Title
SICKLE is a putative splicing-associated protein required for normal intron lariat debranching

Author
Kovak, Emma E

Publication Date
2019

Peer reviewed|Thesis/dissertation

SICKLE is a putative splicing-associated protein required for normal intron lariat debranching

By

Emma E. Kovak

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Frank Harmon, Chair
Professor Sarah Hake
Professor Benjamin Blackman
Professor Iswar Hariharan

Spring 2019

Abstract

SICKLE is a putative splicing-associated protein required for normal intron lariat debranching

by

Emma E. Kovak

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Frank Harmon, Chair

The *Arabidopsis thaliana* SICKLE (SIC) protein is proline-rich but largely uncharacterized. *sic-1* mutants have a miRNA deficiency, alterations in transcript splicing, and increased intron accumulation. *sic-3* mutants have global changes in transcript splicing and circadian clock defects likely caused by altered splicing of circadian clock transcripts. Here we show that 13 of 20 putative SIC-interacting proteins are splicing-associated, and that Arabidopsis SIC and DBR1 interact directly. DBR1 is the only lariat debranching enzyme in Arabidopsis and knockout mutants are embryo lethal. The weak *dbr1-2* mutant accumulates intron lariats, which interfere with miRNA processing, and has morphological phenotypes similar to *sic-1*. Human cells deficient in DBR1 also have lariat accumulation and this causes exon skipping. Here we show that *sic-1*, *sic-3*, and a new *dbr1* mutant allele, *dbr1-3*, accumulate lariats like the *dbr1-2* mutant, and that *sic-3* and *dbr1-2* have similar changes in splicing, including increased intron retentions. Therefore, lariat accumulation in the *sic-1* allele may be responsible for its miRNA deficiencies and related morphological phenotypes, as well as the alterations in transcript splicing. We present a unified theory for *sic* mutant phenotypes based on SIC-DBR1 interaction and *sic* intron lariat accumulation.

Chapter 1: SICKLE is a putative splicing-associated protein required for normal intron lariat debranching

ABSTRACT

The *Arabidopsis thaliana* SICKLE (SIC) protein is proline-rich but largely uncharacterized. *sic-1* mutants have a miRNA deficiency, alterations in transcript splicing, and increased intron accumulation. *sic-3* mutants have global changes in transcript splicing and circadian clock defects likely caused by altered splicing of circadian clock transcripts. Here we show that 13 of 20 putative SIC-interacting proteins are splicing-associated, and that Arabidopsis SIC and DBR1 interact directly. DBR1 is the only lariat debranching enzyme in Arabidopsis and knockout mutants are embryo lethal. The weak *dbr1-2* mutant accumulates intron lariats, which interfere with miRNA processing, and has morphological phenotypes similar to *sic-1*. Human cells deficient in DBR1 also have lariat accumulation and this causes exon skipping. Here we show that *sic-1*, *sic-3*, and a new *dbr1* mutant allele, *dbr1-3*, accumulate lariats like the *dbr1-2* mutant, and that *sic-3* and *dbr1-2* have similar changes in splicing, including increased intron retentions. Therefore, lariat accumulation in the *sic-1* allele may be responsible for its miRNA deficiencies and related morphological phenotypes, as well as the alterations in transcript splicing. We present a unified theory for *sic* mutant phenotypes based on SIC-DBR1 interaction and *sic* intron lariat accumulation.

INTRODUCTION

Post-transcriptional regulation is important in many plant processes that affect plant growth and stress responses, including the circadian clock and development. Normally, transcripts with introns are spliced in one way, always using the same splice sites, to form one mature mRNA transcript – this is called constitutive splicing. Alternative splicing is a form of post-transcriptional regulation that makes multiple transcript isoforms from a single gene.

Splicing is a complex process involving removal of introns from a transcript, catalyzed by the spliceosome. Many eukaryotic genes have multiple introns, so there can be multiple alternatively spliced transcripts from different combinations of removed and retained introns. The four main types of alternative splicing events are intron retention, exon skipping, and alternative 5' or 3' splice site usage.

The spliceosome removes introns from eukaryotic mRNA in three main steps: splice site recognition, then the two catalytic steps releasing the 5' then 3' ends of the intron. The first catalytic step cleaves the 5' end of the intron and ligates it to the intron branch point (A nucleotide), forming a lariat; the second step cleaves the 3' end of the intron, releasing a free intron lariat. After the intron is fully removed from the mRNA transcript, the post-catalytic spliceosome (also known as the post-splicing or intron-lariat spliceosome) releases the spliced mRNA and the spliceosome disassembles, releasing the free intron lariat. The spliceosome assembles de-novo for splicing of each intron, even within the same transcript, and spliceosome proteins are recycled after spliceosome disassembly to be used in formation of a new spliceosome. The sequences of splice sites are relatively conserved, but there are small variations within organisms that affect spliceosome ability to splice individual introns; therefore the spliceosome has higher affinity for some introns than others (as reviewed in (Wahl et al., 2009; Reddy et al., 2013; Matera and Wang, 2014)).

The spliceosome remodels between each step of splicing, and different spliceosome components are responsible for each step. The spliceosome is composed of small nuclear ribonucleoproteins (snRNPs) and other splicing proteins, including the Nineteen Complex (NTC, aka Prp19 complex). First, the U1 snRNP recognizes the 5' splice site, while the U2 snRNP recognizes the 3' splice site and the branch point, together forming the intron-defining complex (Complex A). The U4/U6/U5 tri-snRNP joins the intron-defining complex to form the pre-catalytic complex (Complex B). The U1 and U4 snRNPs are released, with the U6 snRNP replacing U1 binding at the 5' splice site; then the NTC joins, forming the activated pre-catalytic complex. The activated spliceosome performs the two catalytic steps: 5' splice site cleavage and lariat formation, then 3' splice site cleavage and exon ligation (Complex B* and Complex C). After the two catalytic steps the spliceosome contains the spliced mRNA with joined exons and the excised intron lariat; then spliced mRNA is released, and the remaining intron lariat spliceosome contains U2, U5, and U6 snRNPs and the NTC (as reviewed in (Wahl et al., 2009; Reddy et al., 2013; Matera and Wang, 2014)).

Intron lariat accumulation is common in mutants that affect spliceosome disassembly (Hogg et al., 2010). Intron lariats do not accumulate if splicing is blocked at or before the first step (Brenner and Guthrie, 2005) because formation of

the lariat structure is the first step of the splicing reaction. Splicing is a reversible process with competition between the forward and reverse reactions at several steps that serve as quality control checks (Hoskins and Moore, 2012). Therefore, splicing of an intron can be reversed any time before mRNA release and spliceosome disassembly.

The NineTeen Related (NTR) complex is responsible for spliceosome disassembly. Three proteins, Cwc23, Ntr1, and Prp43 are core members of the NTR complex, and both *CWC23* (Sahi et al., 2010) and *PRP43* (Arenas and Abelson, 1997) mutants accumulate lariats. A yeast *PRP43* mutant accumulates un-spliced pre-mRNAs and excised intron lariats in spliceosomes (Arenas and Abelson, 1997). A partial-loss-of-function *CWC23* mutant has global splicing defects and accumulates un-spliced pre-mRNA and excised intron lariats (Sahi et al., 2010). It is possible that mutants in other splicing genes besides DBR1 and members of the NTR complex also accumulate lariats, but examples of this have not been published.

After the spliced intron lariat is released from the spliceosome, it must be debranched before it can be degraded. RNA debranching enzyme 1 (DBR1) is the only known intron lariat debranching enzyme in Arabidopsis, and is the rate-limiting step in intron degradation. The first published viable plant mutant, *dbr1-2*, exhibits intron lariat accumulation and miRNA deficiency, with additional morphological phenotypes including curly and serrated leaf margins, increased branching, reduced height, and reduced fertility (Li et al., 2016). Work on this *dbr1-2* mutant showed the cause of the miRNA deficiency is intron lariat accumulation. Lariats compete with pri-miRNAs for binding to the HYL1/DCL complex, which reduces miRNA processing (Li et al., 2016).

Lariat accumulation in *dbr1* mutants may also change alternative splicing. Work in human cells shows that insufficient DBR1 activity leads to changes in alternative splicing of transcripts, primarily increased exon skipping (Han et al., 2017). In DBR1-deficient human cells, un-degraded intron lariats are stuck in stalled post-splicing complexes that contain snRNPs. Sequestration of snRNPs blocks their reuse in the next round of spliceosome complex formation, effectively causing deficiency in these snRNPs. With fewer spliceosome components available, less active spliceosomes associate with introns that have weak splice sites, which causes exon skipping (Han et al., 2017). This is supported by other work indicating that decreased recycling of spliceosome components could affect splicing efficiency (Arenas and Abelson, 1997; Bitton et al., 2014).

Apart from their impact on miRNA processing and transcript splicing, little is known about processing of lariats and any alternate functions of circular RNAs. Most lariats are degraded within seconds to minutes of splicing, but some lariats persist for much longer in WT plants. For example, the Arabidopsis *dbr1-2* mutant had 1534 detectable lariats (different lariats by sequence), and 23% of these, a total of 360, were also detectable in WT plants (Li et al., 2016). In vertebrates, DBR1 appears to have no mechanism for selecting certain lariats as substrates, but lariat sequence features such as a less common cytosine branchpoint (adenosine is most common) are more difficult for DBR1 to debranch (Talhouarne and Gall, 2018). Some intron lariats are actively transported from the nucleus to the cytoplasm

(Talhouarne and Gall, 2018), and at least one has been shown to globally regulate gene expression (Cheng et al., 2018).

The debranching activity of DBR1-type enzymes is well studied, but the larger context of its activity is unclear in any organism. For example, it remains unclear whether DBR1 is a component of any spliceosome complex. Research in human HeLA cell extracts examined the post-splicing Intron Large (IL) and Intron Small (IS) complexes, and showed that DBR1 was able to debranch intron lariats only after the IL complex was converted to the IS complex; however, DBR1 does not appear in immunoprecipitation of either the IL or IS complex (Yoshimoto et al., 2009). The IL complex includes U2, U5, and U6 snRNPs and some NTC proteins, while the IS complex is made up of fewer proteins, and does not include known snRNPs or NTC components. Human DBR1 has been shown to interact with XAB2, which is a component of both the NTC and the IL complexes (Masaki et al., 2015).

SICKLE (SIC) is a small 35kd proline-rich protein that is conserved across the angiosperm lineage. The protein has predicted domains for protein binding, RNA binding, and DNA binding, as well as a nuclear localization sequence and an MPLKIP motif (Marshall et al., 2016). The subcellular distribution of SIC is in the nucleus within sub-nuclear foci (Zhan et al., 2012; Marshall et al., 2016), which is an accumulation pattern common to splicing-related proteins (Lamond and Spector, 2003). A previous study performed IP-MS and pulled SIC down with 20 other proteins, including SIC itself (Karampelias et al., 2016).

Previous work has characterized *sic* mutant phenotypes including globally increased intron retention, reduced levels of several miRNAs, altered splicing of circadian clock transcripts, a cool temperature-sensitive circadian clock, and impaired auxin-signaling (Zhan et al., 2012; Karampelias et al., 2016; Marshall et al., 2016). Here we investigate proteins that interact with SIC based on previous work that identified 20 putative SIC interactors. Of these, 13 proteins are splicing-associated and we demonstrate that SIC directly interacts with one of these, Arabidopsis DBR1. We also show that *sic* mutant alleles *sic-1* and *sic-3* accumulate lariats like the weak *dbr1-2* mutant allele and a novel *dbr1* mutant allele, *dbr1-3*. Furthermore, the changes in transcript splicing observed in *sic-3* share similarities with alternative splicing observed in *dbr1-2*. Given the shared phenotypes of *sic* and weak *dbr1* mutants, we present a unified theory for *sic* mutant phenotypes based on SIC-DBR1 interaction and *sic* intron lariat accumulation.

RESULTS

SIC interacts with many splicing proteins, including homologs of NTC and IL components

In order to learn what processes SIC is involved in, we used a prior immunoprecipitation-mass spectrometry study that described proteins associated with SIC (Karampelias et al., 2016). We predicted functions of SIC-associated proteins from this experiment based on literature curation. Of the 20 SIC-associated proteins in this experiment, 13 were known spliceosome- or NTC-associated proteins or homologs of proteins participating in these complexes in other species (Table 1).

The human PP2A complex de-phosphorylates U2 and U5 snRNPs that are part of the spliceosome in order to facilitate structural changes necessary for the transition from the first to the second step of splicing. Depletion of both PP2A and PP1 phosphatases in HeLa cells blocks the second step of splicing (Shi et al., 2006).

The PRP19/NTC assembles into the U5 snRNP to form an activated spliceosome, and is essential for steps 1 and 2 of intron removal, providing support for snRNPs and spliceosome conformational changes. The NTC is conserved from yeast to humans, and little studied in plants. The NTR (NTC-related) complex contains some proteins from the NTC, and catalyzes spliceosome disassembly, though the NTC itself may be required as well. Availability of snRNPs is affected by spliceosome disassembly and recycling. AtXAB2 (At5g28740) is a core component of the Arabidopsis NTC complex (Ganpudi and Schroeder, 2011).

Finally, DBR1 and PRMT4A and PRMT4B are not part of either the PP2A or NTC complex, but are splicing-associated proteins. Homologs of Arabidopsis PRMT4A and PRMT4B methylate parts of the spliceosome in mammals (Cheng et al., 2007; Kuhn et al., 2011). DBR1 is a single copy gene in Arabidopsis and all eukaryotes (Ooi et al., 2001).

SIC interacts directly with DBR1

From the previously published list of proteins pulled down with SIC using IP-MS (Karampelias et al., 2016), we chose to further characterize SIC interaction with DBR1. In order to confirm interaction between SIC and DBR1, we performed co-immunoprecipitation (CoIP), yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFC).

CoIP experiments employed transient co-expression of epitope-tagged SIC and DBR in leaves of *Nicotiana benthamiana*. After co-infiltration of DBR1-CFP and SIC-FLAG, western blotting with anti-GFP and anti-FLAG showed high levels of DBR1-CFP and very little SIC-FLAG in the input lane; FLAG pulldown and subsequent Western blotting showed enriched SIC-FLAG and about 1/3 of the DBR1-GFP present in the input associated with this SIC (Figure 1A). The same co-infiltration of DBR1-CFP and SIC-FLAG after GFP pulldown showed slight enrichment of DBR1-CFP compared to the input, and a little SIC-FLAG associated with this DBR1 (Figure 1A). Negative controls with single infiltrated constructs showed no signal for protein associated with the protein that was pulled down. This indicates that SIC and DBR1 interact directly when overexpressed in *Nicotiana benthamiana*.

For Y2H, in addition to full-length SIC we also included $SIC_{\Delta 29\text{-}104}$, which is the predicted protein product of one of three transcript spliceforms in *sic-3*. In Y2H the full-length SIC protein had no detectable interaction with DBR1 (Figure 1B). However, SIC with an internal deletion, $SIC_{\Delta 29\text{-}104}$, had a strong interaction with DBR1 (Figure 1B). We suspect that $SIC_{\Delta 29\text{-}104}$ mimics a SIC protein modification that doesn't happen in the Y2H heterologous system, but does happen in *Nicotiana benthamiana* and Arabidopsis.

In addition to Y2H interaction data, we also examined localization of these SIC and DBR1. Not all proteins that have a speckled nuclear distribution are nuclear

speckles (interchromatin granule clusters), therefore it is crucial to co-localize with a known splicing component (Lamond and Spector, 2003). Here we demonstrated that SIC and DBR1 co-localize in the nucleus, sometimes in punctate nuclear speckles (Figure 2).

Five intron lariats accumulate in both *sic* and *dbr1-3*

To examine common phenotypes between *sic* and *dbr1* mutants, we identified a novel weak *dbr1* mutant allele designated *dbr1-3*. The *dbr1-3* mutant is from the SALK T-DNA mutant collection (Alonso et al., 2003) and has a T-DNA insertion in the 5'UTR of the *DBR1* gene (Supplemental Figure 1A). Homozygous *dbr1-3* plants exhibit a substantial reduction in *DBR1* expression (Figure 3B) and high levels of intron lariats (Figure 3A) similar to published results for *dbr1-2* (Li et al., 2016). The *dbr1-3* and *sic-1* mutants have compact rosettes (Supplemental Figure 1B), similar to that described for *dbr1-2*.

Because SIC physically interacts with DBR1, we investigated whether intron lariats accumulate in *sic* mutants as they do in the *dbr1-3* mutant. We assayed *sic-1*, *sic-3*, and our *dbr1-3* mutant allele for the presence of five intron lariats previously shown to accumulate to high levels in *dbr1-2* but not WT Col-0 (Li et al., 2016). Intron lariats were detected by PCR using lariat-specific primers as described previously (Ohi et al., 2007). Intron lariats were largely undetectable in WT for intron 1 of *LHY* (*LHYI1*), intron 4 of *CCA1* (*CCA1I4*), intron 1 of *RVE4* (*RVE4I1*) and intron 1 of *SIT4* (*SIT4I1*), and detectable at low levels for intron 2 of *CAX1* (*CAX1I2*) (Figure 3A). As expected, the weak *dbr1-3* mutant allele exhibited intron lariats derived from each of these introns, except *CCA1I4* (Figure 3A). The *sic-3* and *sic-1* mutants also accumulated these intron lariats. Interestingly, intron lariats were more apparent in the stronger *sic-1* mutant than in *dbr1-3* (Figure 3A). Thus, these *sic* alleles have stable intron lariats, similar to a weak *dbr1* mutant allele. We were not able to evaluate lariat accumulation in a double *sic dbr1* mutant because it appears to be lethal. However, this suggests that a double mutant has more lariat accumulation than either single mutant, because it is lethal like the strong Arabidopsis *dbr1* mutant (Wang et al., 2004).

Increased intron lariat accumulation in the *sic* mutants was not due to a change in *DBR1* expression, as the levels of *DBR1* transcript in *sic-1* and *sic-3* were close to those in WT (Figure 3B). To confirm that higher lariat accumulation in *sic* mutants was not due to increased levels of the source transcript, we measured total *LHY* transcript in each mutant. *LHY* expression was lower in sic mutants and equal to WT in *dbr1-3* (Figure 3C). Therefore lariat accumulation in *sic* mutants is likely due to decreased lariat debranching as in *dbr1*.

Both *sic-1* and *dbr1-3* have impaired lariat degradation

The experiments above measured static levels of intron lariats at a single time point in plants under diurnal conditions. Next we wanted to see whether *sic* and *dbr1* are similarly impaired in intron lariat degradation, rather than *sic* perhaps producing more lariats than *dbr1*. Therefore, we determined the kinetics of lariat degradation in each genetic background during a time course. To do this with a controlled timing and level of transcription, we examined levels of intron lariats

arising from light-induced transcripts produced from light-regulated genes. Intron lariat accumulation was determined in dark-adapted plants before, during, and after a light pulse (Figure 4A).

Expression of *CCA1*, *LHY*, and *RVE4* was very low before the light pulse (Figure 4C), so we assumed that most or all of the intron lariats present before the light pulse remained un-degraded throughout the 2-day dark period. Gene expression was reliably induced by the light pulse (Figure 4C) and intron lariat levels increased in *sic* and *dbr1* mutants, while few lariats were ever detectable in WT (Figure 4B). Levels of lariat accumulation varied between the three lariats, indicating either differential accumulation or differential degradation (Figure 4B). *CCA1I4L* and *RVE4I1L* were detectable in *sic-1* at all time points, indicating they persisted during the 2-day dark period, and for 9 hours after the light pulse; *LHYI1L* were degraded by the 10 hour time point (Figure 4B). In *dbr1-3*, lariats were intermittently detectable, with the highest accumulation or *LHYI1L* during the light pulse; this indicates that *dbr1-3* has a greater ability to debranch lariats than *sic-1*. Thus, intron lariats are stable in *sic-1*, indicating inhibited intron lariat degradation. The pattern of *RVE4I1L* accumulation in the weaker *sic-3* mutant allele was similar to *sic-1*, but the *LHYI1L* and *CCA1I4L* lariats were not detectable (Figure 4B). Expression of *LHY* and *CCA1* was similar in all three mutant genotypes, consistent with the differences in intron lariat accumulation arising from inability to remove the intron lariat, not elevated gene expression in the mutants (Figure 4C).

sic-3 and *dbr1-2* have global changes in splicing, especially intron retentions

Previous work showed higher levels of alternative splice variants of some circadian clock transcripts in *sic-3* compared to WT, and a further increase of these variants at cold temperatures (Marshall et al., 2016). At 16°C when the *sic-3* phenotypes are most apparent, the *sic-3* mutant had massive changes in differential alternative splicing compared to WT – almost 20% of the transcriptome (Marshall, 2017). Other work showed a global increase in intron sequences in *sic-1* compared to WT (Zhan et al., 2012). I chose to further analyze this previously published RNA-Seq data of *sic-3* and WT at 16°C (Marshall, 2017) and previously published RNA-Seq data comparing *dbr1-2* to WT (Li et al., 2016) by comparing types of alternative splicing events within and between *sic-3* and *dbr1-2*.

Each type of alternative splicing event with DU (compared to WT in each experiment) in *sic-3* and *dbr1-2* made up a very similar percentage of the total number of alternative splicing events with DU; *sic-3* just had a little more exon skipping and *dbr1-2* had a little more intron retention (Figure 5A and Table 3). Intron retention (IR) made up the highest percentage of total alternative splicing affected in both mutants, 38% in *sic-3* and 48% in *dbr1-2* (Figure 5A and Table 3). We also calculated what percentage of each type of AS event was upregulated vs. downregulated. Remarkably high percentages of each type of AS event were upregulated in *dbr1-2* (82% for Alt3ss, over 90% for IR, ES, and Alt3SS) (Figure 5A and Table 3). The percentages of ES and Alt3ss events that were upregulated in both mutants were almost identical, and *sic-3* had less upregulation of IR and Alt5ss (Figure 5A and Table 3). In both *sic-3* and *dbr1-2*, intron retentions were the type of

alternative splicing most affected, and ~70-90% of affected intron retentions were upregulated.

Intron retentions affected in *sic-3* have non-consensus splice sites

Since both *sic-3* and *dbr1-2* had increased intron retention compared to WT, and previous research showed increased exon skipping at weak splice sites in *dbr1*-deficient human cells, we wanted to know if these *sic-3* intron retentions also have weak splice sites. We compared the previously published consensus 5' splice site sequence for all Arabidopsis introns (GT_AG_U2 introns) (Hernando et al., 2015) (Figure 5B) to the 5' splice sites of intron retentions that were affected in the *sic-3* mutant (Figure 5C).

The 5' splice sites of *sic-3* intron retentions were different than the consensus at several positions before and after the conserved GT. The consensus A nucleotide at position 7 was most underrepresented in *sic-3* retained introns, and positions 3, 5, and 6 were also affected (Figure 5C). For intron retention events affected in either *prmt5* or both *prmt5* and *prmt4a;4b*, the 5' splice site sequence was significantly different than the consensus sequence for all Arabidopsis introns; the consensus A and G at positions -1 and -2 were underrepresented (Hernando et al., 2015). The *sic-3* retained introns had different underrepresented nucleotides than the *prmt5* and *prmt4a;4b* retained introns.

Splicing of *UBP12* mRNA is similarly affected in *sic-3* and *dbr1-2*

To investigate gene expression and splicing changes associated with *sic* phenotypes, we compared changes in gene expression and splicing of transcripts associated with the 20 SIC-interacting proteins across four mutants: *sic-3*, *dbr1-2*, *prmt4a;4b*, and *prmt5*. Of the 20 SIC-interacting proteins, associated transcripts were differently affected across these four mutants. Transcripts associated with 10 of the genes were affected in *sic-3* and some other combination of mutants: four were affected only in *sic-3*, three were affected in all four mutants, two were affected only in *sic-3* and *dbr1-2*, and one was affected in all mutants but *dbr1-2* (Table 3).

The transcripts for the two SIC-interacting proteins affected only in *sic-3* and *dbr1-2*, but not in *prmt5* or *prmt4a;4b* which don't have lariat accumulation, may be involved in the shared lariat accumulation phenotype in *dbr1* and *sic*. *AQUARIUS* (*AQR*) was slightly upregulated in both *sic-3* and *dbr1-2* (DE with 0.5-0.6 $\log_2$FC) and one intronic bin was downregulated in *sic-3* (DU -1 $\log_2$FC) (Table 3). Changes in *UBP12*, the other gene with transcripts affected in *sic-3* and *dbr1-2*, were much more pronounced: in both mutants, DU of 6-12 bins had large changes in usage (Table 3). Additionally, *UBP13* was affected in all 3 mutants but much more so in *sic-3* and *dbr1-2* (Table 3).

UBP12 encodes a ubiquitin-specific protease that de-ubiquitinates proteins and processes poly-ubiquitin precursors. Ubiquitination is a reversible post-translational protein modification that affects many cellular processes including protein degradation and trafficking, cell signaling, and DNA damage response. Arabidopsis UBP12 and UBP13 can de-ubiquitinate proteins in-vitro (Cui et al., 2013) and are implicated in plant immunity regulation, flowering, and seed development.

Finally, the four proteins with transcripts affected only in *sic-3* were ISY1, XAB2, EXP4, and PDF1/PP2AA2. *EXP4* expression was downregulated, intronic bins for *PDF1* and *XAB2* were downregulated, and an intronic bin for *ISY1* was upregulated. These changes in expression and splicing may be specific to the *sic* mechanism for lariat accumulation, rather than shared between lariat accumulation in *sic* and *dbr1*.

Shared biological processes are affected in *sic-3* and *dbr1-2*

Finally we analyzed *sic-3* RNA-Seq data (Marshall, 2017) using Gene Ontology (GO) representation analysis for biological process. The Gene Ontology system categorizes genes by the biological processes in which they are involved, and representation analysis calculates the representation of genes from each biological process category in a data set compared to all the annotated genes in the genome. We analyzed the set of all genes that were differentially expressed in *sic-3* compared to WT, and ranked them by how overrepresented they were compared to the set of all Arabidopsis genes. The top 10 overrepresented GO terms for biological process in *sic-3* included five GO terms involving acid catabolism – three for specifically amino acid catabolism – two for light response, and two for starvation response (Supplemental Figure 4). This indicates that the processes most affected in *sic-3* are amino acid catabolism (probably part of starvation response), and light response.

I performed the same type of GO representation analysis for biological process on genes with differential expression compared to WT in *dbr1-2*. In *dbr1-2*, the top 10 overrepresented GO terms included three terms for photosynthesis (possibly five including protein-chromophore linkage and response to low light), two for vitamin E production, and two for circadian rhythms and processes (Supplemental Figure 5). Finally, I performed this analysis for genes that were regulated similarly in *sic-3* and *dbr1-2* (either upregulated in both or downregulated in both). For genes that were similarly regulated in both mutants, the overrepresented GO terms were response to salicylic acid, auxin, external stimulus, and stress (Supplemental Figure 6).

DISCUSSION

We have shown that *sic* and *dbr1* mutants both accumulate intron lariats from some of the same transcripts, which appears to arise from inhibited intron lariat degradation. SIC protein interactions and *sic* mutant phenotypes strongly suggest that SIC participates in some aspect of spliceosome function. SIC function is difficult to predict since SIC shows no homology to known splicing or splicing-related proteins. Thus, SIC is a previously unknown splicing-associated protein in Arabidopsis and in the Angiosperm lineage.

***sic* and *dbr1* mutants share pleiotropic phenotypes associated with miRNA deficiency**

Previous studies have shown that *sic* and *dbr1* both have morphological phenotypes associated with miRNA deficiencies. Additionally, both the *sic* and *dbr1* mutants have decreased miRNA abundance. A *dbr1* mutant has decreased levels of 32 different miRNAs, 17 of which are HYL1-dependent (Li et al., 2016). Additionally,

a subset of miRNAs checked in a *sic* mutant identified 10 miRNAs with deficiencies (Zhan et al., 2012). Comparing these two studies, 9 of the 10 SIC-dependent miRNAs are also reduced in *dbr1*, and all 10 are HYL1-dependent. This overlap between DBR1-dependent and SIC-dependent miRNAs, most of which are also HYL1-dependent, suggests a similar cause for miRNA reduction in *dbr1* and *sic* mutants, possibly through a common pathway involving HYL1. Previous work showed that the *dbr1* miRNA deficiency is caused by lariat accumulation (Li et al., 2016) and we saw that *dbr1* and *sic* mutants have similar lariat accumulation. Therefore, we suspect that lariat accumulation in *sic* causes the *sic* miRNA deficiency by this same HYL1-mediated mechanism shown for *dbr1*.

sic has global changes in splicing, especially intron retentions

We showed that *sic* mutants have both intron lariat accumulation and increased intron retentions, which may be connected. Previous research showed that decreased DBR1 activity traps snRNPs on accumulated intron lariats, which decreases snRNPs available to form new spliceosomes, and reduces spliceosome interactions with weak splice sites; this causes exon skipping following introns with weak splice sites (Han et al., 2017). In this human study, decreased DBR1 activity caused exon skipping; we propose that in plants this may instead lead to intron retention. Exon skipping is the most common type of alternative splicing event in mammals, but the least common in plants; in plants, intron retention is the most common type of alternative splicing event (Braunschweig et al., 2014; Marquez et al., 2014). Accordingly, decreased spliceosome interaction with weak splice sites could cause retention of those introns in plants, as opposed to skipping of the following exons in humans. Therefore, we propose that snRNP sequestration by accumulated intron lariats in *sic* could cause reduced splicing of introns with weak splice sites, leading to increased intron retention in *sic*.

