z. P., Xing, H. L., Dong, L., Zhang, H. Y., Han, C. Y., Wang, X. C., & Chen, Q. J. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome biology*, *16*(1), 144. https://doi.org/10.1186/s13059-015-0715-0

Webb. A. A., (2003). The Physiology of Circadian Rhythms in Plants. *The New Phytologist, 160*(2), 281-303. Retrieved March 18, 2021, from http://www.jstor.org/stable/1514280

Žádníková, P., Smet, D., Zhu, Q., Van Der Straeten, D., & Benková, E. (2015). Strategies of seedlings to overcome their sessile nature: auxin in mobility control. *Frontiers in plant science*, *6*, 218. https://doi.org/10.3389/fpls.2015.00218

Zhang, Q., Xing, H. L., Wang, Z. P., Zhang, H. Y., Yang, F., Wang, X. C., & Chen, Q. J. (2018). Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention. *Plant molecular biology*, *96*(4-5), 445–456. https://doi.org/10.1007/s11103-018-0709-x

Zhang, Xiuren, Rossana Henriques, Shih-Shun Lin, Qi-Wen Niu, and Nam-Hai Chua. 2006. “Agrobacterium-Mediated Transformation of Arabidopsis Thaliana Using the Floral Dip Method.” Nature Protocols 1 (2): 641–46.