Five intron lariats accumulate in both _sic_ and _dbr1_

We showed that *dbr1* and *sic* have intron lariat accumulation for 5 lariats from different genes. Since the only known function of DBR1 is intron lariat debranching, all *dbr1* phenotypes are likely due to decreased intron lariat debranching. Therefore, accumulation of intron lariats in *dbr1* may cause increased intron retentions in *dbr1* as detailed above. If intron lariat accumulation is sufficient to cause intron retentions in *dbr1*, and *sic-3* has similar lariat accumulation and intron retentions to *dbr1*, then intron lariat accumulation may also cause intron retentions in *sic-3*.

Timing of intron lariat removal is similar in _sic_ and _dbr1_ mutants

To date, intron lariat accumulation has been shown in *dbr1* mutants (Li et al., 2016), and in yeast *PRP43* and *CWC23* mutants – members of the NTR complex that are necessary for spliceosome disassembly (Arenas and Abelson, 1997; Sahi et al., 2010). Since the splicing reaction is reversible prior to mRNA release, and intron lariats accumulate in mutants affected during and after spliceosome disassembly, intron lariat accumulation may be limited to mutants affected during or after spliceosome disassembly. Since *sic* mutants accumulate lariats, we hypothesize that

SIC acts during the last steps of spliceosome disassembly or during intron lariat debranching.

Intron lariat intermediates (still bound to adjoining exon at 3' end) can accumulate temporarily in mutants blocked in the second catalytic step of splicing (Burgess and Guthrie, 1993). Therefore we investigated the ability of *sic* mutants to degrade intron lariats of *CCA1*, *LHY*, and *RVE4* long after gene expression turns off: this allows time for splicing of accumulated stalled lariat intermediates to be reversed, or for lariat intermediates to be degraded. We saw that *sic-1* had lariats throughout the time course, suggesting that *sic-1* has accumulated intron lariats that are somehow protected from debranching by DBR1, rather than accumulated lariat intermediates that could be reversed or degraded in 48h in the dark or 9h after the light pulse. It is unlikely that the absence of SIC could block both splicing reversal and degradation of defective splicing intermediates. Alternatively, it is possible that SIC could act in multiple steps of splicing, including spliceosome disassembly. Our list of putative SIC interactors includes proteins that act in the two catalytic steps of splicing (PP2A complex), as well as the NTC (XAB2, ISY1, AQR) which acts in every step from the first catalytic step to the intron lariat /post-catalytic spliceosome, onwards until spliceosome disassembly. SIC interaction with the NTC would be the simplest way for SIC to act in multiple steps of splicing.

We don't know what state intron lariats are in when DBR1 recognizes them for debranching. DBR1 could act on completely naked intron lariats, or on intron lariats associated with a minimal post-splicing spliceosome with or without snRNPs and NTC proteins. In humans, DBR1 debranches lariats bound by the Intron Small complex, which is not known to include any snRNPs or other proteins from the NTC (Yoshimoto et al., 2009). If the plant intron lariat debranching complex is different than the human IS and does include NTC proteins (or if lariats are debranched in an IL-like complex), then SIC interaction with DBR1 and putative interactions with NTC proteins XAB2, AQR, and ISY1 could all happen in this one lariat debranching complex. Alternatively, the spliceosome is remodeled from mRNA release to lariat debranching, and SIC may act in the transitions between complexes and thereby interact with all four proteins. Regardless, the presence of SIC appears to be critical for DBR1 to effectively debranch lariats.

Overall discussion

These two possible causal relationships we propose, that lariat buildup in *sic* causes both the miRNA decrease and increased intron retentions, would explain most of the phenotypes we have observed in the *sic* mutant: the developmental phenotypes similar to *dbr1*, splicing changes, and circadian clock defects. The SIC protein may have other activities besides affecting lariat buildup, which is another way SIC could affect all these processes separately, but ours is a compelling hypothesis by which lariat buildup could cause all the observed *sic* phenotypes. Future research and increased understanding of splicing mechanisms in plants could elucidate a mechanism of action for SIC.

Splicing is still very incompletely understood in plants, especially compared with splicing in humans and yeast. Here we show lariat accumulation in the Arabidopsis *sic* mutant analogous to lariat accumulation already shown for a weak

Arabidopsis *dbr1* mutant. Few lariat-accumulating mutants are known even in yeast. Even in metazoans, where much of splicing is characterized in detail, the post-splicing complexes/intron lariat spliceosome are less well understood. There remains a large gap in knowledge about DBR1 activity, how or whether it interacts with the spliceosome, and when and where it debranches lariats. We showed that SIC is critical for DBR1 function, an effect we predict is mediated through a protein-protein interaction. This work makes it clear that SIC is a splicing-associated protein that is likely to be important for lariat intron debranching in all Angiosperms.

MATERIALS AND METHODS

Plant materials and growth conditions for *Arabidopsis thaliana*

WT was the *Arabidopsis thaliana* Columbia-0 accession, and all mutants were in this background. We used previously-described mutants *sic-1* (Zhan et al., 2012) and *sic-3* (Marshall et al., 2016). The *dbr1-3* mutant is the previously uncharacterized Salk T-DNA insertion line SALK_041038 (Alonso et al., 2003) obtained from the Arabidopsis Biological Resource Center (abrc.osu.edu). This insertion is located in gene AT4G31770 at position 927 in the 5'UTR.

For all experiments, seeds were gas sterilized for 4 hours in a sealed container after combining 100mL bleach and 3mL hydrochloric acid. After gas sterilization, seeds were left in a sterile flow hood for 1 hour, combined with sterile water, and transferred to 4°C for 3 days of stratification in the dark. In this condition, seeds were either incubated at 4°C in water and plated at the end of stratification or incubated at 4°C after plating on sterile MS plates. All seeds were grown on 1X Murashige and Skoog basal salt medium (pH 5.7-5.8) with 0.8% micropropagation agar. After 3 days at 4°C, plates were transferred to a growth chamber with constant light (cool white fluorescent bulbs, 50µmol·m^{-2}·sec^{-}1) and 22°C for 3 days. After germination in LL 22°C, plates were either transferred to entrainment conditions (varied by experiment) or other experiment-specific conditions listed below.

Transient infiltration of *Nicotiana benthamiana*

Agrobacterium tumefaciens strain GV3101 was transformed separately with each construct. Cultures were grown in LB supplemented with 25ug/mL rifampicin and 25ug/mL gentamycin and the appropriate antibiotics for selection of plasmid vectors for 16h at 30°C with shaking, re-suspended in induction medium (50mM MES pH5.6, 0.5% glucose, 1.7mM NaH_2PO_4, 20mM NH_4Cl, 1.2mM $MgSO_4$, 2mM KCl, 17µM $FeSO_4$, 70µM $CaCl_2$, 200µM acetosyringone), and incubated for 4h at 30°C with shaking. Cultures were re-suspended, diluted to OD_{600} = 1 in 10mM MES with 200µM acetosyringone and infiltrated into young *Nicotiana benthamiana* leaves at OD_{600} = 0.25 for each construct.

Protein interaction analysis by co-immunoprecipitation

Expression constructs used were modified pK7-HFc (Huang et al., 2016) with 35s promoter and C-terminal fusion of 6xHis and FLAG tags (both SIC and DBR1), pEearleyGate101 (SIC), and pEearleyGate 201 (DBR1) (Earley et al., 2006). Constructs were infiltrated into *Nicotiana benthamiana* as described above. Immunoprecipitation protocol was generously provided by Mael Baudin, modified and performed as follows. Approximately 3cm^2 leaf samples were harvested from infiltrated leaves and immediately frozen in liquid nitrogen. 1mL of IP1 (50mM Tris-HCl pH8, 150mM NaCl, 0.5mM EDTA pH8, 0.2% Nonidet P-40) was added to ground samples, mixed gently, and incubated on a rotating wheel at 4°C for 15 minutes. Incubated samples were then centrifuged at 13krpm and supernatant was removed. 25uL of antibody-conjugated bead slurry was blocked in IP1 with 2% BSA and added to half of each supernatant. Supernatant and beads were incubated for 2

hours on a rotating wheel at 4°C. After incubation, samples were centrifuged at 5,000xg for 5m and supernatant was removed. Beads were washed 4x in IP2 (50mM Tric-HCl pH8, 150mM NaCl, 0.5mM EDTA pH8, 0.1% Nonidet P-40). Beads were then boiled for 5m at 95°C in SDS sample buffer (recipe) and loaded onto a 12% SDS-PAGE gel.

Protein interaction analysis by yeast two-hybrid

Protein-coding sequences were transformed into pGBKT7 bait and pGADT7 prey vectors. Bait and prey vectors were transformed individually and in combination into yeast strain Y2HGold using yeast transformation protocol from (Clontech, 2009) and plated on SD-Leu-Trp with 50ug/mL Kanamycin. 3 individual transformants for each plasmid combination were grown in SD-Leu-Trp with 50ug/mL Kanamycin to OD_{600} = 1 and spotted with 10uL per spot at OD_{600} = 1 and 5 subsequent 1:10 serial dilutions onto SD-Leu-Trp and SD-Leu-Trp-His-Ade with 50ug/mL Kanamycin, with and without 200ng/mL Aureobasidin.

Co-localization

Coding sequences were cloned into pEG101 for SIC-YFP and pEG201 for DBR1-CFP. Transient infiltrations into *Nicotiana benthamiana* as described earlier in methods. Confocal images were taken of fresh leaf pieces 48h after infiltration.

Growth conditions for lariat time course

Seeds for lariat time course were gas sterilized as described above and plated in closely-spaced spots on MS plates. After 3 days stratification as described above, plates were transferred to constant light and 22°C for 5 days, then dark and 22°C for 2 days. Plants were given a 1h light treatment ($50\mu mol \cdot m^{-2} \cdot sec^{-1}$) and sampled immediately before light treatment, immediately after, 1h after, and 9h after. Transfers between conditions and light treatment all occurred at ZT0.

Gene expression analysis with RT-qPCR

Genomic DNA was removed from total RNA using dsDNase kit (Fisher Scientific Company). First-strand cDNA synthesis was primed with poly(dT) using 1.8ug of DNase-treated RNA. RT-qPCR was performed with CFX Real-Time system (Bio-Rad Laboratories). Cq for each amplification curve was calculated in regression mode using Bio-Rad CFX Manager 3.0 software. Cq values were normalized using the equation $2^{(Cq\ for\ normalizer - Cq\ for\ experimental\ sample)}$ and the "normalizer" was the geometric mean of C_q values for two reference genes. Two technical replicates were averaged for each sample. For all Arabidopsis samples, reference genes were IPP2 and PP2A.

Intron lariat PCR

cDNA was prepared as above except synthesis was primed with random primer. Intron lariat PCR was performed as previously described (Ohi et al., 2007). PCR mix included 10x ExTaq buffer with $MgCl_2$ (Takara), 0.2mM dNTPs, 0.25µM primers, and homemade hTaq. PCR program was 96°C for 2m; 35 cycles of 96°C 20s, annealing 30s, 72°C 20s; 72°C 2m.

RNA-Seq ASpli analysis

The ASpli program (Mancini et al., 2019) was used for analysis of previously-described RNA-Seq data comparing *sic-3* to WT at 16°C (Marshall, 2017) and previously-published RNA-Seq data comparing *dbr1-2* to WT (Li et al., 2016). ASpli was run with settings as in (Hernando et al., 2015).

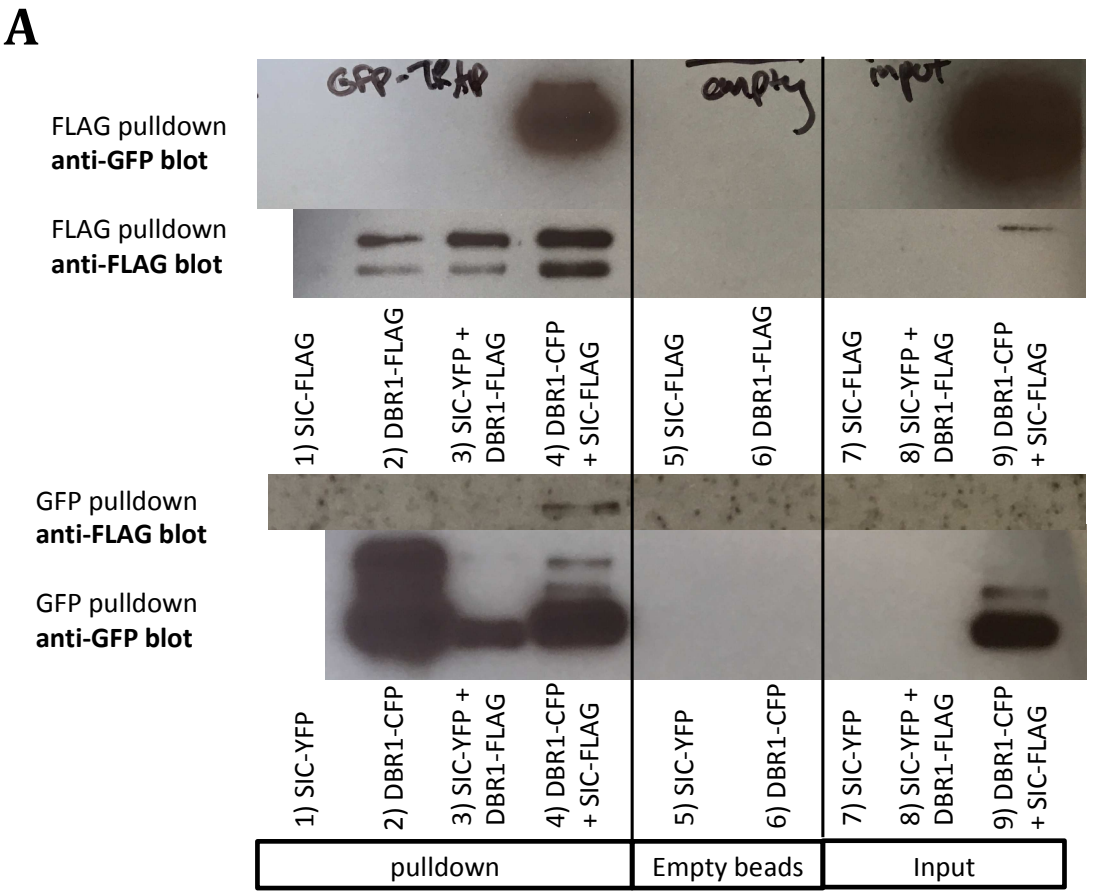
Splice site enrichment analysis

For each of the 43 intron retention event called as differentially utilized in *sic-3* at 16°C compared to WT by ASpli analysis (false discovery rate <0.1, $\log_2$ fold-change >0.58), the sequence 10 nucleotides upstream and downstream of the invariant G in the 5' splice site was retrieved from the *Arabidopsis thaliana* Ensembl 35 genome annotation. These sequences were used together to generate a sequence logo using WebLogo (Crooks et al., 2004), representing the consensus sequence around the 5' splice site in these retained introns. This was compared with a previously-published sequence logo for all Arabidopsis GT_AG_U2 introns (Hernando et al., 2015).

Gene ontology analysis

Gene Ontology enrichment analysis for biological process (The Gene Ontology Consortium et al., 2011; Mi et al., 2017; Carbon et al., 2019) was performed online using differentially expressed genes (false discovery rate <0.05, $\log_2$ fold-change >0.58) from RNA-Seq analysis with ASpli as described above.

Chapter 1 Figures



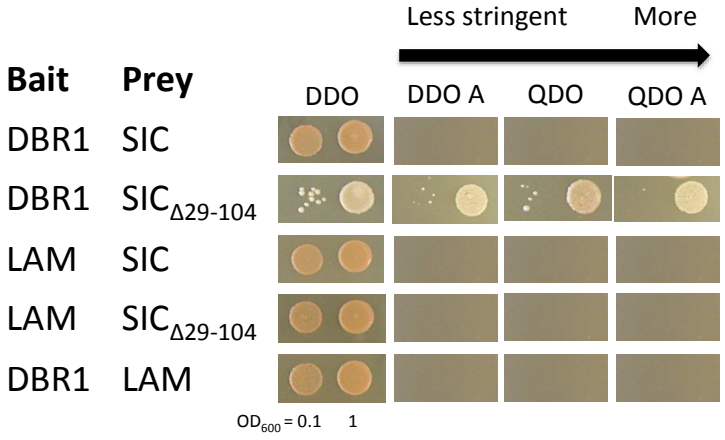


Figure 1. SIC interacts with DBR1 in co-immunoprecipitation and yeast two-hybrid experiments.

A) Western blots of CoIP using single and co-infiltrated tagged constructs for SIC and DBR1. Tissue sampled for CoIP from *Nicotiana benthamiana* leaves 48h after Agrobacterium infiltration with constructs. GFP pulldown using GFP-TRAP magnetic

beads (Chromotek), FLAG pulldown using agarose beads conjugated to anti-OctA sc-807 (Santa Cruz Biotechnology). GFP blot with anti-GFP sc-9996, FLAG blot with anti-OctA sc-807 (Santa Cruz Biotechnology). Empty beads were agarose beads as used for OctA pulldown, but without conjugated antibody. Input was extracted protein before incubation with beads.
B) Yeast two-hybrid assay for protein interaction. Pictures from 4d after spotting cultures on plates. In each picture, yeast spot on right is OD_{600} = 1, and spot on left is 0.1. Media is SD with DDO=double dropout (SD-Leu-Trp), QDO=quadruple dropout (SD-Leu-Trp-His-Ade), A=aureobasidin. Negative-control pairings are with LAM (human lamin C, does not interact with most proteins). Three independent yeast transformants grown per experiment in two independent experiments.

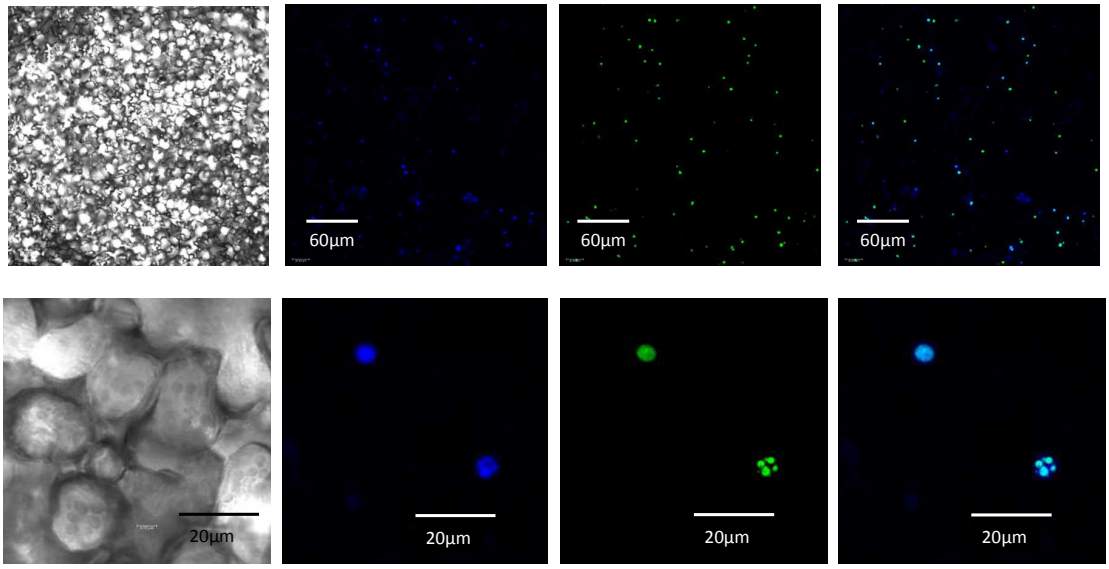


Figure 2. SIC and DBR1 co-localize in nuclear speckles.

Left image is brightfield, then DBR1-CFP, then SIC-YFP, then overlay. Top photos show many nuclei with overlap of CFP and YFP signal for most. Bottom photos show two nuclei with full signal overlap, the top with fluorescence throughout the nucleus and the bottom in nuclear speckles. SIC-YFP and DBR1-CFP transient infiltrations in *Nicotiana benthamiana*. Pictures taken 48h after infiltration. Confocal images of fresh leaves.

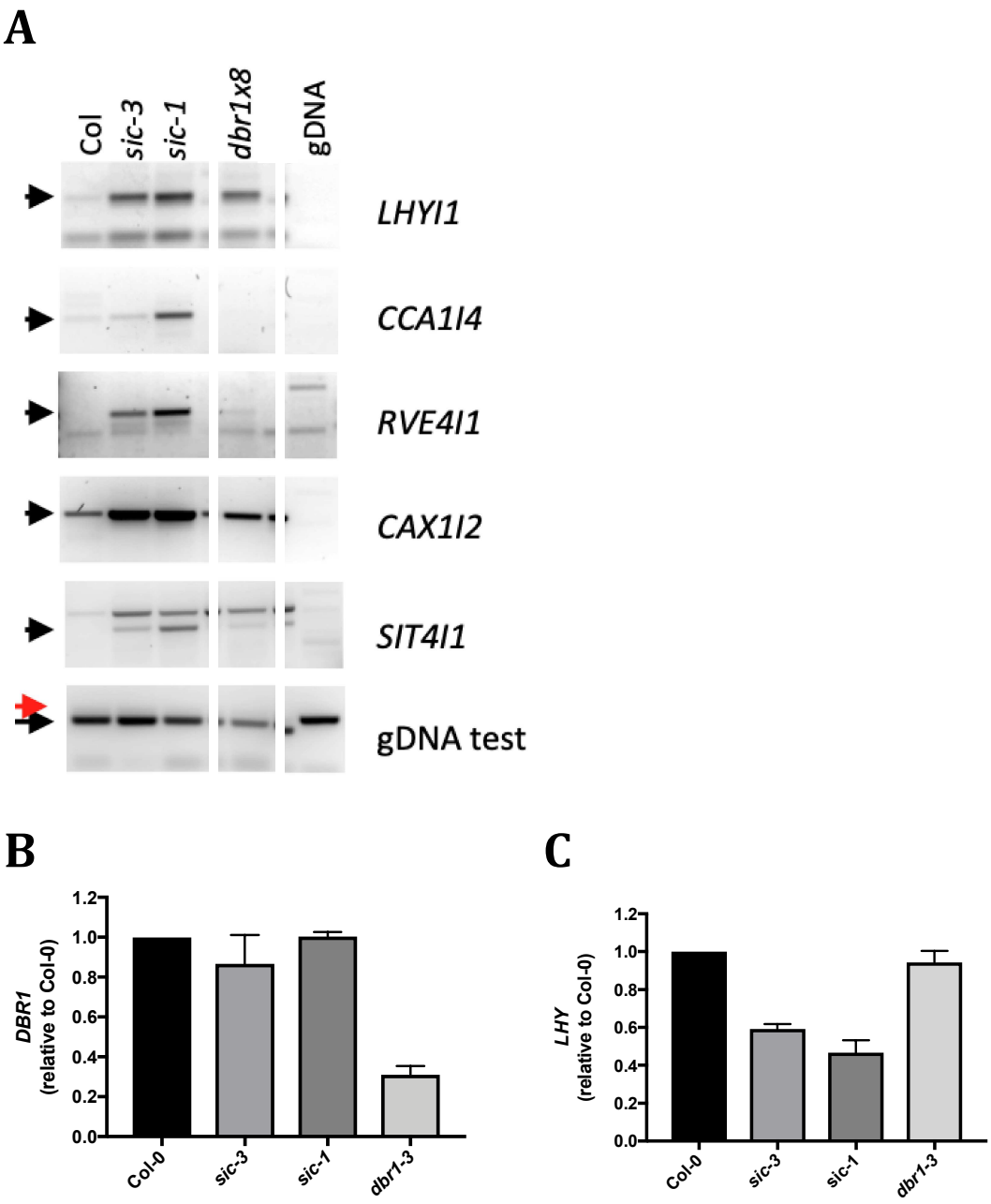


Figure 3. Five intron lariats accumulate in *sic-1* and *dbr1-3.*

A) PCR for lariat sequences from five transcripts in WT, *sic* mutants, *dbr1-3* mutant, and sample containing only gDNA (negative control for primer amplification of gDNA sequences). Images shown are of ethidium bromide stained gels with contrast inverted. Black arrows indicate band size that represents lariat sequence. In bottom panel, bands are from primers meant to amplify both gDNA and cDNA. In "gDNA test" row, smaller band indicated with black arrow is product size for 3'-end of *LHY* cDNA, red arrow indicates larger band size for gDNA. All samples were run on the same gel and intervening lanes were removed to improve clarity.

B) RT-qPCR showing *DBR1* knockdown in *dbr1-3* mutant and **C)** expression level of LHY transcript from which *LHYI1* lariat shown originates. Mean with + SD.

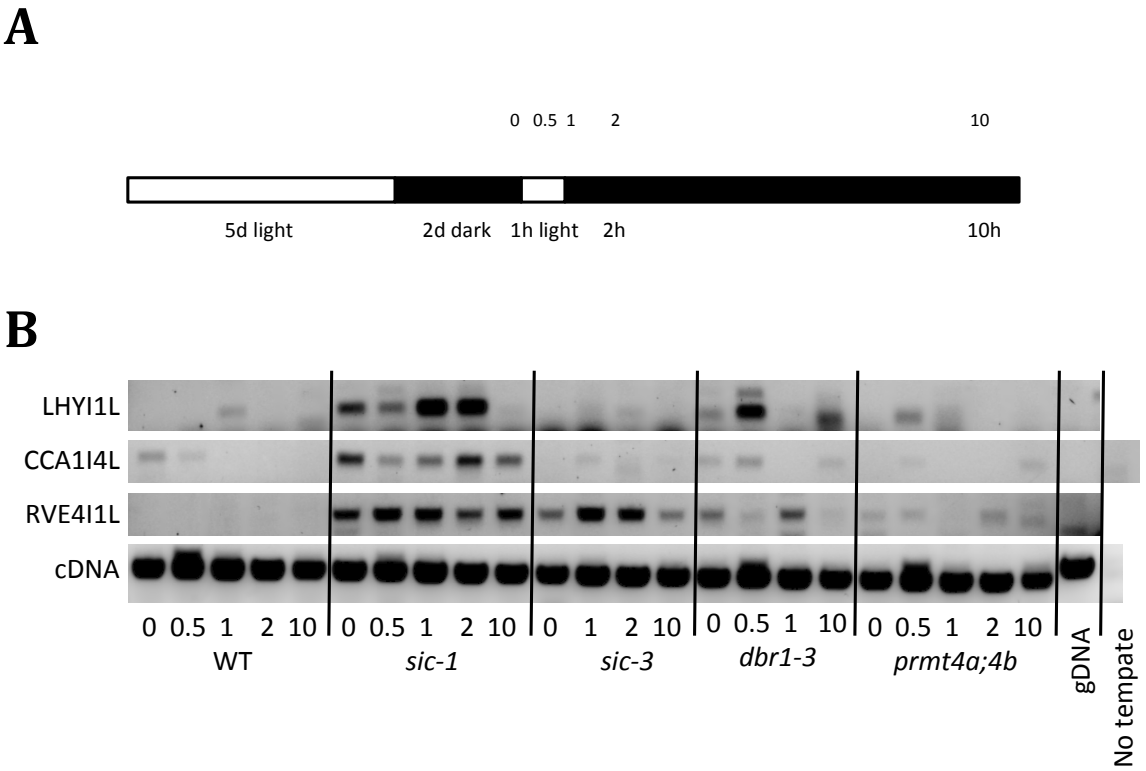


Figure 4. Lariats are more stable in *sic-1* than in *dbr1-3*.

A) Experimental design, where black bars represent dark conditions and white bars represent light conditions, all at 22°C. Blue arrows represent times samples were taken, samples numbered by hours after beginning of light pulse.

B) Gel of lariat PCR products for three different lariats, *LHYI1*, *CCA1I4*, and *RVE4I1*; cDNA control PCR to test for the presence of contaminating gDNA. *sic-3* is missing the 0.5h time point and *dbr1-3* is missing the 2h time point. No-template controls for *LHYI1L* and *CCA1I4L* PCR were combined and are shown in row 2, and no-template controls for *RVE4I1L* and cDNA control were combined and are shown in row 4. No-template controls show no bands in the absence of template.

All samples were run on the same gel and intervening lanes were removed to improve clarity. Images shown are of ethidium bromide stained gels with contrast inverted.

C) RT-qPCR for expression of full *LHY* and *CCA1* transcripts from which the *LHYI1* and *CCA1I4* lariats originate.

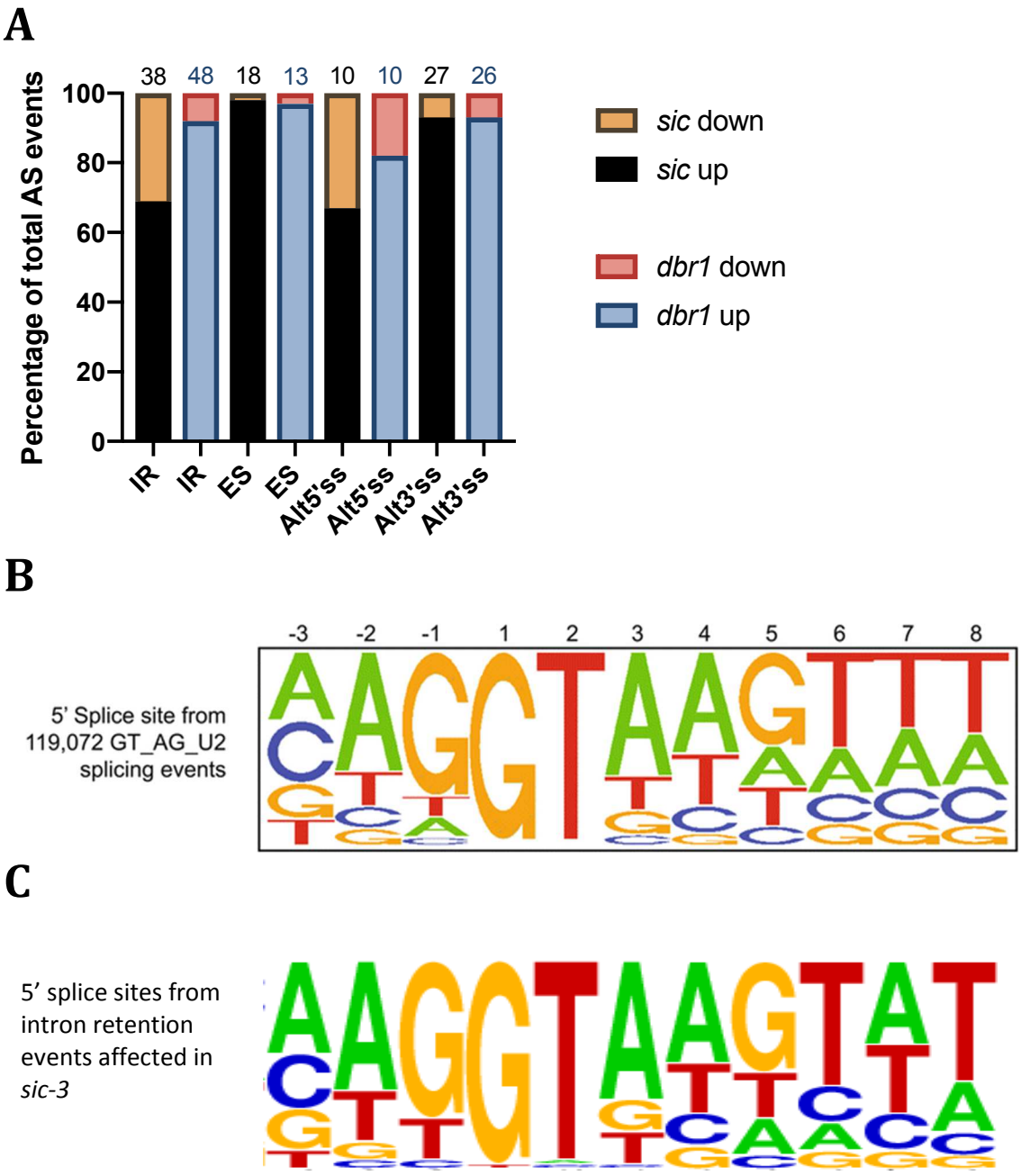


Figure 5. Intron retentions are most affected in *sic-3* and *dbr1-2*, and 5' splice sites for *sic-3* intron retentions deviate from the WT consensus.

A) Data shown is also in Table 2. All alternative splicing events in *sic-3* and *dbr1-2* mutants compared to respective WT controls. Percentage of total splicing events that correspond to each type of event are listed above bars. Percentage of each type of event that is up or downregulated is graphed in black and yellow (*sic-3*) or red and blue (*dbr1-2*).
B,C) Consensus sequences around the 5' splice site in **B)** WT and **C)** *sic-3* represented as nucleotide frequencies at each position visualized with SeqLogo.

B) Previously published consensus 5' splice site sequence for all Arabidopsis GT_AG_U2 introns (Hernando et al., 2015).
C) Nucleotide frequencies from all intron retention events affected in *sic-3* compared to WT.

TABLES

Table 1. Putative SIC-interacting proteins identified by co-immunoprecipitation-mass spectrometry.

Arabidopsis gene	Arabidopsis annotation	Homology	Possible function	PP2A complex	PRP19/ NTC	spliceosome-associated	mRNA processing
AT5G06600	UBP12			x		x	
AT3G11910	UBP13			x		x	
AT1G51690	ATB ALPHA			x		x	
AT1G17720	ATB BETA			x		x	
AT3G58500/AT2G42500	PP2A			x		x	
AT1G25490	RCN1			x		x	
AT3G25800	PDF1, PP2AA2			x		x	
AT5G28740	transcription-coupled DNA repair protein-related	human XAB2			x	x	x
AT2G38770	EMB2765	human AQR	putative RNA helicase; CM predicted intron-binding?		x	x	x
AT4G31770	DBR1					x	x
AT3G06930	PRMT4B					x	x
AT3G18790	unknown	ISY1	TC-NER pre-incision complex, DNA repair; mRNA splicing pathway		x	x	x
AT5G49020	PRMT4A					x	x

SIC-interacting proteins determined by previously published affinity purification mass spectrometry experiment (Karampelias et al., 2016). Literature curation used to identify proteins and their complex identities.

Table 2. Types and percentages of alternative splicing in *sic-3* and *dbr1-2*.

AS with DU compared to WT		
	sic-3	***dbr1-2***
Total Events	537	1291
IR	38%	48%
w/DU +	69%	92%
ES	18%	13%
w/DU +	98%	97%
Alt3SS	10%	10%
w/DU +	93%	93%
Alt5SS	27%	26%
w/DU +	67%	82%

We used ASpli (Mancini et al., 2019) to analyze our RNA-Seq experiment comparing *sic-3* and WT, and previously-published RNA-Seq experiment comparing *dbr1-2* to WT (Li et al., 2016).

At the top is the total number of alternative splicing (AS) events annotated by ASpli analysis that have differential usage (DU) of splice sites in the mutant compared to WT. Below the total is the percentage of total events annotated as each type: IR (intron retention), ES (exon skipping), 3'Alt (alternative 3' splice site usage), 5'Alt (alternative 5' splice site usage), or Multiple (splice junction had multiple types of alternative splicing events). Below the percentage of events of a given type is the percentage of these that had higher differential usage (as opposed to lower) in the mutant compared to WT.

Table 3. Changes in expression and splicing of transcripts of SIC-interacting proteins in *sic-3*, *dbr1-2*, *prmt4a;4b*, and *prmt5* mutants compared to WT.

			Does DU or DE change in mutant compared to WT?				
gene	**name**	**annotations**	***sic-3***	***dbr1-2***	***prmt4a;4b***	***prmt5***	**changed in:**
AT1G17720	ATB BETA	SIC-interacting	Y	Y	Y	Y	all
AT1G51690	ATB ALPHA	SIC-interacting	Y	Y	Y	Y	all
AT3G11910	UBP13	SIC-interacting	Y	Y	Y	Y	all
AT3G58500/ AT2G42500	PP2A (2 isoforms)	SIC-interacting	Y	Y	Y	Y	all
AT2G38770	AQUARIUS (AQR)	SIC-interacting	Y	Y			*sic* & *dbr1*
AT5G06600	UBP12	SIC-interacting	Y	Y			*sic* & *dbr1*
AT5G49020	PRMT4A	SIC-interacting	Y		Y	Y	all but *dbr1*
AT2G39700	EXP4	SIC-interacting	Y				*sic* only
AT3G18790	ISY1	SIC-interacting	Y				*sic* only
AT3G25800	PDF1, PP2AA2	SIC-interacting	Y				*sic* only
AT5G28740	XAB2	SIC-interacting	Y				*sic* only
AT1G25490	RCN1	SIC-interacting		Y	Y		
AT3G06930	PRMT4B	SIC-interacting		Y	Y		
AT5G53620	unknown protein	SIC-interacting		Y		Y	
AT1G42430	unknown protein	SIC-interacting		Y		Y	
AT3G47650	BSD2 family	SIC-interacting		Y			
AT4G31770	DBR1	SIC-interacting		Y			
AT4G29350	PFN2 (PROFILIN 2)	SIC-interacting			Y		

All RNA-Seq experiments analyzed with ASpli. Includes our sic-3 RNA-Seq data and previously-published RNA-Seq data comparing *dbr1-2* to WT (Li et al., 2016) and *prmt4a;4b* and *prmt5* to WT (Hernando et al., 2015).

Table 4. Changes in alternative and constitutive splicing of UBP12 in *sic-3* and *dbr1-2*.

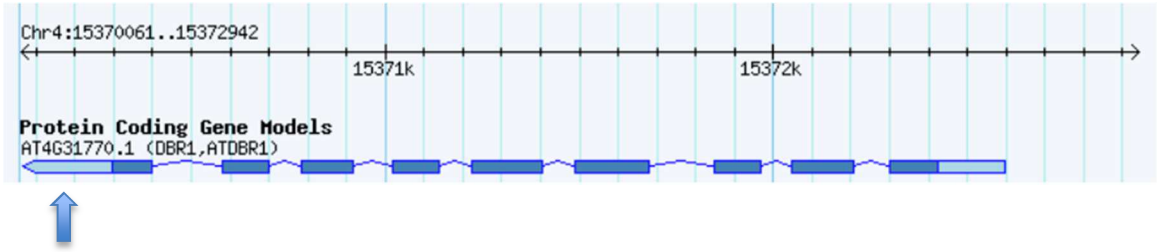
		event	locus	locus_overlap	symbol	gene_coordinates	start	end	length	logFC	pvalue	bin.fdr
dbr1	AT5G06600:E008	IR	AT5G06600	-	AT5G06600	Chr5:2019084-2028116	2020325	2020693	369	0.69	0.00	0.05
sic3	AT5G06600:I006	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2020760	2020909	150	-1.01	0.00	0.02
sic3	AT5G06600:I007	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2020976	2021059	84	-1.30	0.00	0.01
sic3	AT5G06600:I008	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2021132	2021216	85	-1.22	0.00	0.00
sic3	AT5G06600:I014	-	AT5G06600	-	AT5G06600	5:2019107-2027945	**2022464**	**2022672**	209	-1.02	0.00	0.04
dbr1	AT5G06600:I014	-	AT5G06600	-	AT5G06600	Chr5:2019084-2028116	**2022464**	**2022672**	209	2.10	0.00	0.00
sic3	AT5G06600:I022	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2024590	2024871	282	-1.08	0.00	0.01
sic3	AT5G06600:I023	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2024982	2025078	97	-1.20	0.00	0.02
dbr1	AT5G06600:I025	-	AT5G06600	-	AT5G06600	Chr5:2019084-2028116	2025579	2025719	141	-2.45	0.00	0.03
sic3	AT5G06600:E027	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2025720	2025828	109	0.92	0.00	0.07
sic3	AT5G06600:E028	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2025916	2026134	219	0.94	0.00	0.06
sic3	AT5G06600:I027	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2026135	2026231	97	-2.30	0.00	0.03
dbr1	AT5G06600:I028	-	AT5G06600	-	AT5G06600	Chr5:2019084-2028116	2026498	2026818	321	1.03	0.00	0.01
sic3	AT5G06600:I029	-	AT5G06600	-	AT5G06600	5:2019107-2027945	2026977	2027078	102	-1.17	0.00	0.02
sic3	AT5G06600:I030	-	AT5G06600	-	AT5G06600	5:2019107-2027945	**2027230**	**2027435**	206	-0.99	0.00	0.06
dbr1	AT5G06600:I030	-	AT5G06600	-	AT5G06600	Chr5:2019084-2028116	**2027230**	**2027435**	206	1.31	0.00	0.00
sic3	AT5G06600:Io032	-	AT5G06600	-	AT5G06600	5:2019107-2027945	**2027503**	**2027804**	302	1.06	0.00	0.01
dbr1	AT5G06600:I032	-	AT5G06600	-	AT5G06600	Chr5:2019084-2028116	**2027506**	**2027804**	299	1.44	0.00	0.00
sic3	AT5G06600:I032	-	AT5G06600	-	AT5G06600	5:2019107-2027945	**2027506**	**2027804**	299	1.07	0.00	0.02

SUPPLEMENTAL FIGURES

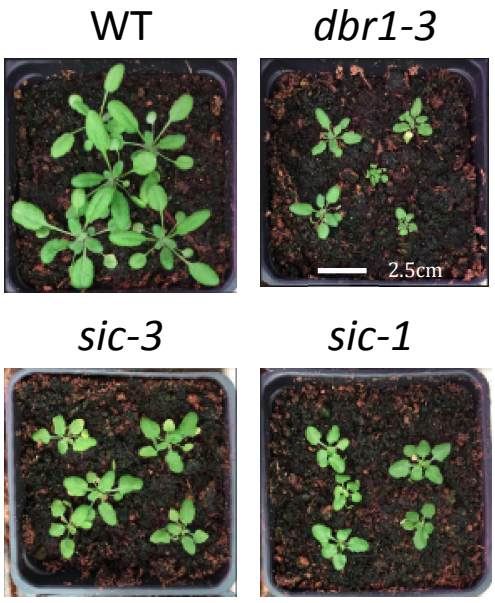
Supplemental Figure 1. DBR1 mutant allele *dbr1-3*.

A) *DBR1* gene model shown with light blue bars for UTRs, dark blue for exons, and lines for introns. *dbr1-3* T-DNA insertion is in 5' UTR, indicated by blue arrow.
B) *dbr1-3* rosettes are similar to *sic-1* and *sic-3* rosettes, and are small compared to WT. All plants photographed at 4 weeks after germination.

A



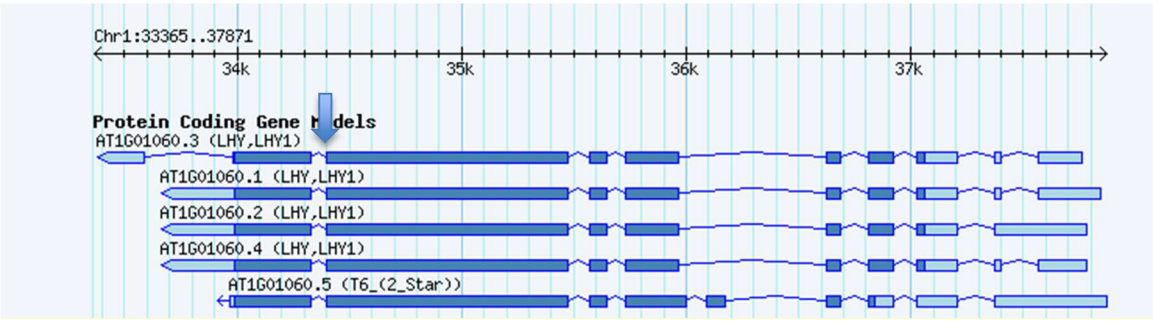
B



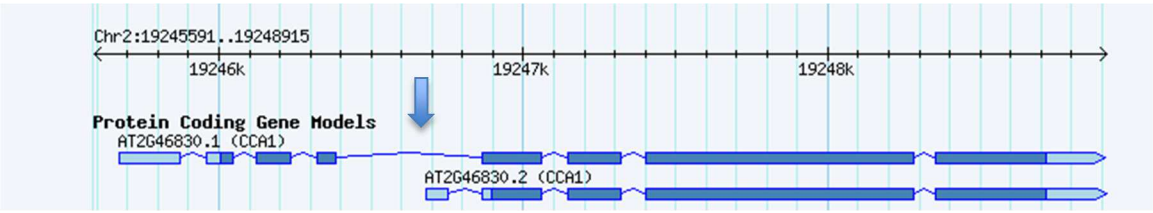
Supplemental Figure 2. Introns from which lariats originate and primer location

Gene models for lariats measured by PCR, with lariat-producing intron indicated.

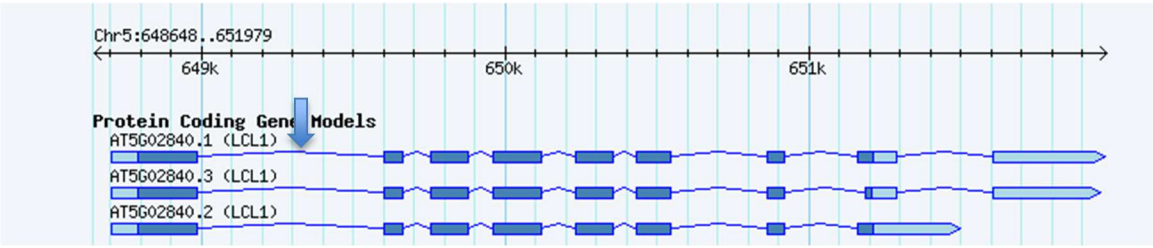
LHYI1 lariat



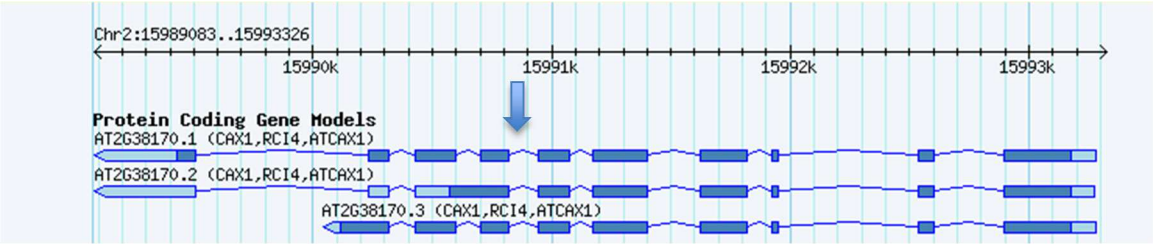
CCA1I4 lariat



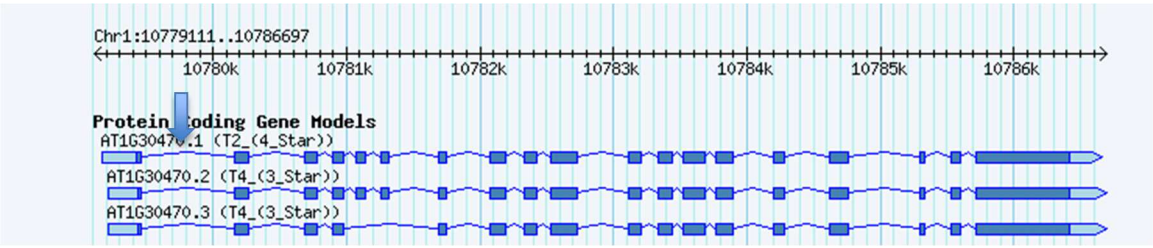
RVE4I1 lariat



CAX1I4 lariat

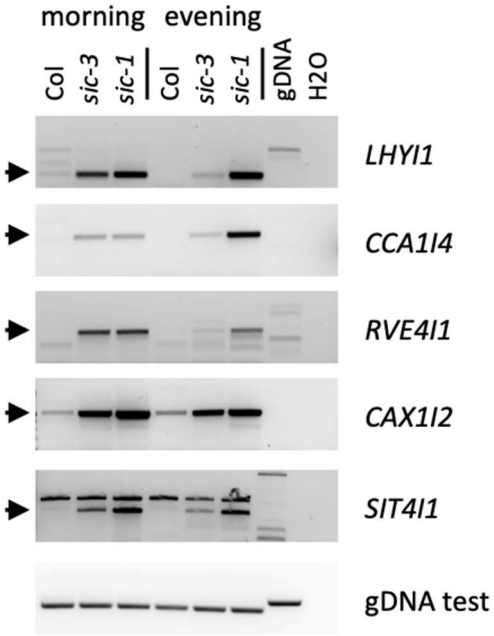


SIT4I1 lariat



Supplemental Figure 3. *sic* mutants exhibit increased accumulation of five lariats.

Lariat endpoint PCR for RNA used in *sic-3* RNA-Seq experiment. Lariat accumulation is similar to separate experiment shown in Figure 3.



Supplemental Figure 4. GO enrichment analysis for biological process using DE genes in *sic-3.*

Top 10 significantly overrepresented genes by fold enrichment shown (p<0.05). Analysis and results summaries below.

	Arabidopsis thaliana (REF)	upload_1 (Hierarchy) NEW!					
GO biological process complete	#	#	expected	▽ Fold Enrichment	+/-	raw P value	FDR
branched-chain amino acid catabolic process	18	10	2.08	4.80	+	3.22E-04	2.24E-02
response to absence of light	41	15	4.75	3.16	+	4.40E-04	2.89E-02
cellular amino acid catabolic process	77	28	8.92	3.14	+	2.04E-06	4.16E-04
alpha-amino acid catabolic process	69	24	7.99	3.00	+	1.94E-05	2.39E-03
organic acid catabolic process	139	39	16.10	2.42	+	5.80E-06	9.03E-04
carboxylic acid catabolic process	139	39	16.10	2.42	+	5.80E-06	8.80E-04
defense response to fungus	232	58	26.87	2.16	+	9.28E-07	2.03E-04
response to starvation	180	44	20.84	2.11	+	3.54E-05	3.61E-03
cellular response to starvation	148	36	17.14	2.10	+	2.16E-04	1.70E-02
response to light intensity	156	37	18.07	2.05	+	2.11E-04	1.69E-02

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20190429)

Annotation Version and Release Date: GO Ontology database Released 2019-02-02

Analyzed List: upload_1 (Arabidopsis thaliana) Change

Reference List: Arabidopsis thaliana (all genes in database) Change

Annotation Data Set: GO biological process complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	27581 out of 27581	3168 out of 3194
Unmapped IDs:	0	494
Multiple mapping information:	0	50

Supplemental Figure 5. GO enrichment analysis for biological process using DE genes in *dbr1.*

Top 10 significantly overrepresented genes by fold enrichment shown (p<0.05). Analysis and results summaries below.

	Arabidopsis thaliana (REF)	upload_1 (Hierarchy) NEW!					
GO biological process complete	#	#	expected	▽ Fold Enrichment	+/-	raw P value	FDR
photosynthesis, light harvesting in photosystem I	24	17	1.65	10.29	+	1.81E-10	5.63E-08
vitamin E metabolic process	8	5	.55	9.08	+	9.17E-04	3.41E-02
vitamin E biosynthetic process	8	5	.55	9.08	+	9.17E-04	3.39E-02
photosynthesis, light harvesting in photosystem II	10	6	.69	8.71	+	3.23E-04	1.42E-02
protein-chromophore linkage	28	16	1.93	8.30	+	6.11E-09	1.20E-06
response to low light intensity stimulus	19	10	1.31	7.64	+	7.74E-06	5.80E-04
starch catabolic process	17	8	1.17	6.83	+	1.17E-04	6.25E-03
syncytium formation	16	7	1.10	6.35	+	4.46E-04	1.87E-02
photosynthesis, light harvesting	46	18	3.17	5.68	+	6.98E-08	9.18E-06
circadian rhythm	108	40	7.44	5.38	+	3.85E-15	3.25E-12
rhythmic process	119	43	8.19	5.25	+	7.39E-16	7.29E-13

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20190429)

Annotation Version and Release Date: GO Ontology database Released 2019-02-02

Analyzed List: upload_1 (Arabidopsis thaliana) Change

Reference List: Arabidopsis thaliana (all genes in database) Change

Annotation Data Set: GO biological process complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	27581 out of 27581	1887 out of 1899
Unmapped IDs:	0	323
Multiple mapping information:	0	56

Supplemental Figure 6. GO enrichment analysis for biological process using DE genes with same sign logFC in *sic-3* and *prmt4a;4b.*

All significantly overrepresented genes shown (p<0.05). Analysis and results summaries below.

	Arabidopsis thaliana (REF)	upload_1 (Hierarchy NEW!)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
response to salicylic acid	200	9	1.15	7.81	+	3.53E-06	2.98E-03
↳response to antibiotic	304	10	1.75	5.71	+	1.44E-05	8.54E-03
↳response to chemical	2776	36	16.00	2.25	+	3.80E-06	2.81E-03
↳response to stimulus	5729	64	33.03	1.94	+	3.62E-08	2.14E-04
↳response to drug	593	13	3.42	3.80	+	4.99E-05	2.46E-02
↳response to organic cyclic compound	327	12	1.89	6.37	+	6.54E-07	1.29E-03
↳response to organic substance	1915	27	11.04	2.45	+	2.52E-05	1.35E-02
↳response to hormone	1599	26	9.22	2.82	+	1.92E-06	2.27E-03
↳response to endogenous stimulus	1615	26	9.31	2.79	+	2.30E-06	2.26E-03
response to auxin	388	14	2.24	6.26	+	9.01E-08	2.67E-04
response to external stimulus	1387	24	8.00	3.00	+	1.76E-06	2.60E-03
response to stress	3157	38	18.20	2.09	+	1.43E-05	9.40E-03

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20190429)

Annotation Version and Release Date: GO Ontology database Released 2019-02-02

Analyzed List: upload_1 (Arabidopsis thaliana) Change

Reference List: Arabidopsis thaliana (all genes in database) Change

Annotation Data Set: GO biological process complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	27581 out of 27581	157 out of 159
Unmapped IDs:	0	12
Multiple mapping information:	0	2

Chapter 2: SIC and PRMT4 interact directly, but have different effects on the circadian clock and salt tolerance

ABSTRACT

Previous research identified PRMT4A and PRMT4B protein arginine methyltransferases as SIC-interacting proteins. PRMT4A and PRMT4B together regulate RNA splicing, possibly through direct methylation of splicing factors. Besides PRMT4A and PRMT4B, all other SIC-interacting proteins that are involved in splicing are known members of a large splicing-associated complex (PP2A or NTC), except for DBR1. We show that SIC and PRMT4A interact directly, and PRMT4B may interact directly but weakly with SIC, or indirectly through heterodimerizing with PRMT4A. Although SIC and PRMT4A interact directly, the *sic* and *prmt4a;4b* mutants have different circadian clock and salt tolerance phenotypes: *sic* has a cool temperature-sensitive long period circadian clock and *prmt4a;4b* has no clock phenotype; *sic* has decreased and *prmt4a;4b* has increased salt tolerance compared to WT. Therefore, SIC and PRMT4A;4B likely do not act together to affect either the circadian clock or salt tolerance. Rather, the SIC-PRMT4A interaction may indicate that SIC is a substrate of PRMT4A methyltransferase activity.

INTRODUCTION

The PRMT family of Protein Arginine Methyltransferases catalyze the addition of methyl groups to proteins, nucleic acids, and other small molecules (Bedford and Richard, 2005). PRMT4A and PRMT4B are type I PRMTs, and are paralogs thought to be redundant in function based on genetic and biochemical tests (Hernando et al., 2015). PRMT4A and PRMT4B are thought to be important for the regulation of RNA splicing, similar to PRMT5, possibly through direct methylation of spliceosomal proteins or by influencing the coupling of transcription and RNA processing (Cheng et al., 2007; Kuhn et al., 2011). In humans, PRMT4 (known as CARM1) can methylate splicing factors including parts of the U1 and U2 snRNPs (Meyer et al., 2015), which interact with other splicing protein-RNA complexes to form the spliceosome.

Generally, arginine methyltransferases can act inside the nucleus to methylate DNA or protein, or outside the nucleus to methylate proteins (McBride and Silver, 2001). In mammals, PRMT1, 3, 6, and 8 but not PRMT4 have been shown to methylate GAR motifs in their target proteins; however, Ahmad et al. (2011) saw all rice PRMTs (OsPRMTs), including OsPRMT4, methylate protein GAR motifs in vitro. GAR motifs are found in many RNA-binding proteins, and are thought to provide these proteins with the structural integrity necessary for their RNA binding function (Ahmad et al., 2011). Arginine methylation can also block or promote protein-protein interactions (McBride and Silver, 2001).

Arabidopsis has seven type I PRMTs (PRMT1A and B, PRMT3, PRMT4A and B, PRMT6, and PRMT10) and one type II (PRMT5) (Hernando et al., 2015). Arabidopsis PRMT4A and PRMT4B, which both form homodimers and together form heterodimers, localize to both the nucleus and cytoplasm (Niu et al., 2008). Though the mechanism is still unclear, the Arabidopsis *prmt4a;4b* mutant does have global changes in gene regulation and splicing (Hernando et al., 2015). Some splicing changes in *prmt4a;4b* may be due to PRMT4A;4B together stabilizing weak RNA-RNA interactions to enable splicing at non-consensus donor splice sites (Hernando et al., 2015).

Arabidopsis SICKLE (SIC) is a protein of unknown function and *sic* mutations globally affect splicing. A hypothesis for SIC function is that it is a splicing factor. Prior work suggests that PRMT4 and SIC occur together in a multiprotein complex of several splicing-associated proteins (Karampelias et al., 2016). It is possible that these two proteins function together through a direct physical interaction, and that a PRMT4-SIC complex plays an important role in splicing. This Chapter describes confirmation that SIC interacts directly with PRMT4A, but not PRMT4B, according to yeast two-hybrid (Y2H). Testing of *sic* and *prmt4a;4b* mutants for common phenotypes found that while *sic* and *prmt4a;4b* share common transcript splice variants, *prmt4a;4b* lacks the circadian clock and salt sensitivity phenotypes of *sic*.

RESULTS

SIC interacts with PRMT4A

We used yeast two-hybrid (Y2H) to investigate whether SIC interacts directly with PRMT4A or PRMT4B, as indicated by a previously published

immunoprecipitation-mass spectrometry (IP-MS) experiment with SIC (Karampelias et al., 2016) (Chapter 1). To test whether any detected SIC-PRMT4 interactions were specific to PRMT4 class enzymes, we also tested PRMT5 for interaction with SIC. In addition to full-length SIC, we also included $SIC_{\Delta 29\text{-}104}$ that is the predicted protein product of one of two unique transcript splice forms in the *sic-3* allele (Chapter 1).

Both SIC and $SIC_{\Delta 29\text{-}104}$ interacted equally strongly with PRMT4A, with similar growth density on all three selective media (Figure 1A); this indicates that the SIC-DBR1 interaction has a special requirement for $SIC_{\Delta 29\text{-}104}$ (Chapter 1). None of the negative-control pairings with human lamin (LAM) as the bait or prey construct grew on any selective media, so this interaction is specific and is not the result of auto-activation from individual constructs (Figure 1A). Neither SIC nor $SIC_{\Delta 29\text{-}104}$ paired with PRMT4B nor PRMT5 interacted and grew on selective media (Figure 1A). Therefore SIC and PRMT4A interact directly, and this interaction is specific to PRMT4A, as there is no interaction between SIC and PRMT4B or PRMT5. Further, PRMT4B likely appeared in the SIC IP-MS experiment because it interacts indirectly with SIC by heterodimerizing with PRMT4A.

PRMT4A may interact with DBR1

Since we found that SIC interacts with both DBR1 and PRMT4A via Y2H, we used Y2H to further investigate whether DBR1 and PRMT4A also interact directly with one another. Indeed, DBR1 and PRMT4A interact in Y2H, and DBR1 does not interact with PRMT4B (Figure 1B). Shown is the appropriate negative-control pairing of DBR1 with LAM, but PRMT4A and PRMT4B are paired with LAM in the opposite vectors (with bait and prey switched compared to the experimental pairings). The DBR1 pairing with PRMT4B showed no growth on selective media and PRMT4A and PRMT4B are very similar, so it seems likely that PRMT4A would show no auto-activation in the prey vector either; however, without a control pairing of LAM with PRMT4A in the prey vector we cannot have full confidence in this direct DBR1-PRMT4A interaction.

***sic-3* and *prmt4a;4b* have overlapping changes in gene expression and splicing**

Previous work has shown global splicing changes in *prmt4a;4b* (Hernando et al., 2015), and we showed this for *sic-3* in Chapter 1. Here we chose to compare gene expression and splicing between the *sic-3* and *prmt4a;4b* mutants in order to characterize their similarities. Alternative splicing from both experiments was analyzed using ASpli (Mancini et al., 2019); we ran ASpli for the *sic-3* RNA-Seq data, and compared this to published ASpli output for *prmt4a;4b* RNA-Seq (Hernando et al., 2015). In our RNA-Seq experiment, *sic-3* had 3362 differentially expressed (DE) genes compared to WT, and in a previously published experiment *prmt4a;4b* had 5504 (both using <0.05 FDR and 0.58 log2 fold-change cutoffs). Due to differences in plant growth conditions between these RNA-Seq experiments, we did not compare the absolute number of differentially expressed genes or bins with differentially utilized splice sites (*sic-3* was entrained in LD 22°C for days 4-8, transferred to 16°C

and collected after 8h, and RNA was pooled for time points across 24h; *prmt4a;4b* grown in 22°C continuous light for 10 days then collected at one time point).

sic-3 and *prmt4a;4b* showed similar overlap for genes with either negative or positive DE, and for opposite-signed DE; all overlap totaled 22% of *sic-3* and 13% of *prmt4a;4b* DE genes (Figure 2A). In terms of alternative splicing, *sic-3* had only 201 genes with differentially utilized (DU, equivalent of DE) introns, exons, and 5' and 3' splice sites, while *prmt4a;4b* had 1066 (Figure 2B); 40 DU genes overlapped between the two mutants, which totaled 20% of *sic-3* DU genes and 4% of *prmt4a;4b* DU genes (Figure 2B). Conversely, *sic-3* had 1942 constitutively spliced genes with differential usage compared to WT, and *prmt4a;4b* had only 998 (Figure 2C). The overlap between mutants for constitutively spliced genes with DU was 172 genes (Figure 2C), comprising 9% and 17% of *sic-3* and *prmt4a;4b* totals, respectively. Additionally, SIC expression was decreased in the *prmt4a;4b* mutant.

Finally, we compared the different types of alternative splicing (AS) events with DU in *sic-3* and *prmt4a;4b*. Alternative splicing events were characterized as either intron retention (IR), exon skipping (ES), 3' or 5' alternative splice site (3'Alt, 5'Alt), or multiple, meaning multiple types of alternative splicing happened at that splice junction. The type of AS most affected in *sic-3* was ES and was 3'Alt in *prmt4a;4b* (Table 1). Between these two mutants the percentages of IR and 5'Alt affected were very similar, whereas the percentages of ES and 3'Alt were different (Table 1). Within each type of alternative splicing, we also calculated the percentage of events that were upregulated in the mutant background (positive DU value). When percentage upregulation is close to 50%, this means that this type of AS was equally upregulated and downregulated. Most types of AS were only slightly upregulated in both mutants (Table 1). The only exceptions were that almost all ES events affected in *sic-3* were upregulated, and the majority of IR events affected in *prmt4a;4b* were downregulated (Table 1). Overall, *sic-3* had more AS event upregulation than *prmt4a;4b* in three categories, except that 3'Alt was higher in *prmt4a;4b* (Table 1). This suggests that while both SIC and PRMT4A;4B affect alternative splicing, they have different effects. SIC mostly upregulates exon skipping, and PRMT4A;4B slightly upregulates 3' alternative splice site usage and mostly downregulates intron retention.

sic-3 and *prmt4a;4b* may have different changes in light response and redox reactions

As shown in Chapter 1, the top 10 overrepresented Gene Ontology (GO) terms for biological process in *sic-3* included five GO terms involving catabolism – three for amino acid catabolism and two for acid catabolism, two for light response, and two for starvation response. This indicates that the processes most affected in *sic-3* are amino acid catabolism (probably part of starvation response), and light response. In *prmt4a;4b*, the top 10 overrepresented GO terms for biological process included photosynthesis, five terms for various protein component biogenesis and metabolism, translation, ribosome and RNP complex biogenesis, and nitrogen compound biosynthesis (Supplemental Figure 1).

We also performed GO term representation analysis for biological process with lists of genes regulated similarly or differently between *sic-3* and *prmt4a;4b*

(both with positive or negative log fold-change values, or with opposite signed log fold-change). There were no significantly overrepresented GO terms for genes regulated similarly in *sic-3* and *prmt4a;4b*. For genes regulated differently, the overrepresented parent GO terms were response to light stimulus, oxidation-reduction reactions, and response to external stimulus (Supplemental Figure 2). This indicates that SIC and PRMT4A;4B may both regulate response to light and oxidation-reduction reactions but in opposing manners.

prmt4a;4b does not have perturbed circadian clock rhythms like *sic* mutants

In order to investigate possible functions of a SIC-PRMT4A interaction, we compared overlap between *sic* and *prmt4a;4b* mutant phenotypes. Previous work showed no *prmt4a;4b* clock phenotype under LL 22°C free run after 16h light/8h dark entrainment (Hernando et al., 2015), but did not investigate the temperature conditions where *sic* mutants exhibit the most pronounced circadian clock phenotypes. In order to compare circadian clock activity in *sic* and *prmt4a;4b* mutants, we measured circadian clock-regulated bioluminescence from the *PRR7:LUC* reporter construct under multiple entrainment and free run conditions in WT, *sic-1*, *sic-3*, and *prmt4a;4b* mutants.

Under temperature entrainment conditions, constant light and cycles of 12h 22°C-12h 16°C, WT and *prmt4a;4b* plants consistently had a 24-25h period with relative amplitude error (RAE) values of 0.3-0.5 (Figure 3A). RAE is calculated during curve fitting of bioluminescence data to estimate period length and values <0.6 are considered rhythmic. In contrast, *sic-1* and *sic-3* mutants had more arrhythmic individuals (RAE >0.6), higher 0.4-0.6 RAE for rhythmic individuals, and slightly shorter 22-24h periods (Figure 3A). Both *sic-1* and *sic-3* had significantly shorter average circadian periods than WT, and significantly higher average RAE values, while *prmt4a;4b* was not significantly different from WT (Figure 3B, 3C).

Under constant light and low temperature 16°C free running conditions after entrainment in constant light and cycles of 12h 22°C/12h 16°C, *sic* clock phenotypes were more apparent (Figure 4A). WT and *prmt4a;4b* individuals still tightly clustered around 24-25h period and 0.2-0.5 RAE, but with a few more arrhythmic individuals than in LL 22°C-16°C. In contrast, more than half of the *sic-1* and *sic-3* individuals were arrhythmic, and their periods ranged from WT 24-25h period all the way up to 34h (Figure 4A). Only the *sic-1* average period was significantly longer than WT, but the *sic-3* average was higher than WT, and *prmt4a;4b* was the same as WT (Figure 4B). Both *sic-1* and *sic-3* average RAE were higher than WT, while *prmt4a;4b* average RAE was actually lower (Figure 4C).

Under constant light and cold 16°C free running conditions after entrainment under constant 22°C and cycles of 12h light/12h dark, *sic* clock phenotypes were very severe (Figure 5A). WT and *prmt4a;4b* clustered less tightly after LD 22°C entrainment than after LL 22°C/16°C entrainment, with a wider range of both period and RAE values; however, *sic-1* and *sic-3* still ranged far more widely. Both *sic-1* and *sic-3* had significantly longer average periods than WT, and *prmt4a;4b* average period was not significantly different than WT (Figure 5B). Only *sic-3* had a significantly higher average RAE than WT under these conditions (Figure 5C).

Under all three conditions tested, *prmt4a;4b* had a similar circadian clock period and RAE compared to WT. As expected, *sic-3* and *sic-1* mutants had long periods and higher RAE values consistent with poor rhythm quality. Because SIC affects the circadian clock but PRMT4A and PRMT4B do not, the SIC-PRMT4A interaction is not involved in SIC's contribution to proper circadian clock function.

Salt sensitivity in *sic-1, sic-3, prmt4ab* mutants

A *prmt4a;4b* mutant was previously described as having elevated sensitivity to salt inhibition of root growth (Hernando et al., 2015). We chose to measure salt tolerance in *sic* mutants to evaluate whether the SIC-PRMT4A interaction contributes to salt tolerance mechanisms. WT, *sic-1*, *sic-3*, and *prmt4a;4b* seedlings were germinated on MS agar plates, then robust 3d-old individuals were transferred to fresh agar plates with either normal MS or MS supplemented with 100mM or 160mM NaCl. Root growth was then measured after 7 days of additional growth. WT showed robust root growth on MS without NaCl, with a large decrease in root growth in the presence of 100mM NaCl and almost no root growth with 160mM NaCl (Figure 6A). Both *sic-1* and *sic-3* mutants had significantly less root growth than WT on normal MS, and no growth on 160mM NaCl. The *sic* mutants differed at 100mM NaCl, where *sic-3* root growth was not significantly different than WT and *sic-1* had significantly less root growth than WT (Figure 6A). In contrast, *prmt4a;4b* mutant root growth with MS was not significantly different from WT, but was significantly greater than WT on both 100mM and 160mM NaCl (Figure 6A). Overall, *sic* mutants had significantly less root growth than WT both with and without salt stress, indicating that root growth in *sic* mutants is generally impaired. The *prmt4a;4b* mutants had the opposite phenotype, with similar root growth to WT in the absence of salt stress but greater root growth compared to WT under salt stress, indicating that *prmt4a;4b* mutants have higher salt tolerance than WT.

In order to see whether *sic* mutants had impaired salt tolerance in addition to generally impaired root growth, we divided mean root growth on 100mM or 160mM NaCl by mean root growth on MS within each genotype (Figure 6B). This root growth ratio for WT exposed to 100mM NaCl was 0.46, indicating that root growth was halved under this degree of salt stress; for 160mM this ratio was essentially 0 (Figure 6B). The root growth ratio for *sic-1* at 100mM was also essentially 0, indicating that *sic-1* root growth is much more impaired by salt stress than WT (Figure 6B). Root growth in the *sic-3* mutant was somewhat less impaired than WT at 100mM NaCl, but similarly impaired or more so at 160mM NaCl (Figure 6B). Finally, *prmt4a;4b* root growth was almost unimpaired at 100mM NaCl, and much less impaired than WT or either *sic* mutant at 160mM NaCl (Figure 6B). These experiments show that *sic* and *prmt4a;4b* mutants have different effects on root growth without stress and in response to salt stress. Therefore, SIC and PRMT4A make different contributions to salt stress tolerance mechanisms controlling root growth.

DISCUSSION

As shown in Chapter 1, a previous IP-MS experiment for SIC yielded PRMT4A and PRMT4B as SIC-interacting proteins (Karampelias et al., 2016). We used Y2H to

show that SIC and PRMT4A can interact directly, while SIC and PRMT4B cannot; PRMT4A and DBR1 may also interact directly. Previous work in our lab used bi-molecular fluorescence complementation (BiFC) with transient infiltration of *Nicotiana benthamiana* to show strong SIC-PRMT4A and PRMT4A-PRMT4B interactions, both on the plasma membrane and in the cytosol, and a similar but weaker SIC-PRMT4B interaction (Marshall, 2017). The differences between these Y2H and BiFC results suggest that Arabidopsis SIC and PRMT4B are able to interact directly in *N. benthamiana* but not in yeast, or that the interaction is weak and was more easily visible by BiFC than Y2H. Since PRMT4A and PRMT4B interact in BiFC, and SIC and PRMT4A but not PRMT4B interact in Y2H, it is likely that PRMT4B appeared as a SIC-interacting protein in IP-MS because of an indirect interaction through PRMT4A. Since we have now shown direct SIC-PRMT4A and SIC-DBR1 interactions (Chapter 1), and possibly DBR1-PRMT4A, it is possible that all three proteins form a complex together or else have separate binary interactions. As discussed in Chapter 1, few DBR1 protein interaction partners have been identified, so discovery of this possible DBR1-PRMT4A interaction adds to our knowledge of DBR1 protein activity.

The next question we asked was what possible functions this SIC-PRMT4A interaction might serve. Toward this end we investigated possible shared phenotypes of *sic* and *prmt4a;4b* by measuring circadian clock rhythms, differential gene expression and mRNA splicing, and salt tolerance. When we investigated circadian rhythms and salt tolerance in the two mutants, we did not see similarities. As shown previously, *sic* mutants had a longer period and higher RAE than WT in cool temperature entrainment and free run (Marshall et al., 2016), but we found no similar phenotype in *prmt4a;4b* under thermocycles or in cool temperature free run. Rather, the *prmt4a;4b* period and RAE were not significantly different from WT under any condition except for lower RAE (higher quality rhythms) in LL16 free run after LL 22C/16C entrainment, which does not indicate any problem with the circadian clock. Salt tolerance also differed between *sic* and *prmt4a;4b*. *sic* had decreased root growth compared to WT, as well as decreased salt tolerance, while *prmt4a;4b* had more root growth under salt stress than WT. These results suggest that SIC and PRMT4A do not act together to affect either circadian rhythms or salt tolerance. Interestingly, this increase in *prmt4a;4b* root growth on MS containing 100mM and 160mM NaCl is the opposite of previously published data showing a root growth decrease; we also saw much less WT growth on 160mM than previously shown (Hernando et al., 2015).

In order to further investigate possible shared phenotypes between *sic* and *prmt4a;4b*, we used previously published data (Marshall, 2017) to measure differential gene expression and mRNA splicing, and overrepresented biological processes. Previous research showed that *prmt4a;4b* has global changes in splicing, and we showed in Chapter 1 that *sic-3* does as well, in addition to previously published increased intron sequences in *sic-1* (Zhan et al., 2012). Here we further compared differential expression, alternative splicing, and constitutive splicing between the two mutants. In all three categories, *sic-3* and *prmt4a;4b* had up to 20% overlap of genes affected. SIC expression was lower in *prmt4a;4b* than in WT, suggesting PRMT4A;4B may positively regulate SIC expression. Changes in

alternative splicing suggested that while both SIC and PRMT4A;4B affect alternative splicing, SIC mostly upregulates exon skipping, and PRMT4A;4B slightly upregulates 3' alternative splice site choice and mostly downregulates intron retention. Therefore, SIC and PRMT4A;4B may not act together in alternative splicing.

In a list of genes with opposite signs of $\log_2$ fold-change for differential expression in *sic-3* and *prmt4a;4b*, GO terms for the biological processes light response and oxidation-reduction reactions were overrepresented. This indicated that these two processes may be oppositely regulated between *sic-3* and *prmt4a;4b*. We know that in *sic-3* circadian rhythms are perturbed and this can involve response to light. Previous research has shown that PRMT4B interacts with Paraquat Tolerance 3 (PQT3), and thereby positively regulates oxidative stress (Luo et al., 2016).

None of the experiments presented in this chapter provided clear answers about the possible role of interaction between SIC and PRMT4A. The *sic-3* and *prmt4a;4b* mutants had different phenotypes for circadian clock rhythms, root growth, salt tolerance, mRNA splicing, and GO representation analysis showed no significant similarly affected pathways between the two mutants. This brings us back to the possibility that SIC and PRMT4A directly interact only because SIC is a substrate of PRMT4A;4B. Although SIC does not contain a GAR motif, PRMTs do not solely methylate GAR-containing proteins. Future experiments could test whether PRMT4A;4B can methylate SIC in-vitro, or whether SIC is methylated in-vivo.

MATERIALS AND METHODS

All experiments were performed as described in Chapter 1, with the exception of those below.

RNA-Seq analysis using ASpli

We used ASpli 2019 to analyze *sic-3* RNA-Seq, and Hernando et al. (2015) used an older version. The main difference is in annotation of alternative splicing bins with multiple types of events as "*" in ASpli 2019 vs. "Multiple" in older versions. In order to make these two sets of results comparable, we changed AS events annotated with "*" by ASpli 2019 to "Multiple". This is also why *sic-3* alternative splicing data differs between Chapters 1 and 2: Chapter 1 *sic-3* and *dbr1* results have the unaltered ASpli 2019 "*" annotation.

Circadian rhythm analysis by bioluminescence assay

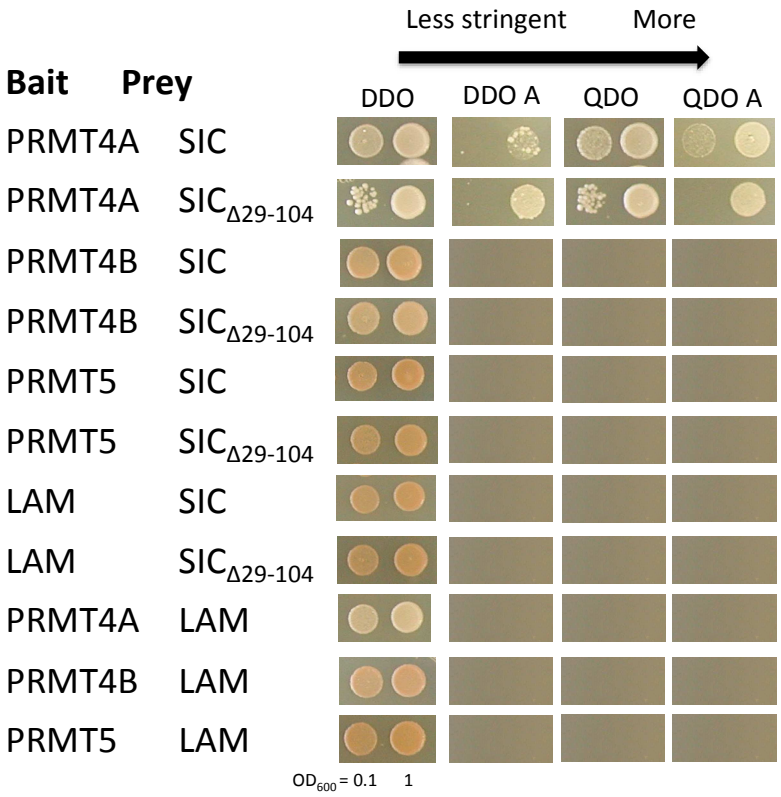
As previously described (Marshall et al., 2016).

Salt tolerance assay for root growth

Seeds were gas sterilized, plated, stratified, and germinated as described in Chapter 1. The most robust individuals for each genotype were transferred to new MS agar plates with either 0mM, 100mM, or 160mM NaCl, the end of each root was marked, and plates were grown upright for 7d. After 7d on new plates (10d total), plates were photographed and root length was analyzed with ImageJ (Schneider et al., 2012).

Chapter 2 Figures

A


Less stringent
More
Bait
Prey
DDO
DDO A
QDO
QDO A
PRMT4A SIC
PRMT4A SIC$_{\Delta 29\text{-}104}$
PRMT4B SIC
PRMT4B SIC$_{\Delta 29\text{-}104}$
PRMT5 SIC
PRMT5 SIC$_{\Delta 29\text{-}104}$
LAM SIC
LAM SIC$_{\Delta 29\text{-}104}$
PRMT4A LAM
PRMT4B LAM
PRMT5 LAM
OD$_{600}$ = 0.1 1

B

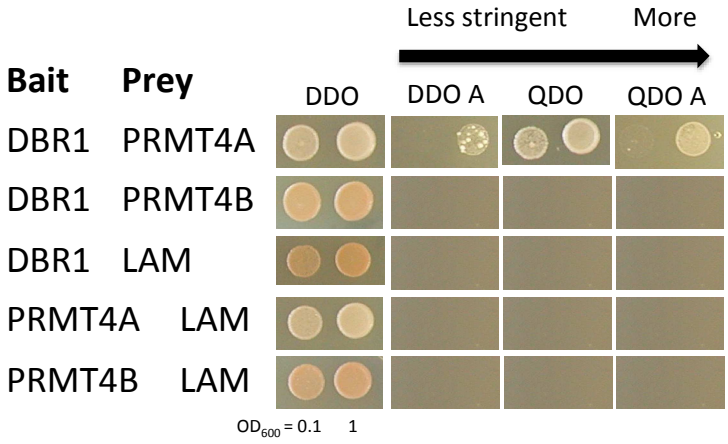

Less stringent
More
Bait
Prey
DDO
DDO A
QDO
QDO A
DBR1 PRMT4A
DBR1 PRMT4B
DBR1 LAM
PRMT4A LAM
PRMT4B LAM
OD$_{600}$ = 0.1 1

Figure 1. SIC interacts with PRMT4A and PRMT4A interacts with DBR1.

Yeast two-hybrid assay for protein interaction between **A)** SIC and PRMT4A and **B)** PRMT4A and DBR1. In each picture, yeast spot on right is OD_{600} = 1, and spot on left is 0.1. Media is SD with DDO=double dropout (SD-Leu-Trp), QDO=quadruple dropout (SD-Leu-Trp-His-Ade), A=aureobasidin. Negative-control pairings are with LAM, human lamin C. Three independent yeast transformants were grown per experiment. Panel **A)** is representative of two independent experiments and panel **B)** is from one independent experiment.

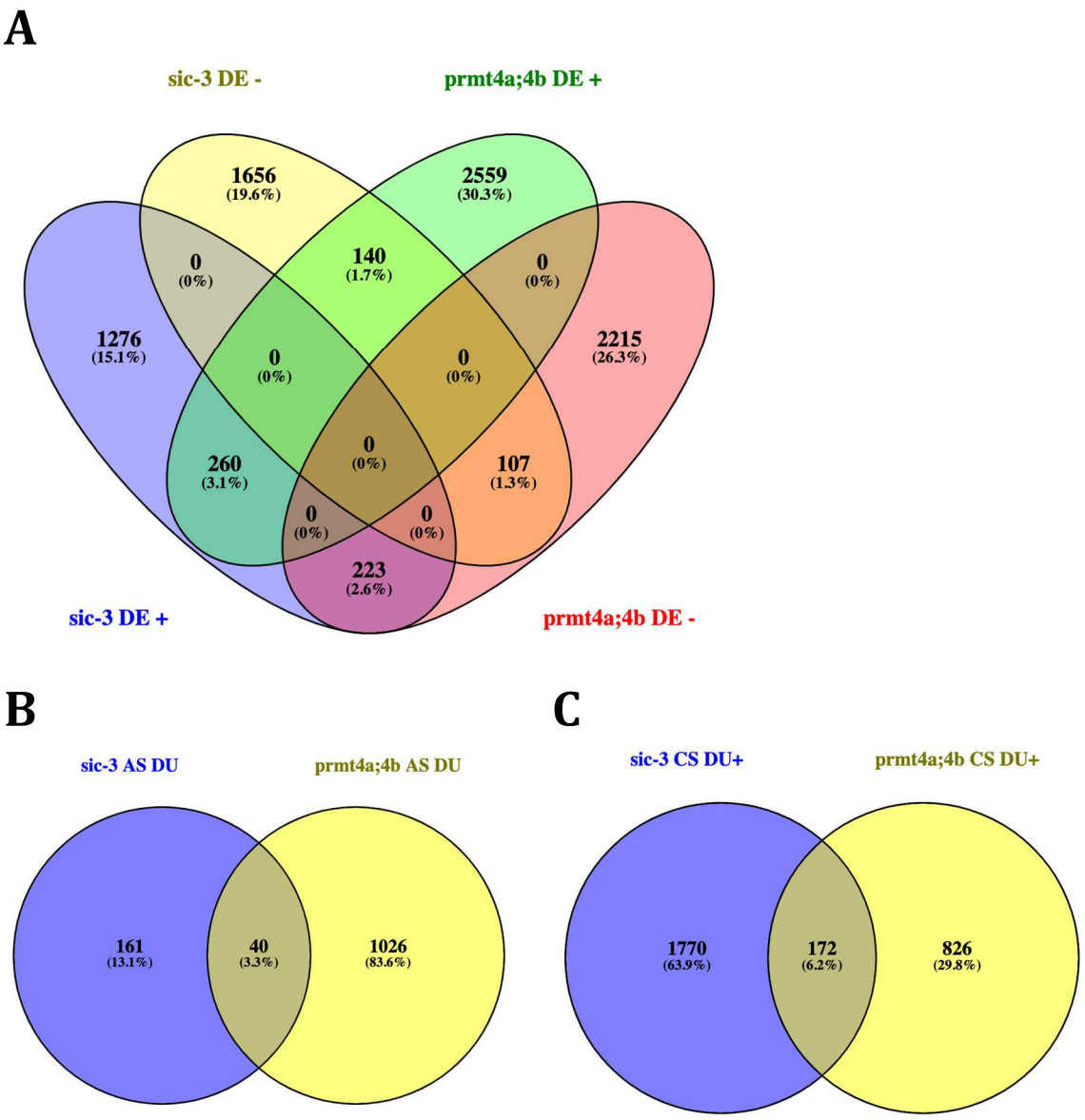


Figure 2. Differential expression and differential usage of splice variants in *sic-3* and *prmt4a;4b*.

Cutoffs were FDR<0.05 for differential expression (DE) and FDR<0.1 for differential usage (DU) of alternatively spliced (AS) transcripts, and $\log_2$FC<0.58 for each. Percentages are out of total number of genes represented in entire Venn diagram. Venn diagrams generated using Venny (Oliveros). + represents a positive $\log_2$ fold-change value and – a negative value.

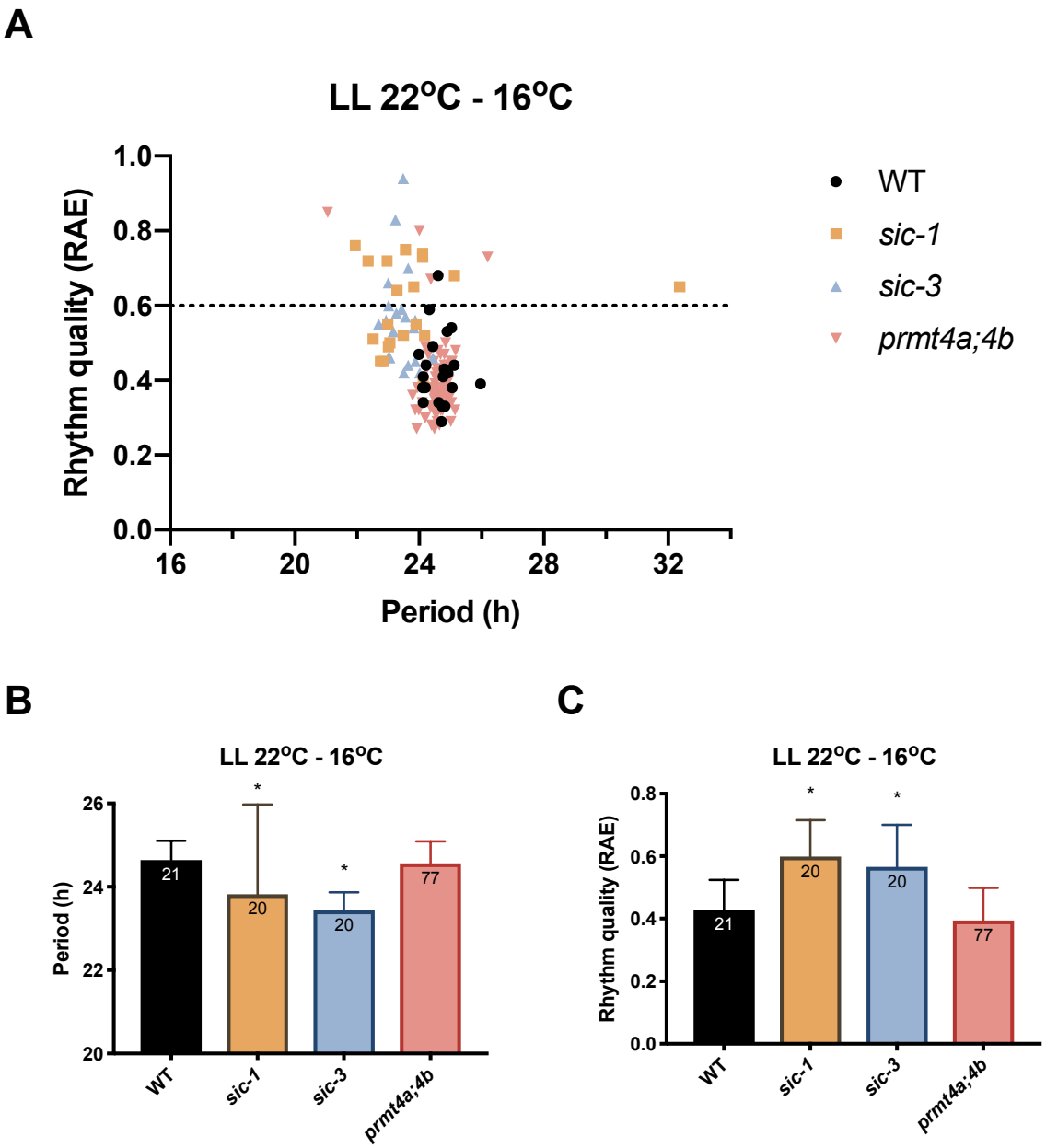


Figure 3. *prmt4a;4b* circadian period and RAE are comparable to WT under 22°C-16°C temperature entrainment.

A) Estimated circadian period plotted against quality of circadian rhythms (lower Relative Amplitude Error (RAE) indicates higher quality rhythms, rhythms with RAE <0.6 are considered rhythmic). Each point represents an individual seedling. Imaging occurred while plants were under constant light and cycles of 12h 22°C-12h 16°C beginning when seedlings were 3 days-old. Data are from one experiment.

B) Average estimated period and **C)** average RAE for plants shown in **A)**. Error bars are standard deviation. Bars are labeled at the top with number of individual plants. Asterisks indicate significant difference from WT by one-way ANOVA with Dunnett's multiple comparisons test ($p<0.05$).

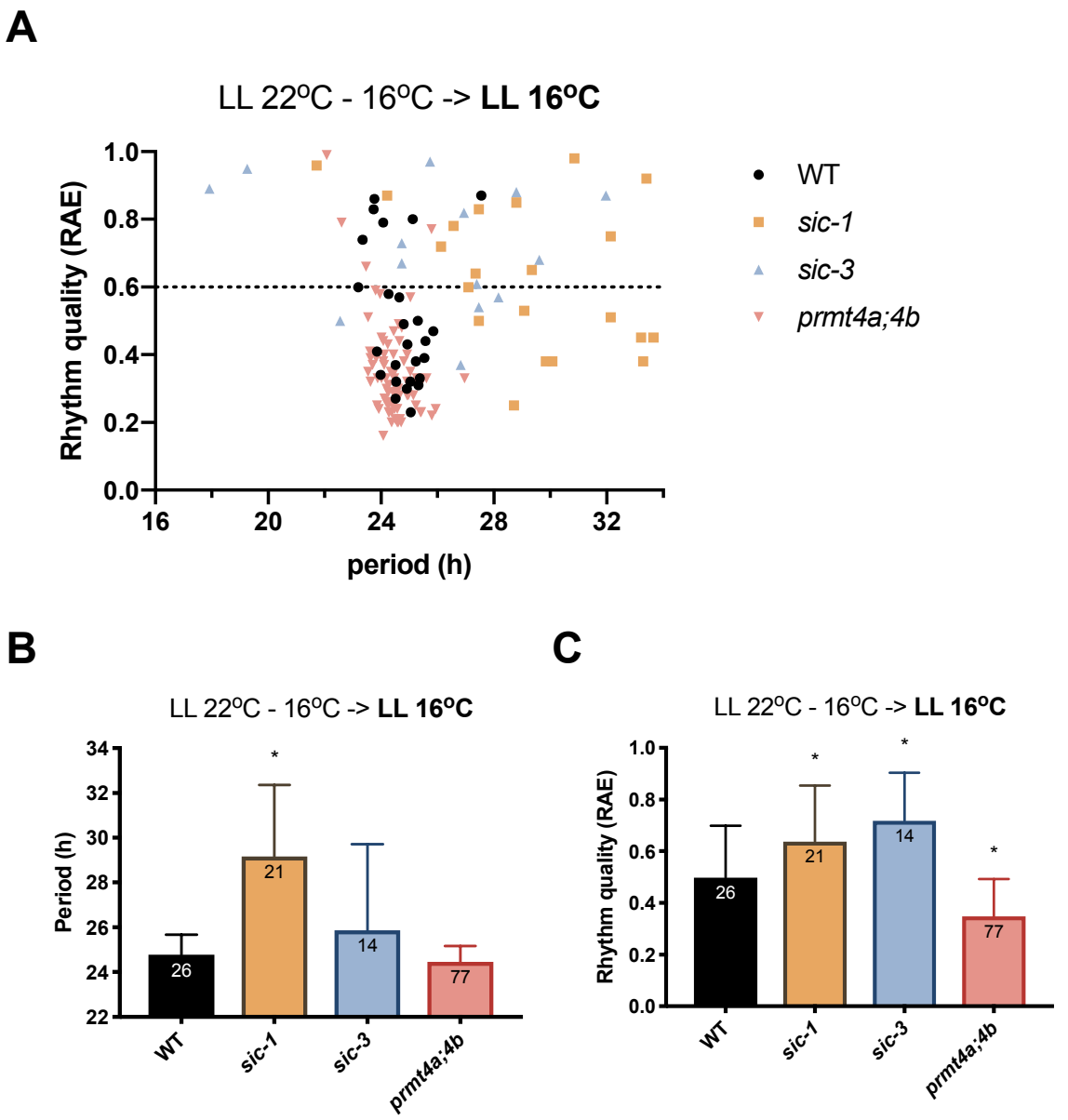


Figure 4. *prmt4a;4b* circadian period matches WT in cool temperature free run after 22°C-16°C thermocycle entrainment, but RAE is lower.

A) Estimated circadian period plotted against RAE (lower RAE indicates higher quality rhythms, rhythms with RAE <0.6 are considered rhythmic). Each point represents an individual seedling. Plants under constant light at 16C after entrainment for 5 days in constant light and cycles of 12h 22C-12h 16C. Data are from one experiment.

B) Estimated period and **C)** RAE of plants shown in **A)**. Bar is the mean and error bars are SD. Bars are labeled at the top with number of individual plants. Asterisks indicate significant difference from WT by one-way ANOVA with Dunnett's multiple comparisons test ($p<0.05$).

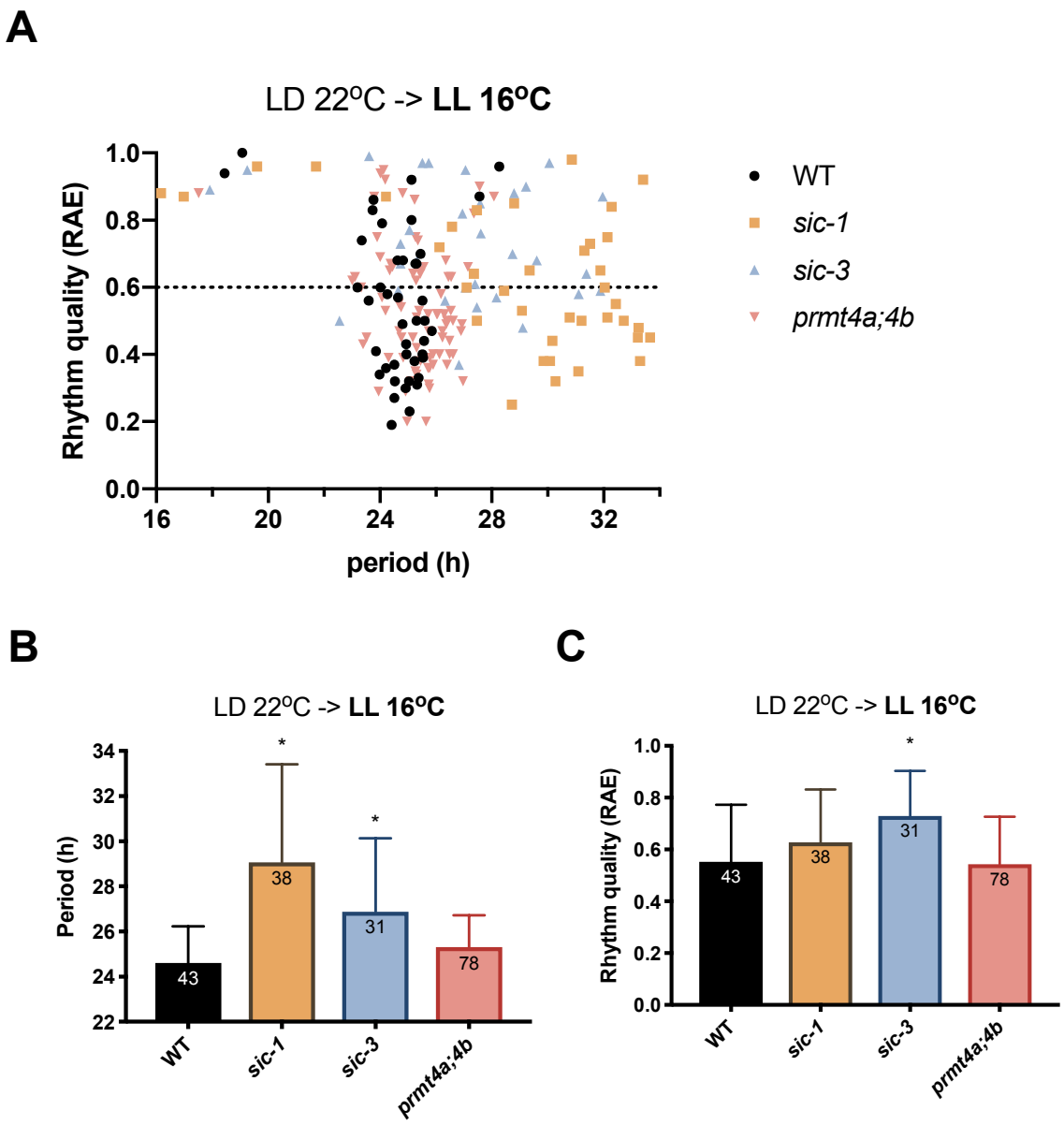


Figure 5. *prmt4a;4b* circadian period and RAE match WT under cool temperature free run after light-dark entrainment.

A) Estimated circadian period plotted against quality of circadian rhythms (lower RAE indicates higher quality rhythms, rhythms with RAE over 0.6 are considered arrhythmic). Each point represents an individual seedling. Plants in constant light at 16C after entrainment in constant 22C and cycles of 12h light/12h dark. Data are from one independent experiment.

B) Average estimated period and **C)** average RAE for plants shown in **A)**. Error bars are standard deviation. Bars are labeled at the top with number of individual plants. Asterisks indicate significant difference from WT by one-way ANOVA with Dunnett's multiple comparisons test ($p<0.05$).

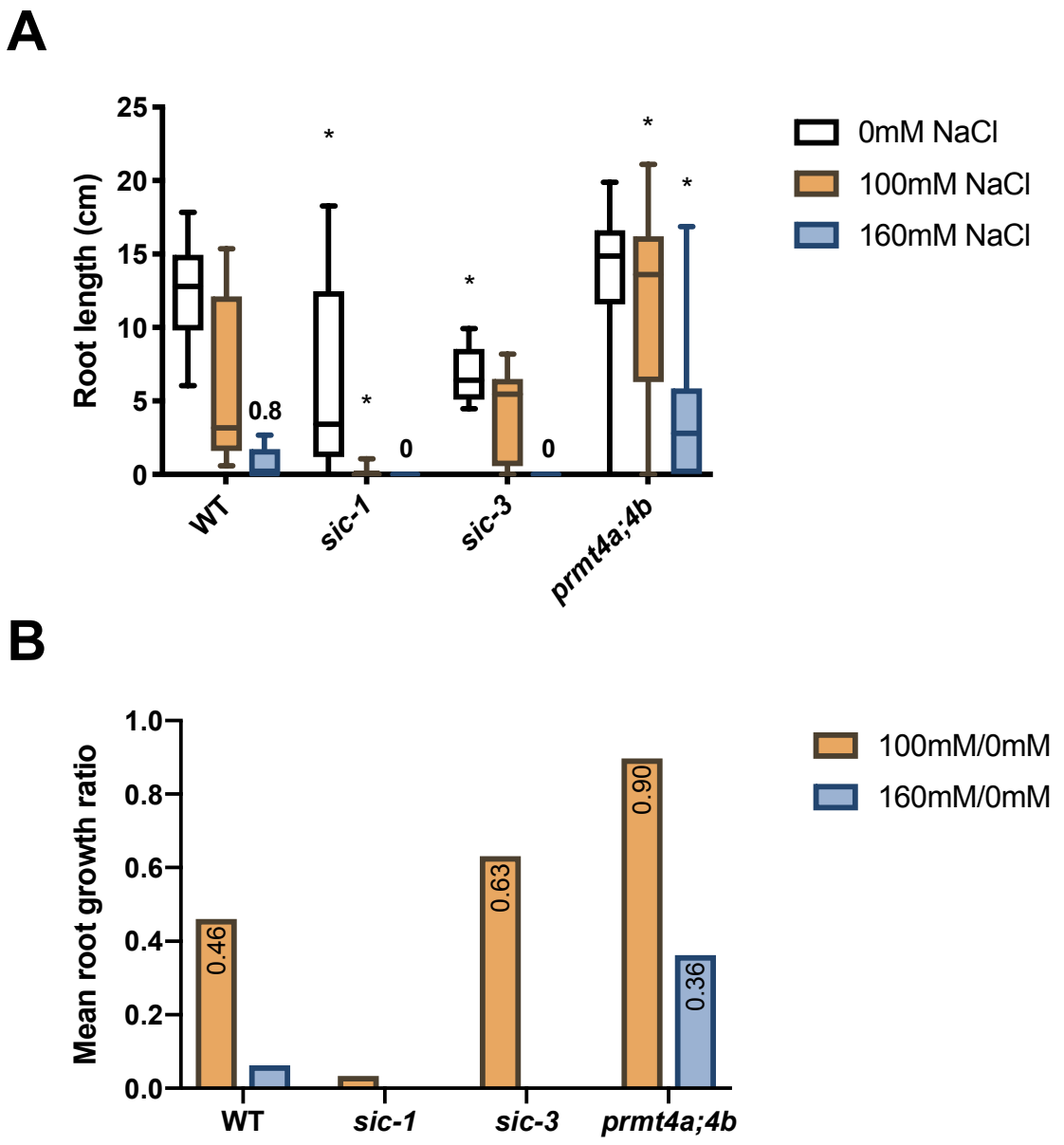


Figure 6. *sic-1* and *sic-3* have lower salt tolerance than WT, while *prmt4a;4b* is more tolerant.

A) Root length of 10d-old plants after 7d on MS plates (plants were 3d-old when plated on salt) with either 0, 100, or 160mM NaCl. N=10 for WT, *sic-1*, *sic-3*, 45 for *prmt4a;4b*. Asterisks indicate significant differences from WT treated with same salt concentration by 2-way ANOVA with Dunnett's multiple comparisons test ($p<0.05$). Box plot shows average with line, box for SD, error bars for min max. Small numbers above some bars list value of mean for that sample.

B) Root growth ratio is average root length for each genotype on 0mM NaCl divided by average root length for the same genotype on 100mM or 160mM NaCl.

TABLES

Table 1. IR, ES, and 3'Alt alternative splicing events are affected differently in *sic-3* and *prmt4a;4b* mutants, and 5'Alt is affected similarly.

Alternative splicing w/ DU		
	sic-3	*prmt4a;4b*
Total Events	225	1262
IR	24%	24%
w/ DU +	51%	27%
ES	34%	8%
w/ DU +	92%	57%
3'Alt	7%	36%
w/ DU +	56%	65%
5'Alt	17%	15%
w/ DU +	63%	55%
Multiple	18%	18%

Differential usage of splice variants between *sic-3* and WT was identified with ASpli (Mancini et al., 2019) and these results were compared to previously-published analysis of alternative splicing in *prmt4a;4b* to WT (Hernando et al., 2015). At the top is the total number of alternative splicing (AS) events annotated by ASpli analysis that have differential usage (DU) of splice sites in the mutant compared to WT. Below the total is the percentage of total events annotated as each type: IR (intron retention), ES (exon skipping), 3'Alt (alternative 3' splice site usage), 5'Alt (alternative 5' splice site usage), or Multiple (splice junction had multiple types of alternative splicing events). Below the percentage of events of a given type is the percentage of these that had higher differential usage (as opposed to lower) in the mutant compared to WT.

SUPPLEMENTAL FIGURES

Supplemental Figure 1. GO enrichment analysis for biological process using DE genes in *prmt4a;4b.*

	Arabidopsis thaliana (REF)	upload 1 (Hierarchy) NEW!					
GO biological process complete	#	#	expected	▽ Fold Enrichment	+/-	raw P value	FDR
photosynthesis	189	67	35.88	1.87	+	3.27E-05	7.75E-03
generation of precursor metabolites and energy	352	116	66.82	1.74	+	8.16E-07	2.68E-04
peptide biosynthetic process	557	180	105.74	1.70	+	2.08E-09	2.05E-06
translation	552	178	104.79	1.70	+	3.28E-09	2.77E-06
amide biosynthetic process	615	196	116.75	1.68	+	1.01E-09	2.00E-06
peptide metabolic process	633	192	120.17	1.60	+	3.45E-08	2.04E-05
ribosome biogenesis	381	110	72.33	1.52	+	1.83E-04	3.39E-02
ribonucleoprotein complex biogenesis	463	133	87.90	1.51	+	4.50E-05	1.02E-02
cellular amide metabolic process	755	215	143.33	1.50	+	2.81E-07	1.04E-04
cellular nitrogen compound biosynthetic process	1211	344	229.90	1.50	+	7.23E-11	2.14E-07
organonitrogen compound biosynthetic process	1320	372	250.59	1.48	+	2.57E-11	1.52E-07
protein-containing complex assembly	486	133	92.26	1.44	+	2.48E-04	4.19E-02

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20190429)

Annotation Version and Release Date: GO Ontology database Released 2019-02-02

Analyzed List: upload_1 (Arabidopsis thaliana) Change

Reference List: Arabidopsis thaliana (all genes in database) Change

Annotation Data Set: GO biological process complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	27581 out of 27581	5199 out of 5236
Unmapped IDs:	0	305
Multiple mapping information:	0	144

Top 10 significantly overrepresented genes by fold enrichment shown ($p<0.05$). Analysis and results summaries.

Supplemental Figure 2. GO enrichment analysis for biological process using DE genes with opposite sign logFC in *sic-3* and *prmt4a;4b.*

	Arabidopsis thaliana (REF)	upload_1 (▽ Hierarchy NEW!)					
GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
response to light stimulus	739	25	9.27	2.70	+	1.18E-05	1.74E-02
↳response to radiation	760	25	9.53	2.62	+	1.86E-05	2.19E-02
↳response to abiotic stimulus	2099	52	26.33	1.97	+	3.50E-06	1.04E-02
↳response to stimulus	5729	116	71.87	1.61	+	5.33E-08	3.15E-04
oxidation-reduction process	942	30	11.82	2.54	+	4.95E-06	9.75E-03
response to external stimulus	1387	37	17.40	2.13	+	2.13E-05	2.10E-02

Analysis Summary: Please report in publication

Analysis Type: PANTHER Overrepresentation Test (Released 20190429)

Annotation Version and Release Date: GO Ontology database Released 2019-02-02

Analyzed List: upload_1 (Arabidopsis thaliana) Change

Reference List: Arabidopsis thaliana (all genes in database) Change

Annotation Data Set: GO biological process complete

Test Type: Fisher's Exact Binomial

Correction: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing No correction

Results

	Reference list	upload_1
Uniquely Mapped IDS:	27581 out of 27581	344 out of 346
Unmapped IDs:	0	19
Multiple mapping information:	0	2

All significantly overrepresented genes shown ($p<0.05$). Analysis and results summaries.

Chapter 3: Maize *lhy* affects the circadian clock, plant growth, and hybrid vigor

ABSTRACT

Historically, circadian clock research in *Arabidopsis thaliana* has been more focused on the activity of *CCA1* than on that of its homolog *LHY*. However, CCA1 is not present in Angiosperms outside the Brassicaceae, necessitating study of *LHY* in other Angiosperms. Research in Arabidopsis suggests a role for *CCA1* and *LHY* in metabolic aspects of hybrid vigor. Maize hybrids are widely used in agricultural production because of the positive agronomic traits conferred by hybrid vigor. This Chapter describes experiments testing the two maize *lhy* paralogs for a role in hybrid vigor. We constructed transgenic *lhy* misexpression lines in order to ask what role *lhy* plays in the maize circadian clock, whether *lhy* contributes to maize growth, what patterns of *lhy* expression characterize hybrids, and whether changes in *lhy* expression contribute to vigor. We found that *lhy* negatively regulates *toc1* in maize, and changes in *lhy* have surprisingly small effects on growth. In addition, *lhy* expression at ZT6 was most correlated with hybrid vigor, and hybridization was sufficient to change *lhy* expression and increase plant height. Overall, our results suggest that although the maize *lhy* genes affect the circadian clock, these genes do not significantly contribute to growth and hybrid vigor.

INTRODUCTION

CCA1 encodes a major transcription factor acting at the core of the *Arabidopsis thaliana* circadian clock. As a consequence it is one of the most extensively-studied genes in the plant circadian clock (Wang and Tobin, 1998; Green and Tobin, 1999; Alabadí et al., 2001; Alabadí et al., 2002; Mizoguchi et al., 2002; Gutiérrez et al., 2008; Nagel et al., 2015). Arabidopsis has a single copy of *CCA1*. Arabidopsis also has a single copy of *LHY*, which encodes a core circadian clock transcription factor homologous to *CCA1* that has partially overlapping functions with *CCA1*. The transcription factors encoded by *CCA1* and *LHY* together negatively regulate the *TOC1* gene. Despite its thorough study in Arabidopsis, *CCA1* is only present in the Brassicaceae, while *LHY* is present (in varying copy numbers) in Angiosperms all the way through the basal Angiosperm *Amborella trichopoda* (Lou et al., 2012; Bendix, 2015). Therefore, we must study the function of *LHY* outside of the Brassicaceae in order to more fully understand circadian clock mechanisms in the majority of Angiosperms.

Arabidopsis mutants with altered circadian clock function often have changes in development, growth, and metabolism. This has led to an interest in the involvement of the circadian clock in hybrid vigor. Work in Arabidopsis has shown that changes in expression of *CCA1* and *LHY* affect growth vigor in Arabidopsis allopolyploids and hybrids. *CCA1* expression was lower in hybrids than their inbred parents at midday (ZT6, 9, and 12; ZT defined as hours after entraining cue, in this case dawn), and higher in the late evening (ZT15 and 18); these changes in *CCA1* caused upregulation of *TOC1* at ZT6 and 9, and downregulation of *TOC1* at ZT15 and 18 (Ni et al., 2009). Selective knockdown of *CCA1* transcript at ZT6-ZT9 by RNA interference (RNAi) caused increased *TOC1* expression similar to that in hybrids. This change in *CCA1* expression also increased plant starch content (Ni et al., 2009). Of genes with altered expression in Arabidopsis hybrids, about 2/3 contained a CCA1-binding site or an evening element, suggesting many are targets of CCA1 (Ni et al., 2009). Therefore downregulation of *CCA1* at midday may cause growth vigor in hybrids via upregulation of metabolic genes (Figure 1A).

The potential role of the circadian clock in mechanisms contributing to hybrid vigor in Arabidopsis provides further motivation to study the circadian clock in Angiosperms outside of the Brassicaceae. The discovery and agricultural exploitation of maize hybrid vigor starting in the 1930s caused one of the largest increases in crop yield in the history of modern agriculture. Maize hybrid vigor was so revolutionary that its adoption in US agriculture was one of the fastest implementations of new technology (Manuelli and Seshadri, 2014). However, the genetic and molecular basis of hybrid vigor is still do not well understood, including hybrid vigor in maize. Since maize hybrid vigor is so agriculturally important it is an interesting and potentially high-impact plant in which to study how the circadian clock affects hybrid vigor in a non-Brassica angiosperm.

The goals of this project were to answer four questions: 1) How do the two maize *lhy* paralogs (*lhy1* and *lhy2*) affect the circadian clock in maize, 2) does *lhy* contribute to maize growth, 3) do maize hybrids have the same altered *lhy* expression pattern as Arabidopsis hybrids (*LHY* downregulation at ZT6 and 9), and 4) do changes in maize *lhy* expression contribute to vigor? To address the first two

questions, we thoroughly characterized transgenic maize *lhy* misexpression lines. We found that each maize *lhy* seems to negatively regulate *toc1* paralogs. Small, but inconsistent, changes in growth result from either increasing expression of one *lhy* gene or decreasing *lhy1* and *lhy2* expression together. To address the third question, we measured *lhy* expression at ZT6 and 10 in maize hybrid combinations with and without vigor. These experiments showed some correlation between *lhy* expression and the extent of hybrid vigor. To address the fourth question, we made hybrids with transgenic *lhy* overexpression or *lhy* knockdown lines and evaluated them for growth phenotypes and hybrid vigor. These experiments showed that hybrid vigor can change *lhy* expression and drastically increase plant height.

RESULTS

Making transgenic *lhy2*-OX, RNAi:*lhy1*;*2* lines and *lhy2* Mu lines

In order to perturb maize *lhy1* and *lhy2* expression in multiple ways, we constructed *lhy1* and *lhy2*-overexpression lines (*lhy1*- and *lhy2*-OX), and *lhy* RNA interference (RNAi:*lhy*) lines to target both *lhy* paralogs. We made RNAi:*lhy* lines with two different promoters, the maize *UBIQUITIN* (*UBQ)* promoter for robust constitutive expression (p*UBQ*:RNAi:*lhy*) and the native *lhy* promoter for increased expression only at the normal times of day (p*lhy*:RNAi:*lhy*). Transgenic lines are numbered 2 for *lhy1*-OX, 3 for *lhy2*-OX, 4 for p*UBQ*:RNAi-*lhy*, and 5 for p*lhy*:RNAi:*lhy*; within lines, numbering is of individual transformation events, eg. 313 is *lhy2*-OX individual T0 plant 13. All transgenic lines were maintained as hemizygotes to reduce possible silencing of the transgene. Unless stated otherwise, "WT" for transgenic plants is non-transgenic siblings in the same family and grow out; also abbreviated "+" for transgenic and "-" for non-transgenic. All RNAi and overexpression lines were made by technician Dominica Rohozinski, and she did the initial characterization of T0 and T1 generation of *lhy1*-OX and *lhy2*-OX lines. We also obtained *lhy1* and *lhy2* UniformMu insertion lines, a maize population with Mutator (Mu) transposable element insertions in a W22 inbred background (Portwood et al., 2018). *lhy* Mu insertion locations are shown in Figure 1B. Since we tried genotyping the *lhy1* Mu lines with little success, we thereafter characterized only the *lhy2* Mu lines.

Maize transgenic *lhy* misexpression lines have a range of effects on *lhy* expression in the T0 generation

Peak (ZT3) *lhy1* and *lhy2* expression in the set of RNAi:*lhy* T0 individuals obtained from the transformation facility varied from very low expression due to successful *lhy* knockdown to expression much higher than WT (Figure 1C). Within a given line, the effect on *lhy1* and *lhy2* expression was similar, indicating the RNAi was equally successful at targeting both paralogs (Figure 1C). Transformation with each construct produced transgenic lines with relative *lhy* expression less than 50% of WT B104, but there were twice as many p*UBQ*:RNAi:*lhy* individuals as p*lhy*:RNAi:*lhy* individuals with this level of knockdown.

We chose lines with the lowest *lhy* expression in the T0 generation, crossed them to WT B104, and measured expression again in the T1 generation to

determine if knockdown remained consistent. Lines 451 and 423 resulted from these crosses and each had the lowest and most consistent *lhy1* and *2* knockdown (Table 1), so we focused on characterizing these.

p*UBQ:*RNAi:*lhy* transgene copy number measurement using ddPCR

To determine whether transgene copy number correlated with transgene expression or resulting phenotypes, we used a novel droplet digital PCR-based method to measure transgene copy number (Collier et al., 2017). Of the main RNAi-*lhy* lines we focused on, 451 had three transgene copies and 423 had two (Figure 12). We measured multiple individual plants from each line, and the range for 423 was larger than for 451, which clustered more tightly.

***lhy* misexpression affects *toc1* expression**

In order to see whether *lhy* knockdown was sufficient to perturb the circadian clock, we measured expression of the core clock gene paralogs *toc1a* and *toc1b*, which are predicted to be negatively regulated by *lhy*. We initially measured *lhy* and *toc1* expression over 24h in plants grown in the field under normal light and temperature cycles. In order to analyze changes in circadian gene expression, we considered the phase and amplitude of rhythms – phase is the number of hours in a full cycle of expression, or the timing of a specific point of the oscillation, and amplitude is the distance between the midpoint of the circadian oscillation to a peak or trough of expression.

p*UBQ*:RNAi:*lhy* line 451, one of two strongest lines when tested at ZT3 in the T0 generation, had about 25% of WT *lhy1* and *lhy2* expression; this knockdown persisted through ZT0-9, and the phase was unchanged (Figure 2A). In this same line, expression of *toc1a* and *toc1b* in transgenic plants had a 1-3h phase advance compared to WT (Figure 2B). As a negative control, we also included p*lhy*:RNAi-*lhy* line 593, which had very low *lhy* expression in the T0 generation which rose to non-transgenic 593 levels in the T1 generation (Table 1); accordingly, this line showed no change in *lhy* or *toc1* expression throughout the 24h time course (Figure 2C,D).

We also measured gene expression in knock down and overexpression lines under free-running conditions (free run occurs under constant conditions in the absence of entraining cues), in order to study endogenous circadian rhythms and better see phase changes. *lhy2*-OX line 313 had the highest *lhy* expression of all *lhy*-OX lines. In this line, *lhy2* expression was arrhythmic and *lhy1*, *toc1*, and *toc1b* all had decreased amplitude and delayed phase of expression (Figure 3). The amplitude of *gi1* expression was not decreased in 313, but the phase was slightly delayed (Figure 3). p*UBQ*:RNAi:*lhy* line 423 had consistent *lhy1* and *lhy2* knockdown to 20-30% of WT expression (Figure 4). This *lhy* knockdown had no effect on the magnitude of *toc1a*, *toc1b*, and *gi1* expression, but slightly advanced the phase of all three (Figure 4). These results indicate that *lhy* plays a role in regulation of *toc1a* and *toc1b* expression in maize.

RNAi:*lhy* knockdown is correlated with *toc1* expression in multiple lines

According to the work above *lhy* expression in WT peaks around ZT3 (Figure 2A,C) and *toc1* expression peaks around ZT8-12 (Figure 2B,D). When *lhy* expression

decreased in p*UBQ*:RNAi:*lhy* 451, the *toc1* peak advanced toward ZT8 (Figure 2B). If peak *lhy* expression affects peak *toc1* expression, we would expect *lhy* ZT3 expression to predict *toc1* ZT6 expression better than *toc1* ZT3 expression, because ZT6 is closer to the *toc1* peak. In order to test this, we plotted *lhy1* and *lhy2* expression at ZT3 and ZT6 against *toc1a* and *toc1b* expression at these same time points and performed a regression analysis on data from 14 RNAi:*lhy* lines. If *lhy* and *toc1* are correlated, then values would fit a line and regression analysis would show a significant non-zero slope.

Three of the pairings of genes and timepoints for which we performed regression analysis showed significant correlations, and as predicted the two strongest significant correlations ($R2 = 0.84$ and 0.68, $p<0.05$) were *lhy1* and *lhy2* ZT3 vs. *toc1a* ZT6 (Figure 5A,C). Both of these correlations were negative, meaning higher peak *lhy* expression was correlated with lower peak *toc1* expression, as we would expect based on known negative regulation of *toc1* by *lhy*. The weakest significant correlation ($R2=0.65$, $p<0.05$) was *lhy1* ZT6 vs. *toc1a* ZT3; this was a positive correlation, which is the opposite of expected. There was no significant correlation between *lhy2* at ZT6 and any *toc1* expression (Figure 5D).

RNAi:*lhy* and *lhy2*-OX have weak and inconsistent growth phenotypes

Above we showed that some RNAi:*lhy* and *lhy*-OX misexpression lines have consistent effects on *lhy* expression, and that these in turn change *toc1* expression amplitude and phase. Next we wanted to see whether this perturbation of circadian clock gene expression affected circadian clock output pathways involved in plant growth. We measured plant growth by final plant height, the length of each internode on the mature plant, leaf chlorophyll content, and average seed weight. Overall, a few of these measures were significantly different between transgenic plants and WT for the two strongest RNAi:*lhy* and *lhy2*-OX lines.

We measured total chlorophyll in *lhy*-OX lines 3101 (Figure 6A) and 3149 (Figure 6B) across a developmental progression, and the average for transgenics was lower than WT at most time points but never significantly different. For chlorophyll A, B, and total chlorophyll in p*UBQ*:RNAi:*lhy* lines 423, 451, and p*lhy*:RNAi:*lhy* 571, the average for transgenic plants was slightly higher than for non-transgenic plants, and significantly higher for total chlorophyll in line 423 (Figure 6C). In line 593, which did not exhibit *lhy* knockdown beyond the T0 generation, average chlorophyll was slightly lower for transgenic plants than for non-transgenics (Figure 6C). For 423 plants in another growout, average total chlorophyll was slightly higher than WT, but the difference was not significant (Figure 6D).

When RNAi:*lhy* plants were grown outside in the field, leaf chlorophyll content varied between plants in different field locations, sometimes showing more variation than transgenic vs. WT plants. Chlorophyll content was significantly different between 423 plants grown in two separate rows in Oxford field for both transgenic (Figure 7A), and WT plants (Figure 7B). However, there were no significant differences in average chlorophyll content between transgenic and WT plants (Figure 7C-E). Conversely, for 451 plants grown at the same time as these 423 plants, there were no significant differences between the two rows (Figure

8A,B), but there was a significant difference between transgenic and WT plants from both rows (Figure 8E), though not within each individual row (Figure C,D).

The second phenotype we measured was final plant height (after silking), as well as the length of each internode in some trials. In p*UBQ*:RNAi:*lhy* families 423 and 451, average final plant height was slightly higher in transgenic plants than their non-transgenic siblings, but neither difference was significant (Figure 9A,B). When comparing internode length, the transgenic plant averages from p*UBQ*:RNAi:*lhy* families 423 and 485 were the same as their non-transgenic siblings (Figure 9C). For five selected *lhy1*-OX families with the highest *lhy1* expression, average internode length was much shorter than the family-nonspecific WT control for all but the first four internodes (Figure 10A), with non-overlapping 95% confidence intervals for the area under each curve (Figure 10B). For seven selected *lhy2*-OX families with the highest *lhy2* expression, many had average shorter internode lengths for some of the internodes measured (Figure 10C). Many *lhy2*-OX lines (3112, 3149, 3204, 381) had shorter internodes than WT (family-nonspecific) for all internodes in similar patterns (Figure 10C). 3101 internodes were only shorter than WT up to internode #12; and 342 internodes were the same as WT until #7, #8-11 were equal, and #9 and up were longer than the control; 313 had drastically shorter internodes than any other line (Figure 10C). Some of this variation may be due to having WT controls from only one line for comparison, so transgenic families were not compared to non-transgenic siblings in exactly the same genetic background. When comparing 95% confidence intervals for the area under each curve, the CIs for 3112 and 342 overlapped with the non-transgenic control, while 3101, 3149, 3204, and 381 were smaller and did not overlap with the non-transgenic control, and 313 was much smaller (Figure 10D).

The third phenotype we measured was average seed weight, a classical yield metric. Seed weight measures a different allocation of resources than plant height. Compared to non-transgenic siblings, the average seed weight was lower for 423 and 451, and significantly so for 423 (Figure 11).

lhy2-OX 313 growth phenotypes are different in inbreds vs. hybrids

In order to test the ability of hybridization to change *lhy* expression and increase plant size and yield, we crossed *lhy2*-OX 313, the *lhy* misexpression line with the strongest phenotypes, in its B104 inbred background to Mo17 (a well characterized inbred, often used to study hybrid vigor) to make hybrid transgenic families. No other *lhy* misexpression line had consistent growth phenotypes, and we wanted to see if hybridization could rescue any growth phenotypes. As explained in the introduction, changes in *lhy* expression have been shown to affect hybrid vigor in Arabidopsis, and monitoring *lhy* expression and growth phenotypes in 313 was one way to test this relationship.

Both the transgenic and non-transgenic plants were taller in a hybrid background than in an inbred background – the transgenic average doubled while the non-transgenic average was 1.4x higher (Figure 14A). Because the 313 transgenic hybrid was nearly equal in height to the nontransgenic hybrid, corresponding to a height ratio of nearly 1 (Figure 14C), the transgenic hybrid had a greater increase in height; this was likely because the 313 transgenic inbreds were

much shorter than non-transgenics (Figure 14A). On the other hand, transgenic leaf chlorophyll A and total chlorophyll content was significantly lower in both 313 transgenic inbreds and hybrids (Figure 14B, 14C).

In 313 hybrid non-transgenics, peak *lhy* expression shifted closer to ZT6 (Figure 14E) versus ZT3 in inbreds (Figure 14D). In 313 transgenics, *lhy1* expression didn't change between inbreds and hybrids, and the peak was consistently at ZT3 (Figure 14D,E). Conversely, *lhy2* expression did change: in inbreds it was constitutive with no peak (Figure 14D), and in hybrids there was a pronounced peak at ZT3 (Figure 14E). These results suggest that hybridization can in fact change *lhy* expression and increase plant growth.

Hybrid vigor correlates with *lhy* expression at some time points

In the last section, we examined one specific hybrid combination with altered *lhy* expression. Next, we wanted to look for more general patterns of *lhy* expression that correlate with hybrid vigor. In order to do this, we measured *lhy* expression and three physiological characteristics – stem height, leaf length, and leaf width – in 4 sets of inbreds and their reciprocal F1 generation hybrid combinations. We calculated the mid-parental value (MPV) for each measure, which is the mean of the two parental values, then compared the two reciprocal hybrids to the MPV. The only hybrid combination that showed a statistically significant increase from the MPV was B73xA632 for stem height and leaf length (Figure 15). Averages for Mo17xB73 and B73xPalomero hybrids were also higher than the MPV, but differences were not significant (Figure 15).

When comparing *lhy1* and *lhy2* expression at ZT10 between inbred parents and their reciprocal hybrid offspring, we examined two different leaves to include developmental variation. Between ZT10 *lhy1* and *lhy2* expression in leaves 2 and 3, the hybrid/MPV ratio varied from 0.25-1: the hybrid/MPV ratio more than doubled from leaf 2 to leaf 3 in Mo17xB73, halved in B73xMo17 and Mo17xB73, and jumped from ~0 to 0.5-0.75 in PalxB73 (Figure 16). It is unlikely that these drastic changes between leaves were random, because they were similar for *lhy1* and *lhy2* in each hybrid (Figure 16).

For the B73xA632 hybrid, the inbred combination with the most vigor in physical measurements, *lhy1* and *lhy2* expression was undetectable at ZT6, then similar to the MPV in both leaf 2 and 3 at ZT10 (Figure 16A). For the reciprocal hybrid A632xB73, which showed less vigor, *lhy2* expression was low at ZT6 and *lhy1* was slightly higher. At ZT10, A632xB73 expression was slightly to 1.3x higher than B73xA632. For B104xMo17 and its reciprocal hybrid, neither of which showed vigor, *lhy* expression in 2/6 measurements was close to the MPV, halved for *lhy2* at ZT6 and 10 in leaves 3 and 2 respectively, and doubled for *lhy1* and *2* at ZT10in leaf 3 (Figure 16B). For Mo17xB73, which showed some vigor, and B73xMo17 which showed less, *lhy* expression was similarly halved or almost 0 in 4/6 measurements, then close to the MPV in two, but somewhat lower in MxB which showed more vigor (Figure 16C). Finally, B73xPalomero showed slightly more vigor than its reciprocal hybrid, though both showed some. These reciprocal hybrids had by far the greatest differences in *lhy* expression (and the greatest genetic distance), but were similarly close in physical vigor to other inbred combinations. With the exception of

PalomeroxB73 in 2/6 measurements, both reciprocal hybrids had *lhy* expression less than half the MPV, and PalomeroxB73expression was much higher than B73xPalomero (Figure 16D).

In order to simplify this analysis, we plotted *lhy* expression/MPV against physical hybrid vigor/MPV in various combinations and analyzed each with a linear regression. Then we plotted the p-value for the slope of each combination (Figure 16E). Hybrid vigor for height was by far best predicted by the sum of *lhy1* and *lhy2* expression at ZT6, and this correlation was significant. Conversely, hybrid vigor for leaf width was best predicted by the sum of *lhy* expression at ZT10, also significant. The correlation between *lhy* expression and hybrid vigor was never significant for all three measures of vigor combined, or leaf length, or the summed and averaged *lhy* expression across timepoints and leaf samples (Figure 16E). Hybrid vigor of all three measures combined and hybrid vigor for leaf length were both best predicted by the combined ZT6 and ZT10 samples for leaf 3. Overall, *lhy1* and *2* expression at ZT6 best predicted hybrid vigor.

lhy2 Mu lines knockdown *lhy2* expression 30-60%

Both Mu insertions in *lhy2* caused a significant decrease in *lhy2* expression at the ZT3 peak, either when homozygous, or when both homozygous and heterozygous. In *lhy2*m154, *lhy2* expression was about 30% of WT, and *lhy2*m160 was 60% of WT (Figure 13B,D). In both mutants, *lhy1* expression was decreased in a similar pattern to *lhy2* expression, but none of the differences were significant (Figure 13A,C). This knockdown was much smaller than expected for Mu insertional mutations, and therefore we did not further characterize these lines.

DISCUSSION

lhy negatively regulates *toc1* expression in maize and *lhy* contributes to maize growth

The goal of this project was to answer four questions. The first two questions were: How do the maize *lhy* orthologs affect the circadian clock in maize? Does LHY contribute to maize growth? In order to answer these questions, we thoroughly characterized transgenic maize *lhy* misexpression lines, and saw that *lhy* seems to negatively regulate *toc1* in maize; additionally, altering *lhy* expression had surprisingly small and inconsistent effects on plant growth.

The plant circadian clock model includes negative regulation of *TOC1* by CCA1 and LHY. Here we showed that this is still the case in maize although maize has two paralogs of *lhy* and no homolog of CCA1. We saw that decreased *lhy* expression resulted in an advanced phase of *toc1* and *gi*, which has been previously shown in an Arabidopsis *cca1 lhy* double mutant (Mizoguchi et al., 2002). We also observed that the 313 *lhy2*-OX line caused a phase delay and decreased amplitude of *lhy1* (Figure 3), which suggests that the two *lhy* paralogs reciprocally regulate one another. This has been shown in previous studies in Arabidopsis as well (Wang and Tobin, 1998). Based on *cca1* and *lhy* mutants and overexpression lines in Arabidopsis, we expected to see significant perturbation of the circadian clock and measurable growth phenotypes in maize *lhy* misexpression lines (Schaffer et al.,

1998; Wang and Tobin, 1998; Green and Tobin, 1999). With our many individual transgenic lines for RNAi:*lhy*, *lhy1*-OX, and *lhy2*-OX, we did occasionally see significant changes in chlorophyll content, internode length, plant height, and seed weight. However, with the exception of *lhy2*-OX 313 which I will discuss in detail in Chapter 4, these effects were small, inconsistent, and occasionally significant.

Leaf chlorophyll content was higher than WT on average in p*UBQ*:RNAi:*lhy* 4 lines, but only significantly higher once in 423 (Figure 6C) and lower once in 451 (Figure 8E). Internode length and height were slightly higher than WT on average in p*UBQ*:RNAi:*lhy* lines, but again these differences were never significant (Figure 9A,B,C). Average seed weight was significantly lower than WT in 423 (Figure 11A). In the T1 generation, internode length for multiple *lhy1*-OX and *lhy2*-OX lines was significantly decreased (Figure 10), but this only persisted in later generations in 313. Leaf chlorophyll content was never significantly different in any *lhy*-OX line except 313.

We also measured transgene copy number by ddPCR to see if it correlated with transgene expression level growth effects. p*UBQ*:RNAi:*lhy* lines 423 and 451 had similar growth phenotypes; both had significantly more chlorophyll than WT in one experiment, but 451 also had significantly lower average seed weight than WT. They had very comparable *lhy* expression, each with 3-4x knockdown compared to WT, and a 1-3h *toc1* phase advance. The slightly stronger phenotypes in 451 correlated with its higher copy number (3 vs. 2), but we were not able to do ddPCR for more p*UBQ*:RNAi:*lhy* lines to look for a more widespread pattern. In published data, our transgenic line with the biggest change in *lhy* expression (not necessarily the highest transgene expression), 313, had a copy number of one, while weaker lines 2149 and 3101 had copy numbers of four and two, respectively (Collier et al., 2017). Therefore, transgene copy number does not seem to be correlated with transgene expression or growth effects.

There are a few possible explanations for these weak effects of *lhy* expression perturbation on growth. One is that we failed to create a sufficient degree of *lhy* knockdown or overexpression. The *lhy2* Mu insertion lines did not appear to have a large effect on *lhy* expression, so we were never able to characterize a full *lhy* knockout. The primers we used to detect lhy expression could theoretically amplify from transcripts containing Mu sequence, which should be unable to make functional protein. However, we did not successfully make a native antibody for measuring *lhy* protein abundance in the Mu insertion lines to answer this question. With the exception of 313, no transgenic line had more than a 3-4-fold change in *lhy* expression. However, we expected that in p*UBQ*:RNAi:*lhy* 423 and 451 – with 3-4-fold *lhy1* and *lhy2* knockdown and a 1-3h *toc1a* and *toc1b* phase advance in driven conditions and free run respectively – we would see larger effects on plant growth. The gene expression data showed that these decreases in *lhy* expression are large enough to perturb the circadian clock as measured by *toc1* expression, so we would expect a change in target gene expression and thereby output pathways which in turn affect growth.

Another possible explanation for seeing only weak effects of *lhy* misexpression on growth is that these *lhy* expression changes and growth phenotypes were variable and environmentally-responsive. One example of

phenotype variation in response to the environment was an experiment measuring chlorophyll content in field-grown plants – averages were significantly different between plants of the same genotype grown in slightly different field locations (Figure 7). This variation in response to environmental variables could have obscured small differences between transgenic plants and WT.

Finally, it is possible that the lack of large growth effects in transgenics with perturbed *lhy* expression reflects a biological reality. The maize circadian clock contains much more redundancy than Arabidopsis and most other grasses (Bendix, 2015), so it is possible that other *rve* genes besides the *lhy1* and *lhy2* paralogs have overlapping functions. In this case, even though we knocked down expression of both *lhy1* and *2*, there could be still more redundant genes to cover *lhy* functions. Alternatively, some research has shown that the monocot circadian clock, unlike the dicot clock, may be uncoupled from growth: maize, rice, and Brachypodium growth rates oscillate in temperature but not light cycles, and do not cycle in constant light or temperature (Poiré et al., 2010; Matos et al., 2014). In this case, although we changed *lhy* expression enough in some lines to perturb the maize circadian clock, perhaps the clock is not involved in growth output.

Some maize hybrids have increased *lhy* expression at ZT6 and ZT10 compared to inbreds, and changes in *lhy* expression do contribute to hybrid vigor

The second two questions we aimed to answer with this study were: Do maize hybrids have the same altered LHY expression pattern as Arabidopsis hybrids (*lhy* downregulation at ZT6 and 9)? Do changes in LHY expression contribute to vigor? In order to ask these questions, we made hybrid transgenic families with *lhy2*-OX 313 and measured *lhy* expression and growth compared to their inbred parents. We also measured *lhy* expression and growth in four sets of inbreds and their reciprocal hybrid combinations.

After we completed this study, the Harmon and Chen labs published a paper in 2016 arguing that *lhy* (which was called *cca1* in this 2016 paper) was upregulated early in the day at ZT0 and 3 in maize hybrids, unlike in Arabidopsis (Ko et al., 2016). Therefore we may have missed some of the relationship between *lhy* expression and hybrid vigor in these experiments by measuring *lhy* expression at ZT6 and 10 rather than ZT0 and ZT3.

In B73xMo17 hybrid non-transgenic 313 plants, we saw peak *lhy* expression shifted toward ZT6 compared to ZT3 in inbreds. This pattern does not match the published theory that increased *lhy* expression at ZT0 and 3 contributes to maize hybrid vigor, shown particularly in B73xMo17 hybrids (Ko et al., 2016). In transgenic 313 inbreds *lhy2* expression was constitutive, but a pronounced peak appeared at ZT3 in transgenic hybrids. This suggests, remarkably, that hybridization was sufficient to restore *lhy2* rhythmicity in *lhy2*-OX plants.

When we examined *lhy1* and *lhy2* expression in four hybrids at ZT6 and ZT10 in leaves two and three, it was difficult to discern a pattern of expression that correlated with the amount of hybrid vigor exhibited by each inbred combination. However, once we aggregated all this data and graphed the hybrid/MPV ratio for *lhy* expression and physical hybrid vigor, some patterns emerged. This aggregate

approach was more appropriate since we were looking for a general rule of *lhy* expression and hybrid vigor in maize.

For this experiment I chose to examine older plants than in previous studies, which sampled 5-14 DAP (Ko et al., 2016). With these older, much larger, plants we couldn't sample all aboveground tissue and see total gene expression averaged across tissues, so we sampled two individual leaves. Unexpectedly, we saw that *lhy* expression varied considerably even between two mature leaves separated by a single internode. However, without doing a time course we can't know whether the differences between these two leaves were caused by different circadian periods or phases or just different levels of expression.

Another difference between this experiment and previous studies was the amount of hybrid vigor detected. We measured morphological features of older plants, while previous studies measured total dry weight and physiological changes in sugar levels. Ko et al. (2016) detected statistically significant differences between B73xMo17 hybrids and the MPV, while we did not. Finally, we expected more heterosis in the Palomero hybrids than others because these two inbreds (Palomero is technically a landrace, not an inbred) are most genetically distant, and hybrid performance generally increases with the genetic distance of the inbreds crossed (Moll et al., 1965; Frisch et al., 2010).

Overall we found that *lhy1* and *lhy2* expression at ZT6 and ZT10 is somewhat correlated with hybrid vigor for plant height, leaf length and leaf width. In order to better compare this with previous studies we would need to measure *lhy* expression at ZT0 and ZT3 as well as ZT6 and ZT10. Previous studies have focused on the role of *lhy* expression in establishing vigor in young plants, which potentially is self-reinforcing and persists throughout development. In contrast, we found that differences in *lhy* expression exist between inbreds and hybrids even in older plants. Therefore *lhy* expression may play a role in hybrid vigor throughout development.

MATERIALS AND METHODS

Plant materials and growth conditions for maize

Inbred maize lines were obtained from the Maize Genetics Cooperation Stock Center. Plants grown in the greenhouse received supplemental light from red and blue LED and white fluorescent bulbs with ~300 μmol·m^{-2}·sec^{-1}, ensuring a 16h/8h day year-round. Greenhouse temperature setpoints were 78°F daytime and 68°F nighttime. Regular greenhouse fertilizer was NPK 20:20:20 and CalMag 3d/week unless otherwise specified. Plants grown outside in fields were either in Oxford Tract or Gill Tract in Berkeley, CA or in Davis, CA; all field growouts took place between May and September. Plants grown in either field in Berkeley were planted then covered with row cover (Agribon AG-19) for two weeks and unless otherwise specified, and received monthly fertilizer. Plants grown in the Davis field had no row cover and were fertilized pre-planting with 8-24-6 @ 20 gal/acre and sidedressed at 46 gal/acre.

Construction of transgenic lines

lhy-OX transgenic lines were constructed as previously described (Ko et al., 2016). RNAi:*lhy* lines were constructed as follows. For p*UBQ*:RNAi:*lhy* 4: a 600 base pair section of lhy2 coding sequence was amplified, corresponding to nucleotides 1,150 to 1,749 of gene model GRMZM2G014902, from B73 genomic DNA. The *lhy* sequence was moved into pANDA RNAi vector (Miki and Shinamoto, Plant Cell Phys 2004; PMID: 15111724) with GATEWAY LR reaction. This generated an RNAi cassette consisting of the maize *ubq* promoter (maize *ubiquitin1* promoter + 1st intron & splicing acceptor site)-lhy(antisense)-gus linker-*lhy*(sense)-NOS terminator. The RNAi cassette was amplified from pANDA by PCR and cloned into pENTR. RNAi cassette in pENTR was moved into vector pTF101.1gw1 with LR reaction to produce final p*UBQ*:RNAi:*lhy* 4 RNAi construct. For p*lhy*:RNAi:*lhy* 5 construct, the *Ubq* promoter was replaced by fusion of the promoter from maize *lhy* and the minimal promoter from the nopaline synthase (NOS) gene from *Agrobacterium tumefaciens*. The *lhy* promoter was a 4,791 base pair region upstream of the transcription start site of gene model GRMZM2G014902.

Gene expression analysis with RT-qPCR

See Chapter 1. For maize samples, reference genes two of the following three, depending on the experiment: GRMZM5G816228, GRMZM2G064954, and GRMZM2G52666.

Transgene copy number measurement by droplet digital PCR

Performed as previously described (Collier et al., 2017).

Measurement of growth phenotypes

Plant height was measured from the top of the prop roots to the collar of the youngest fully expanded leaf. Internode measurements were measured from the middle of one node to the next, and numbered with the first internode starting at

the top of the prop roots. Internode length was measured for all nodes simultaneously after plants reached maturity.

Leaf chlorophyll content assay

Tissue for chlorophyll analysis was collected in 4 circular 2.5mm samples midway between the leaf tip and base, midway between the leaf mid-vein and edge. Tissue was placed on ice. 1mL of DMSO was added to each sample and incubated at 65°C for 30 minutes. Absorbance of the DMSO solution was measured at 645nm and 663nm, and chlorophyll content was calculated based on Arnon's equation as previously described (Richardson et al., 2002).

Chapter 3 Figures

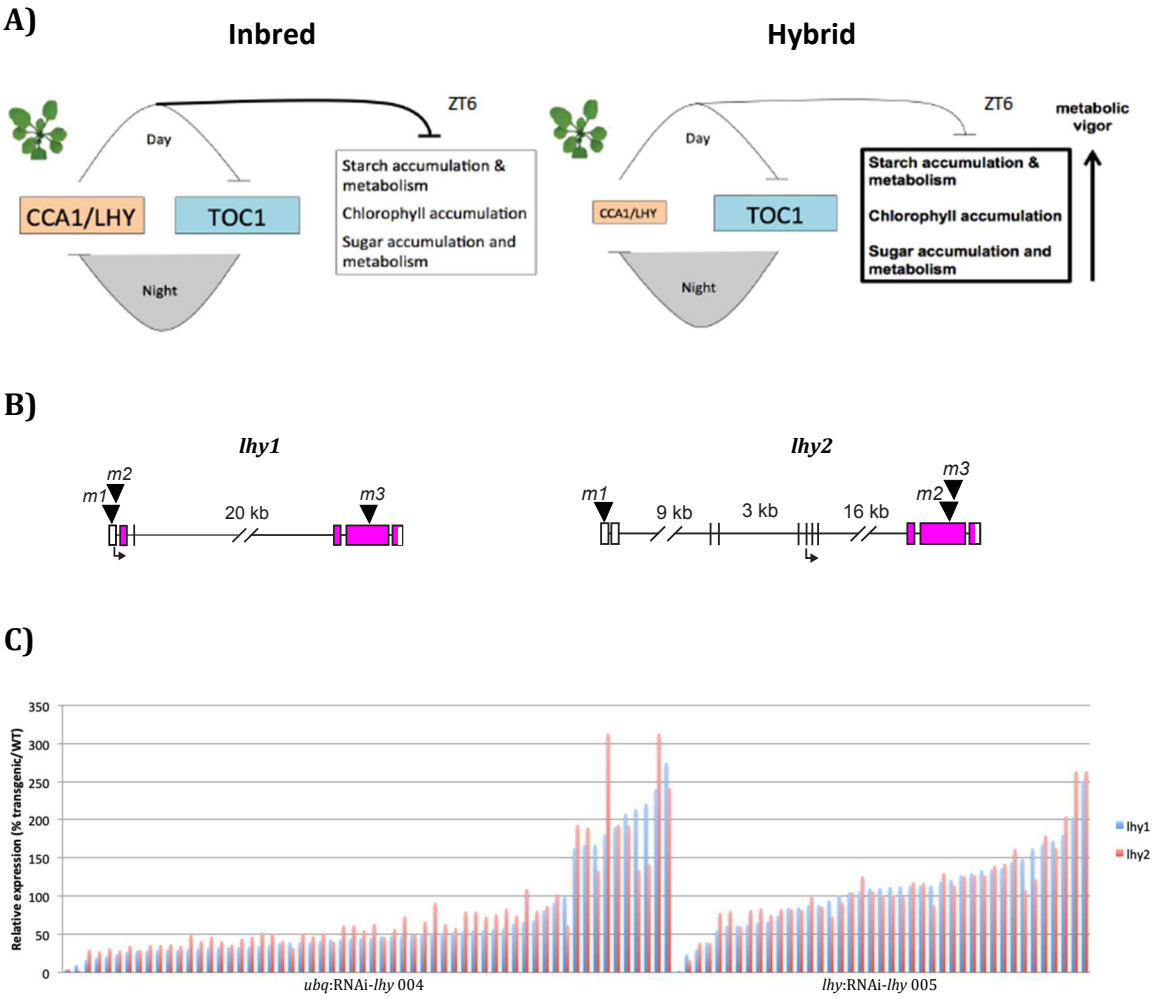


Figure 1. Maize *lhy* expression may be involved in hybrid vigor.

A) Summary of model described in Ni et al. (2009). In Arabidopsis, CCA1 and LHY negatively regulate metabolic processes at ZT6, *CCA1* and *LHY* expression is higher at ZT6 in inbreds than hybrids, metabolic repression is lessened in hybrids due to lower *CCA1* and *LHY* expression, leading to metabolic vigor.
B) Maize *lhy1* and *lhy2* gene models with triangles marking the locations of potential Mu insertions. White boxes are UTRs, pink boxes are CDS, lines are introns.
C) Relative expression (transgenic expression as a percentage of WT expression) of *lhy1* and *lhy2* in maize *RNAi:lhy* transgenic T0 individuals. 004 p*UBQ:*RNAi-*lhy* and 005 p*lhy:*RNAi-*lhy* lines are on the left and right respectively, with values sorted from lowest to highest *lhy1* expression left to right.

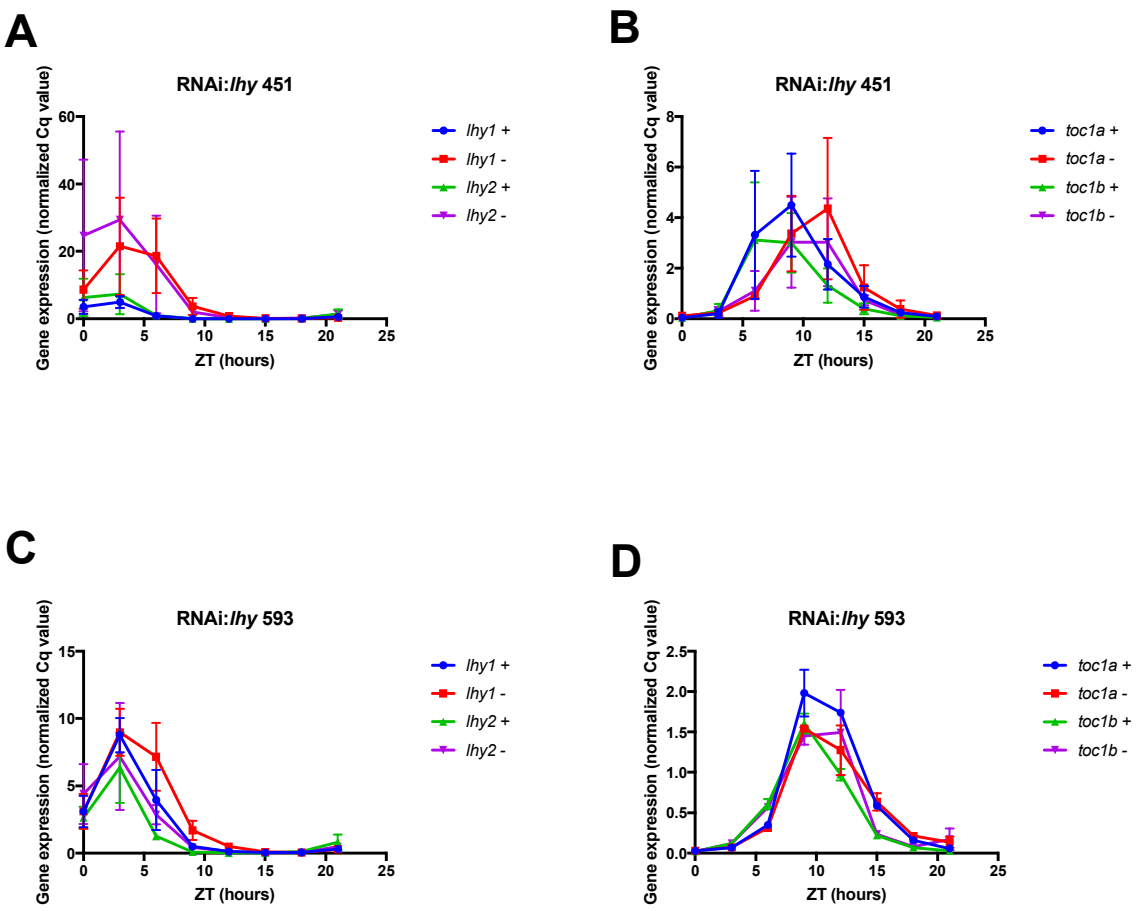


Figure 2. *lhy1*, *lhy2*, *toc1a*, and *toc1b* expression are changed in *pUBQ:RNAi:lhy* 451 but not *plhy:RNAi:lhy* 593.

A)-D) All plants were hemizygotes for the transgene. Gene expression measured by RT-qPCR, normalized to the geometric mean of two reference genes. Points are the average of two technical replicates. Error bars are +- SEM. Plants grown in Oxford field summer 2015. T1 generation. Sampled every 3 hours starting at dawn. Plants sampled at 2 months after planting. 5 plants were sampled per time point per genotype with 2 leaf punches per plant, from youngest fully expanded leaf. Punches taken from leaf tip, moving up leaf with each subsequent time point.

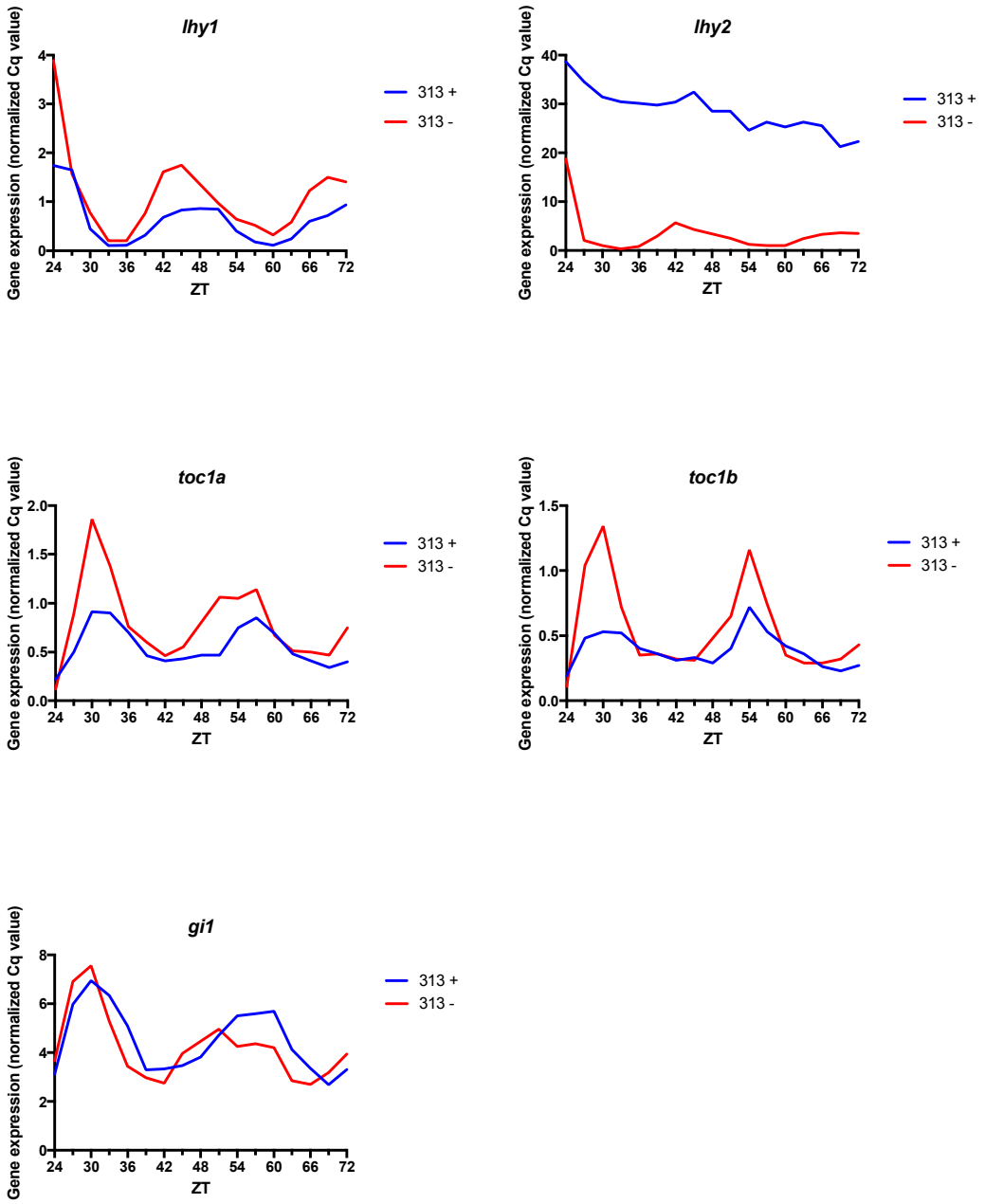


Figure 3. In free run, *lhy2*-OX causes *toc1a* and *toc1b* expression decrease and a phase delay.

Gene expression in *lhy2*-OX line 313 measured by RT-qPCR, normalized to the geometric mean of two reference genes and presented as average of two technical replicates. Plants grown in the greenhouse in late winter with supplemental LED light, released into free run with constant light and temperature at dawn 24 hours before sampling (ZT0 = dawn on the first day in constant light). Sampled every 3 hours starting at dawn, ZT24. Plants sampled at growth stage V8. 5 plants sampled per time point per genotype with 2 leaf punches per plant, one from leaf 7 and one

from leaf 8. Punches taken from leaf tip, moving up leaf until 1/3 of leaf length was sampled.

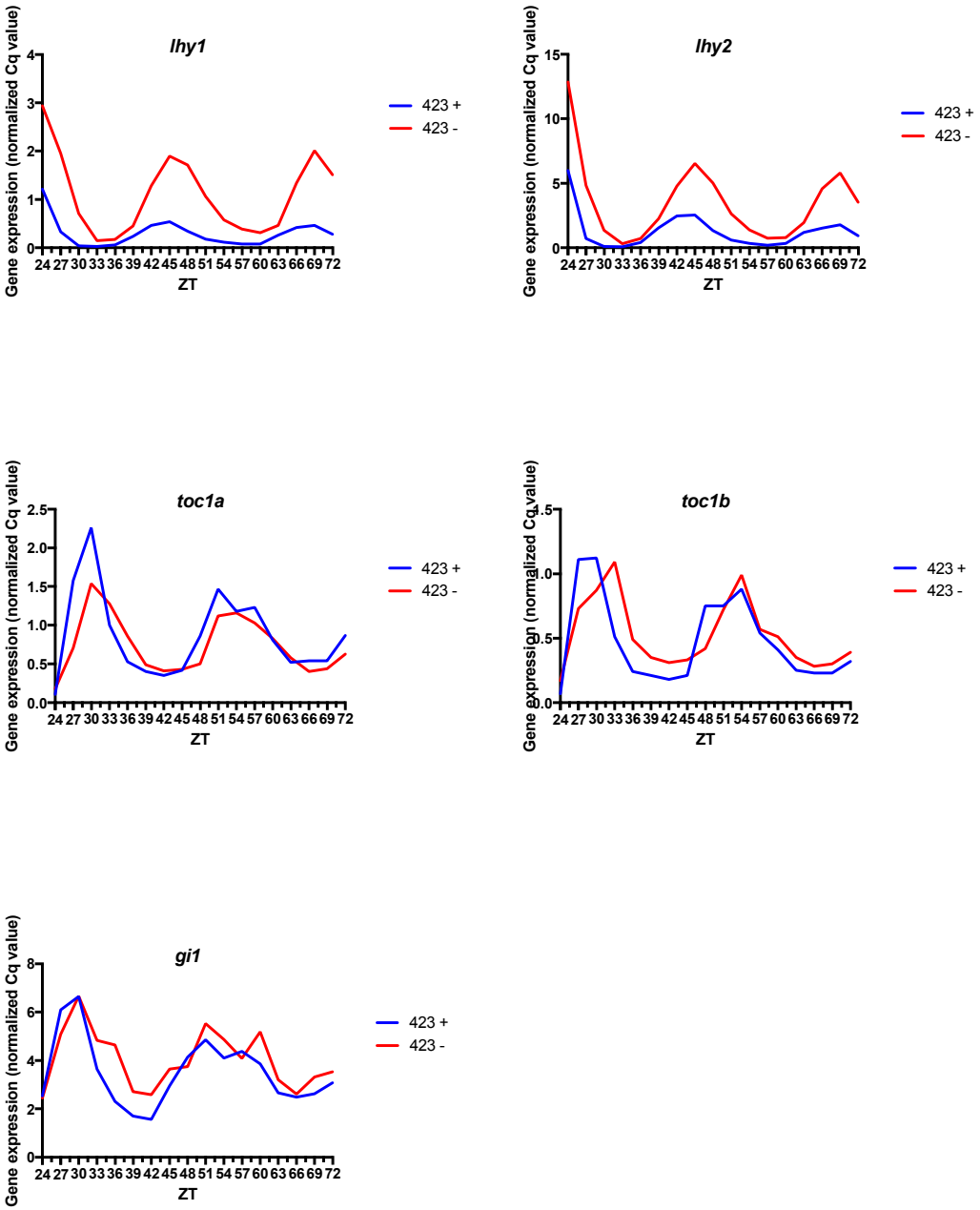


Figure 4. In free run, RNAi-*lhy* causes a phase advance in *toc1a* and *toc1b* expression.

Gene expression in *lhy2*-OX line 313 measured by RT-qPCR, normalized to the geometric mean of two reference genes, presented as the average of two technical replicates. Plants grown in the greenhouse in late winter with supplemental LED light, released into free run with constant light and temperature at dawn 24h before sampling. Sampled every 3 hours starting at dawn. Plants sampled at growth stage V8. 5 plants sampled per time point per genotype with 2 leaf punches per plant, one

from leaf 7 and one from leaf 8. Punches taken from leaf tip, moving up leaf until 1/3 of leaf length sampled.

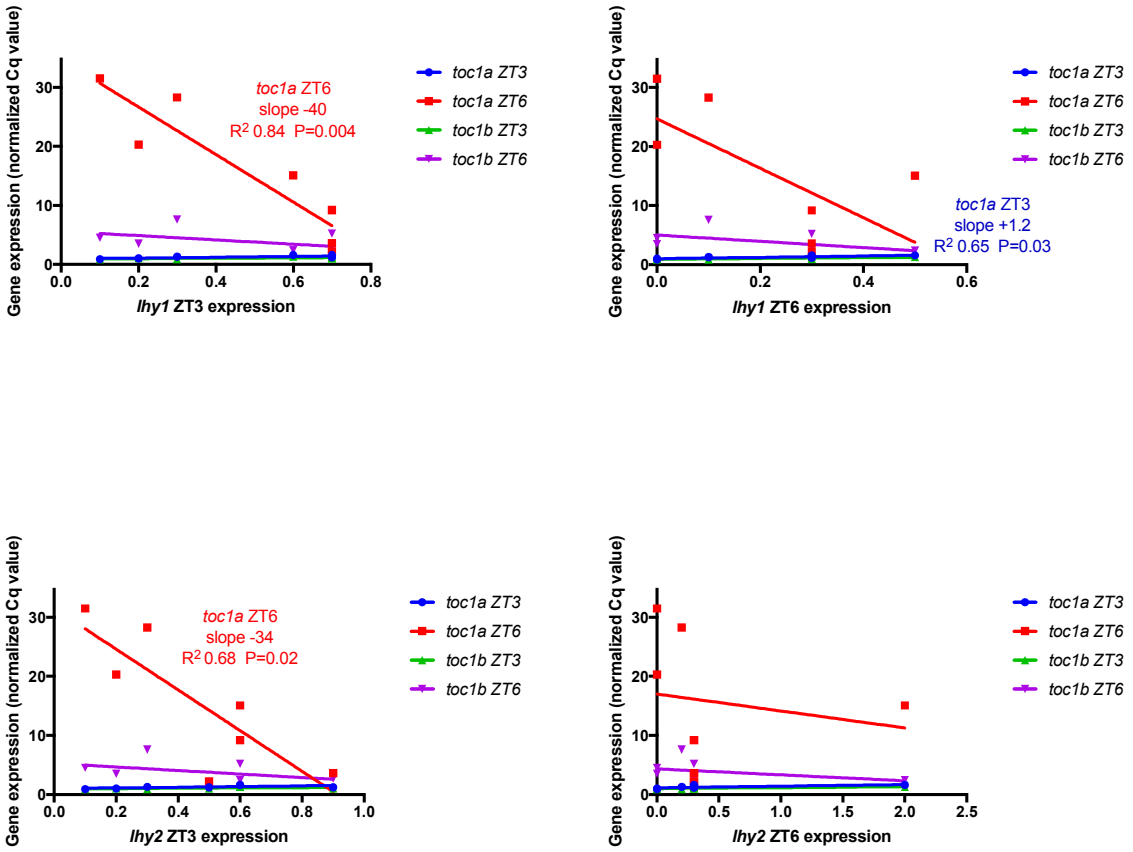


Figure 5. *lhy* expression predicts *toc1a* and *toc1b* expression.

Gene expression for *lhy* plotted against toc1, combinations of ZT3 and ZT6 expression. Linear regressions shown if p<0.05. 7 different transgenic lines, multiple individuals pooled per line. Plants grown in greenhouse.

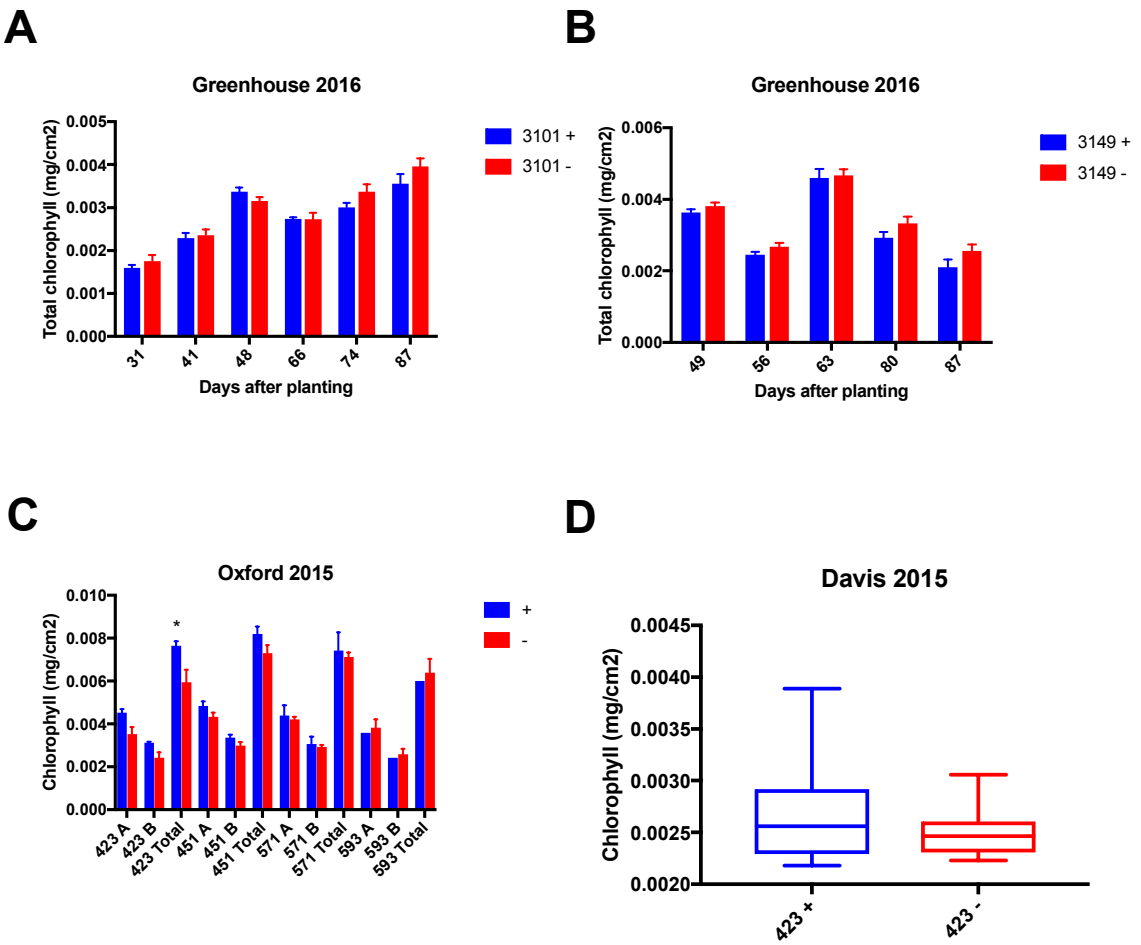


Figure 6. Leaf chlorophyll content is inconsistently affected in RNAi:*lhy* transgenics.

A)-D) Location of plant grow out listed above each graph. "+" and blue for transgenic plants, "-" and red for non-transgenic siblings.
A)-C) Mean with error bars SEM
D) Box plot is line for mean, box for SD, bars for min max.
C),D) T1 generation
N= **A)** 6 **B)** 11, 14 for +, - **C)** 2, except 1 for 593 transgenic **D)** 8, 12 for +, -

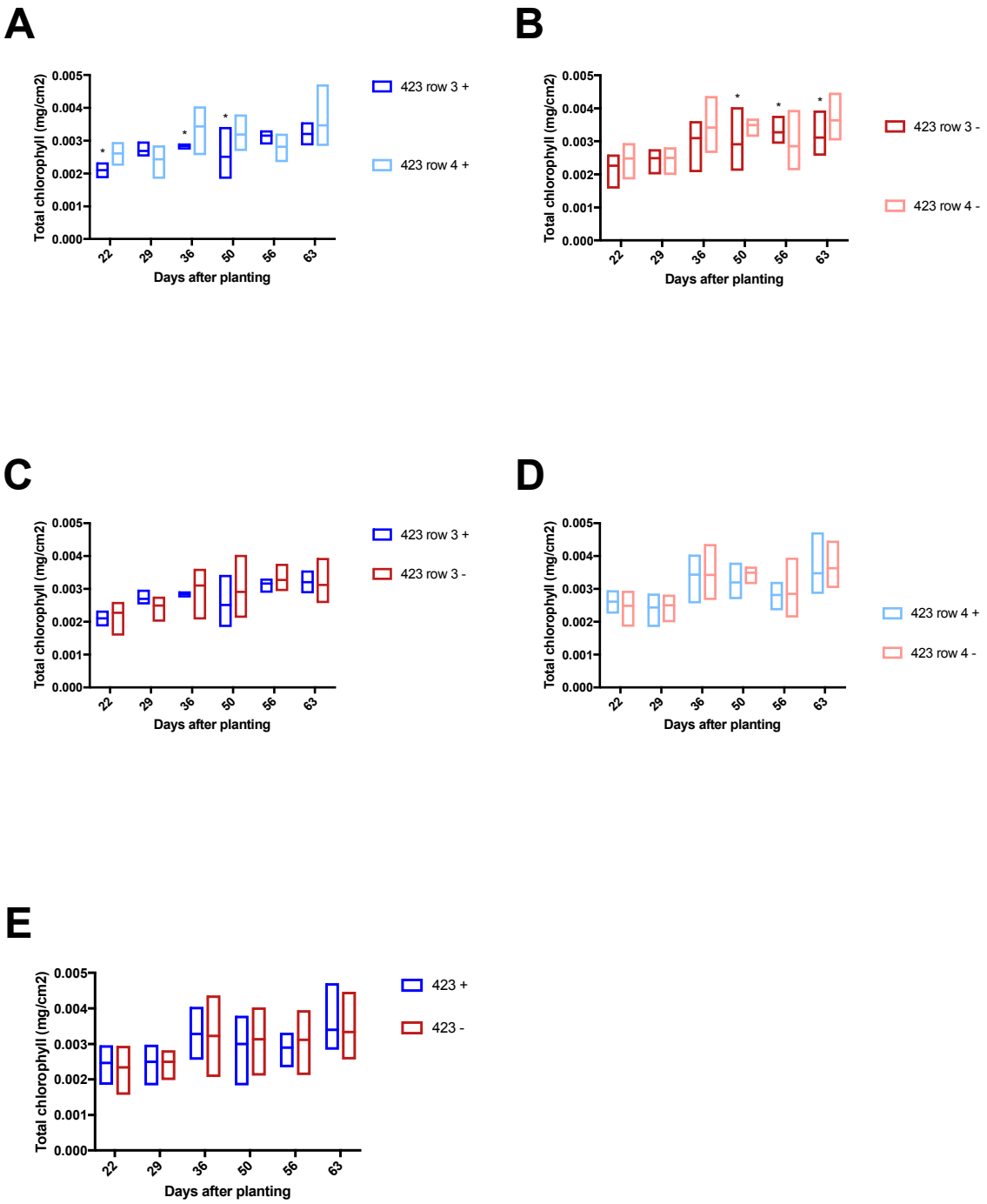


Figure 7. Leaf chlorophyll content is slightly affected in *pUBQ:RNAi:lhy* 423 T2 plants.

A-E) Total chlorophyll in youngest fully expanded leaf at 6 timepoints, measured at the indicated number of days after planting. All plants grown in Oxford field, summer 2016. N=6, 12, 14, 19 for row 3+, -, row 4+, -. Comparisons between: **A)** 423

transgenic or **B)** non-transgenic plants grown in two different locations in Oxford field, rows 3 and 4.

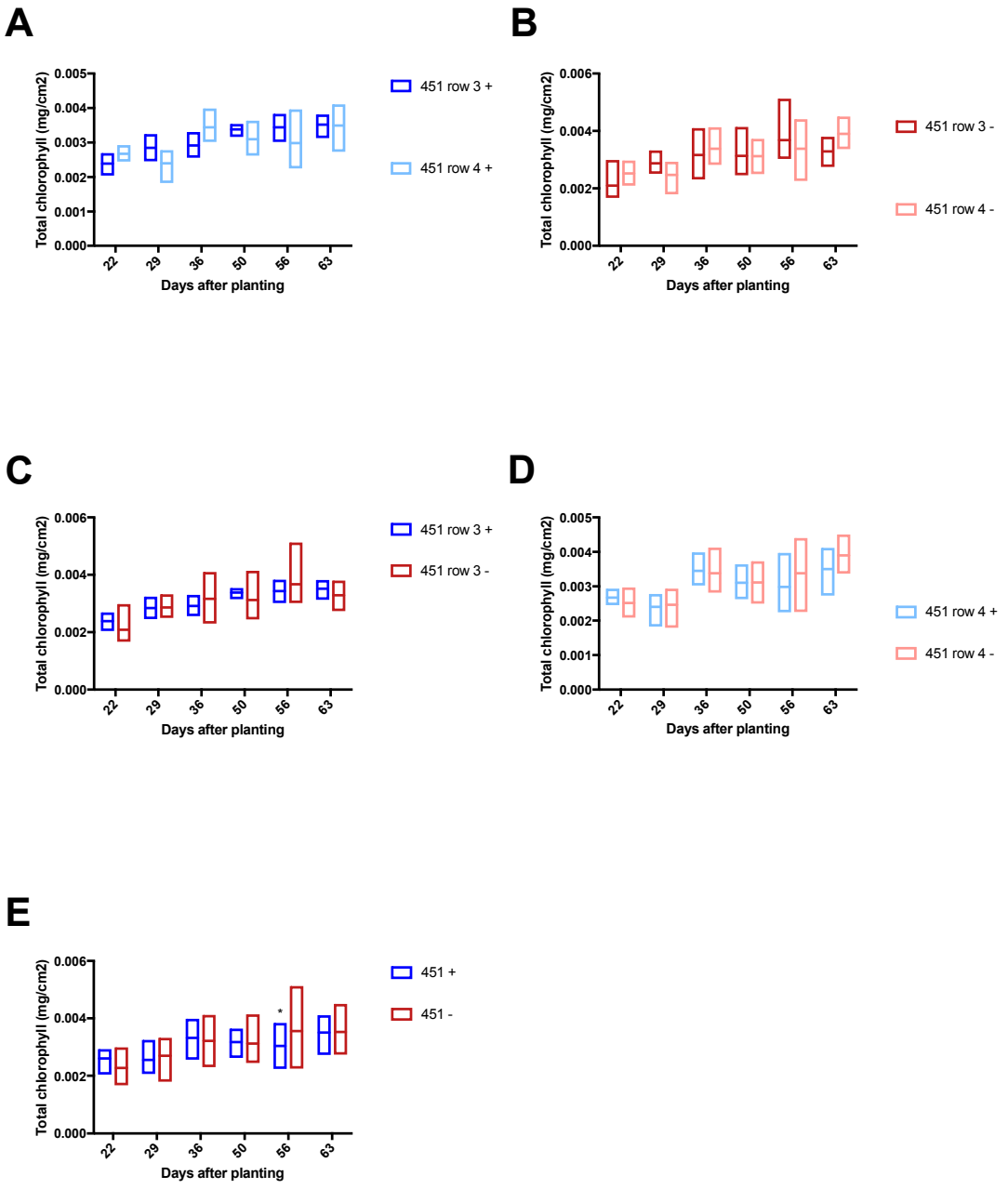


Figure 8. Leaf chlorophyll content is slightly affected in p*UBQ*:RNAi:*lhy* 451 T2 plants.

A)-E) Total chlorophyll in youngest fully expanded leaf at 6 timepoints, measured in days after planting. All plants grown in Oxford field, summer 2016. N=4, 15, 12, 10 for row 3+, -, row 4+, -. Comparisons between **A)** 451 transgenic or **B)** non-transgenic plants grown in two different locations in Oxford field, rows 3 and 4

C) 451 transgenic and non-transgenic siblings grown in the same row 3 or **D)** row 4
E) 451 transgenic and non-transgenic sibling plants, with values pooled from rows 3 and 4

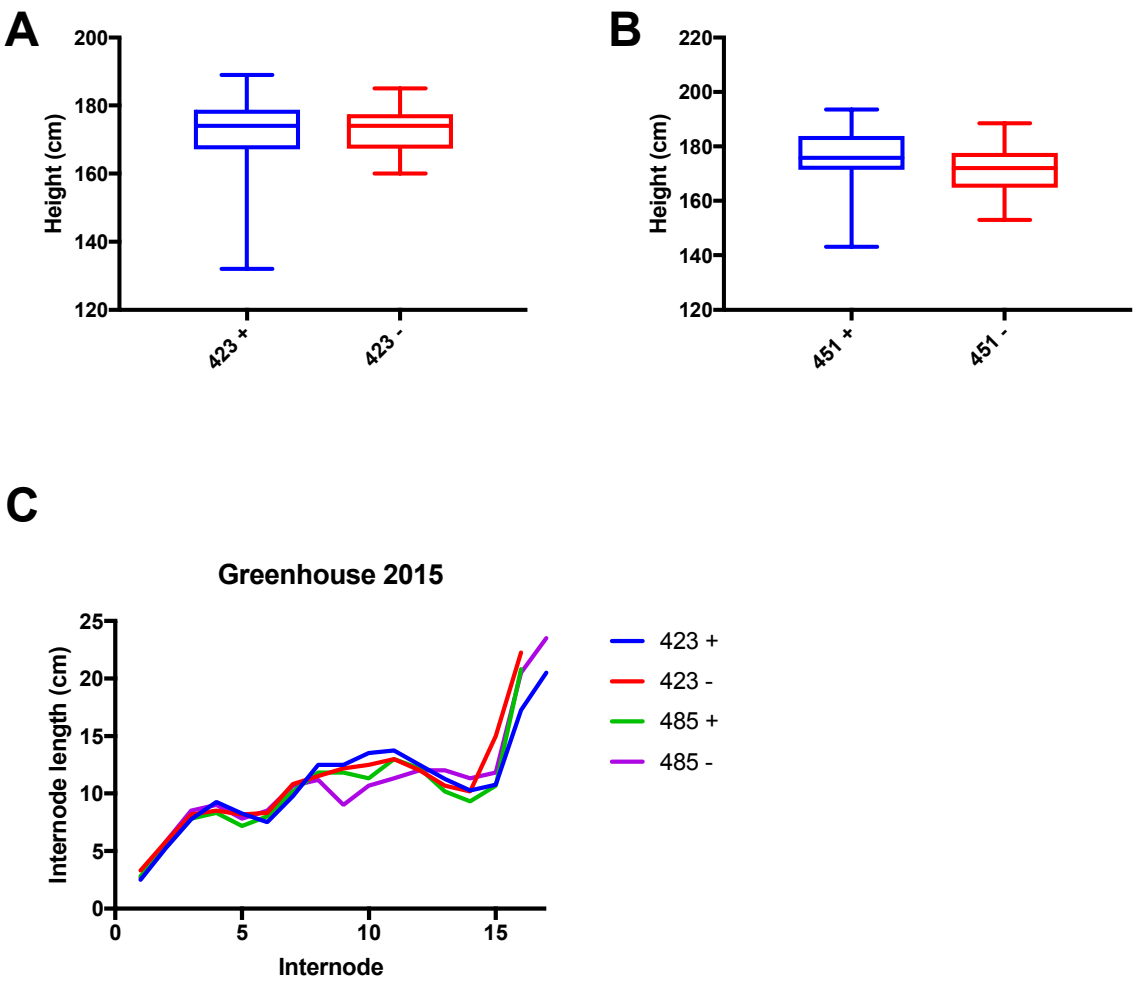


Figure 9. Height and internode length are slightly affected in RNAi:*lhy* and *lhy*-OX lines.

A),B) Box plots are line for mean, box for SD, lines for min and max. Plants grown in Oxford summer 2015. T2 generation.
A) N=20(+), 19(-) **B)** N=16(+), 22(-)

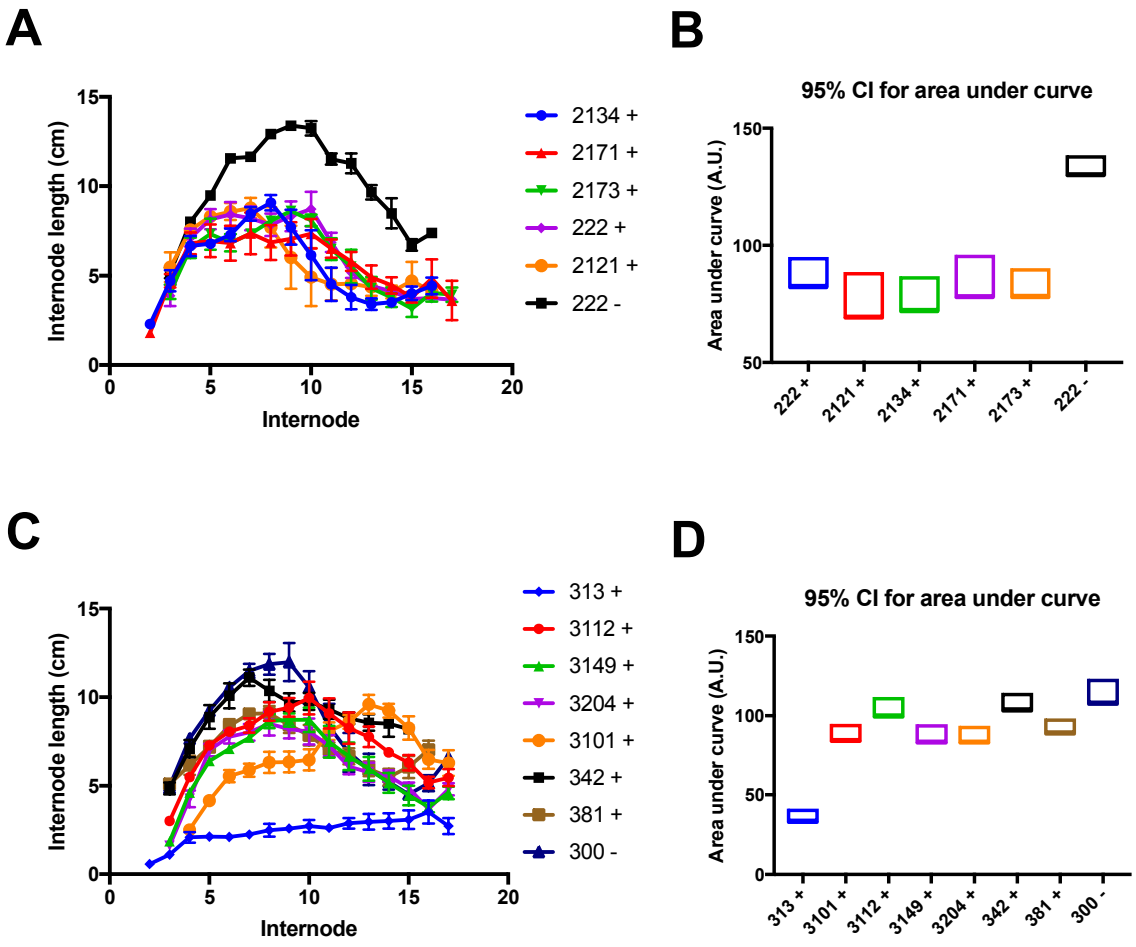


Figure 10. Internode length varies among 002 and 003 lhy-OX lines.

Plants grown in greenhouse 2015. T1 generation.
A),C) Average of N=5 individual plants per line. Error bars are SEM.
B),D) Boxes are for 95% confidence interval for area under curves in **A)** and **C)**

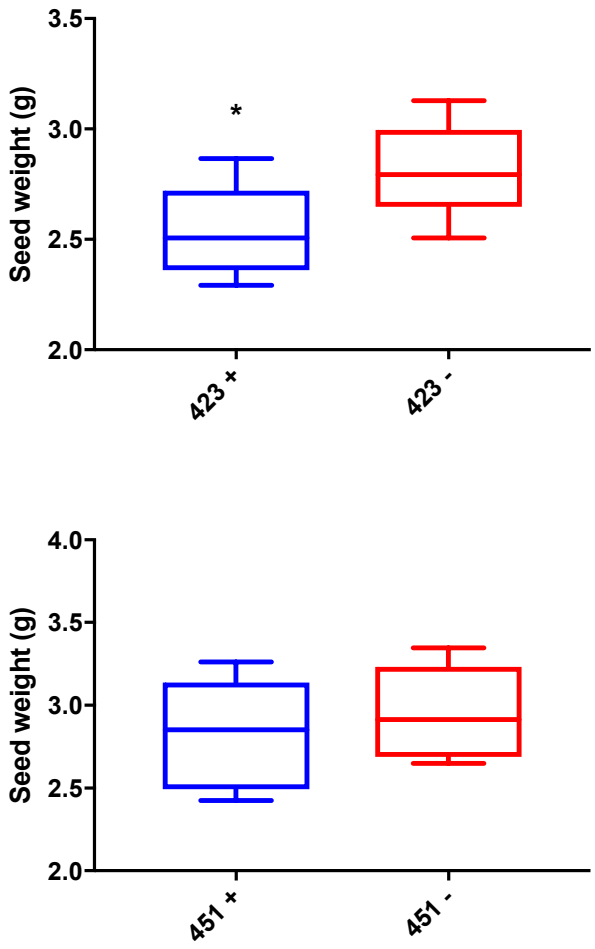


Figure 11. Seed weight is decreased in p*UBQ*:RNAi:*lhy* 423 transgenics, but not 451.

Seed weight is average of 10 kernels. Asterisk marks significant difference between means by Mann-Whitney test. Box plots are line for mean, box for SD, and bars for min max.

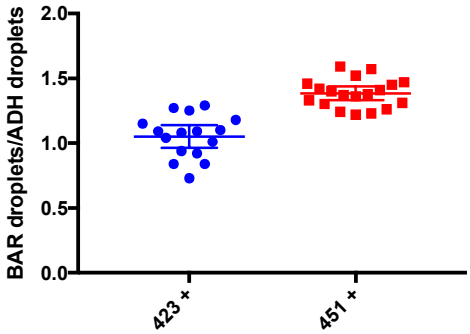


Figure 12. Transgene copy number determination in 423 and 451 lines. *BAR*/*ADH* droplets*2=transgene copy number.

Ratio of droplets from digital droplet PCR. *ADH* (alcohol dehydrogenase) is the single-copy maize reference gene. *BAR* is the transgene marker (encodes a phosphinothricin acetyl transferase, confers resistance to bialaphos). *BAR*/*ADH* droplets*2=transgene copy number, because we maintained these transgenes as hemizygous relative to the reference gene.

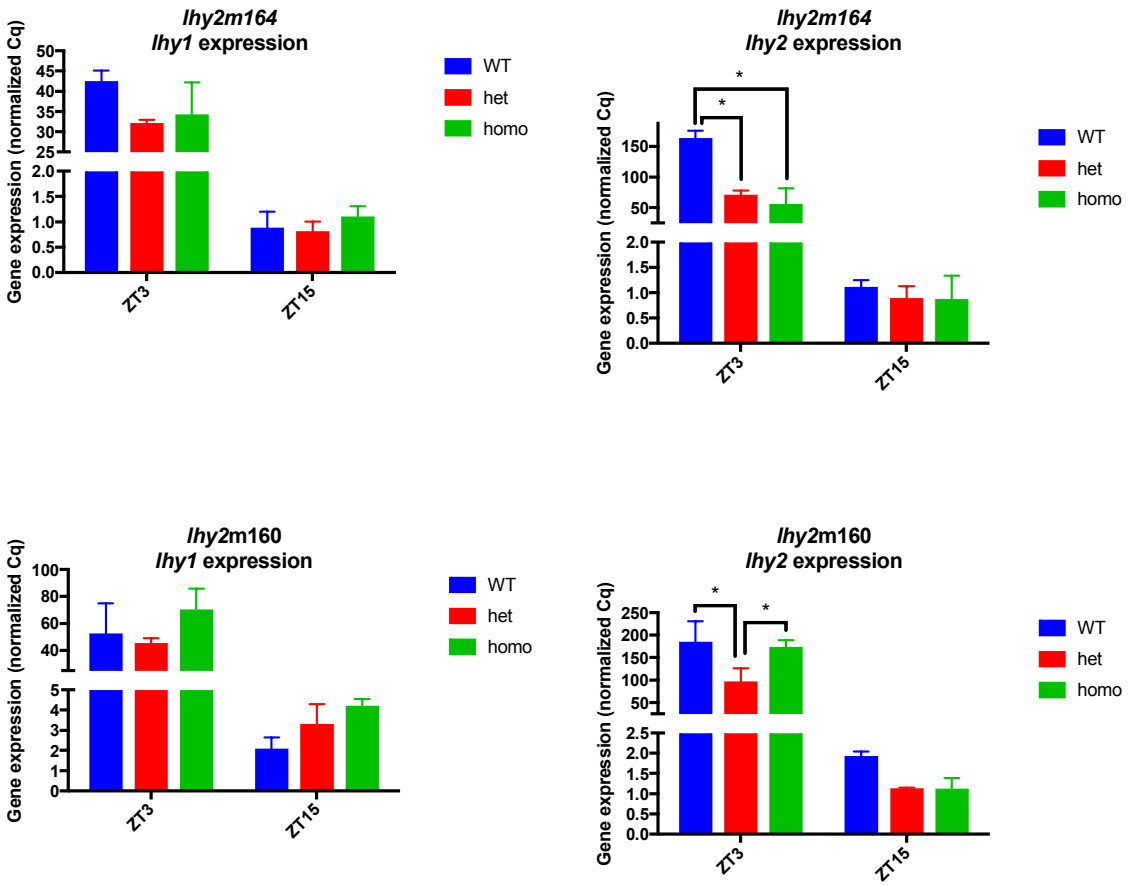


Figure 13. *lhy2* expression is decreased somewhat in *lhy2* Mu lines.

Lhy2 Mu insertions 160 and 164, locations shown in Figure 1B. Gene expression measured after four backcrosses. Gene expression measured by RT-qPCR, normalized to the geometric mean of two reference genes, average of two technical replicates. Plants grown in Gill field summer 2016. Plants sampled near silking. Individual plants sampled with 6 leaf punches per plant. Mean of 3-4 plants per genotype, error bars are SEM. Punches taken from midpoint of leaf.

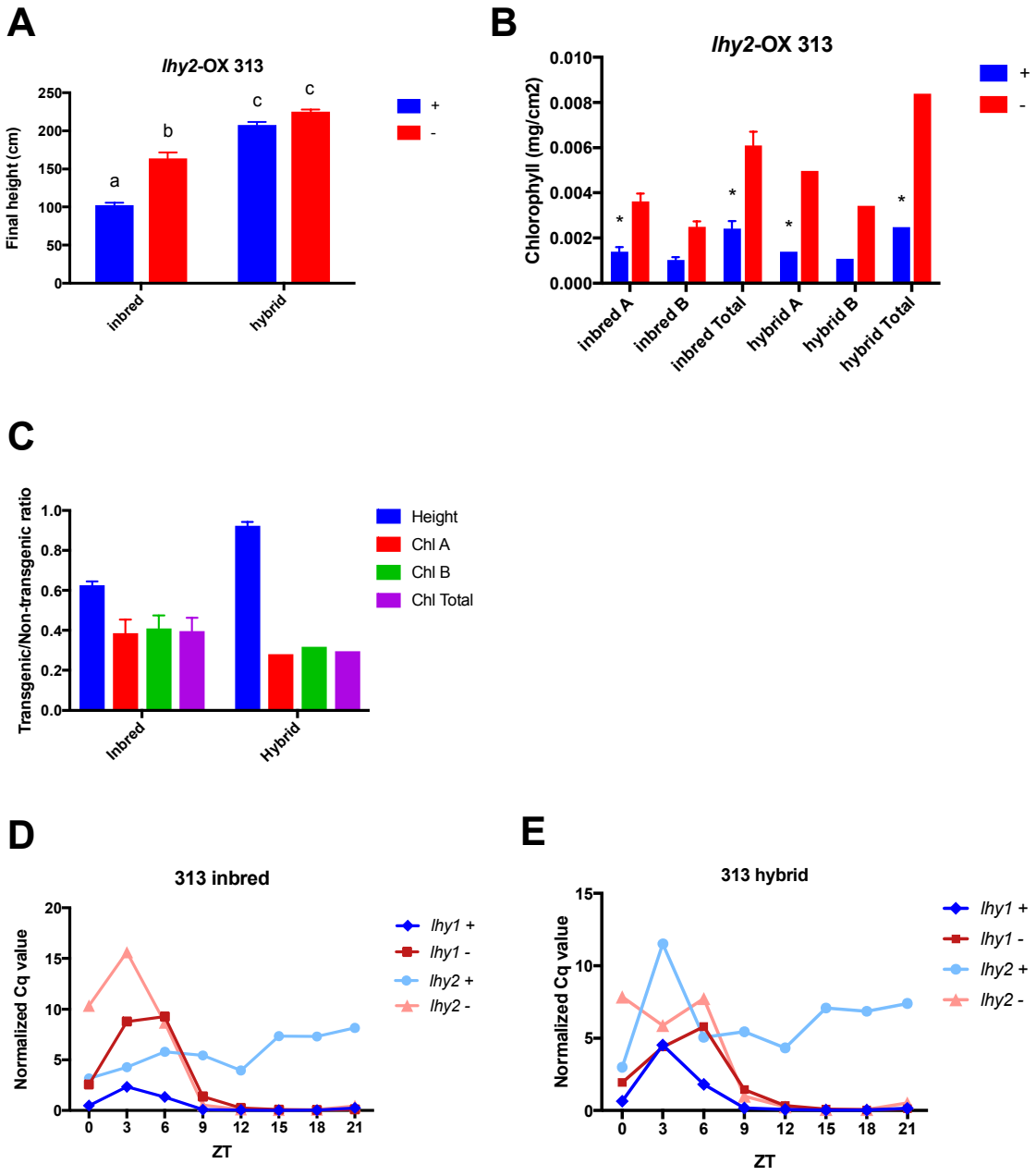


Figure 14. RNAi-*lhy* and *lhy*-OX effects differ in inbred vs. hybrid background.

Plants grown in Oxford field, summer 2015.
A) N=19, 6, 10 14 for inbred +, -, hybrid +, -
B) N=2 for inbreds, 1 for hybrids
C) Transgenic/non-transgenic ratio of height and chlorophyll data from **A)** and **B)**.
Mean, error bars are SEM.

D,E) Gene expression measured by RT-qPCR, normalized to the geometric mean of two reference genes, average of two technical replicates. Sampled every 3 hours starting at dawn, ZT24.

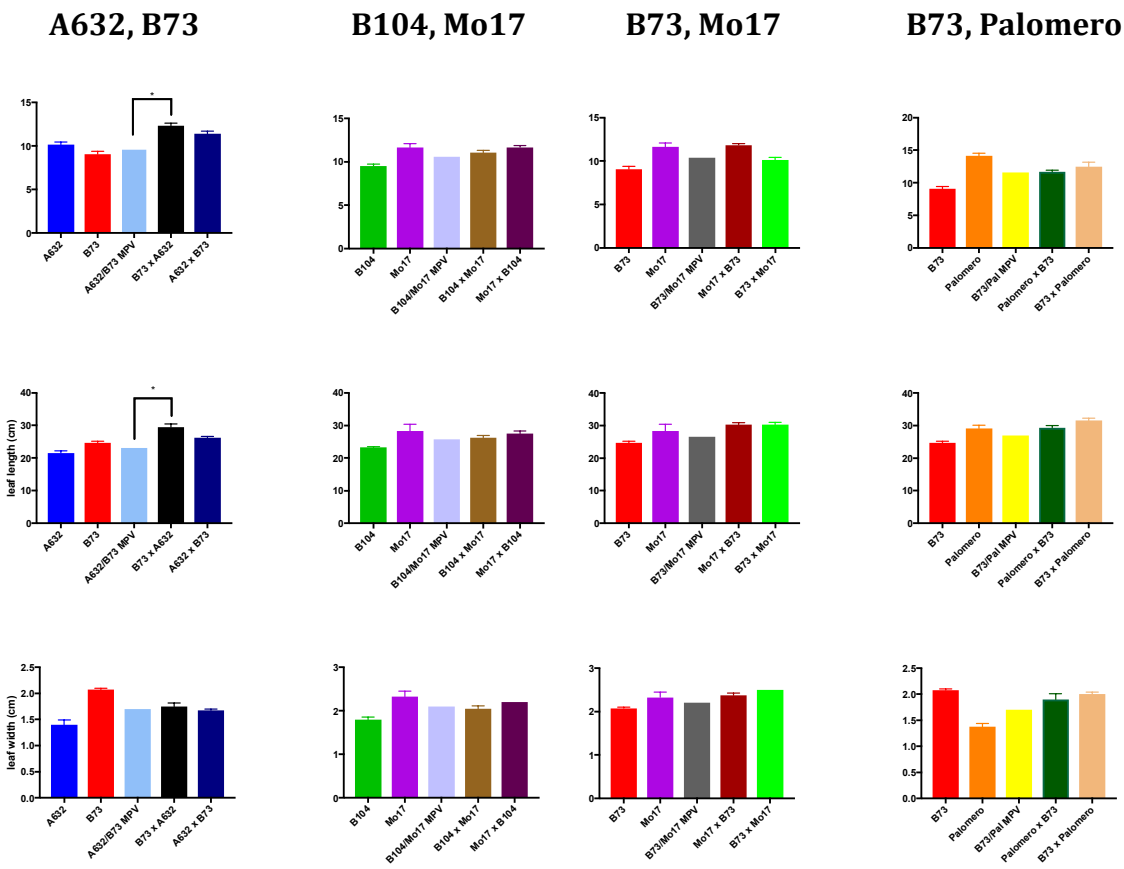


Figure 15. Hybrid vigor varies by hybrid cross and morphological measurement.

Plant height (first row), leaf length (second row), and leaf width (bottom row) in 4 hybrid combinations with reciprocal hybrids (8 total). MPV is average of two inbred parents. Graphed mean with error bars for SEM. Asterisk marks significant difference from WT by student's t-test ($p<0.05$).

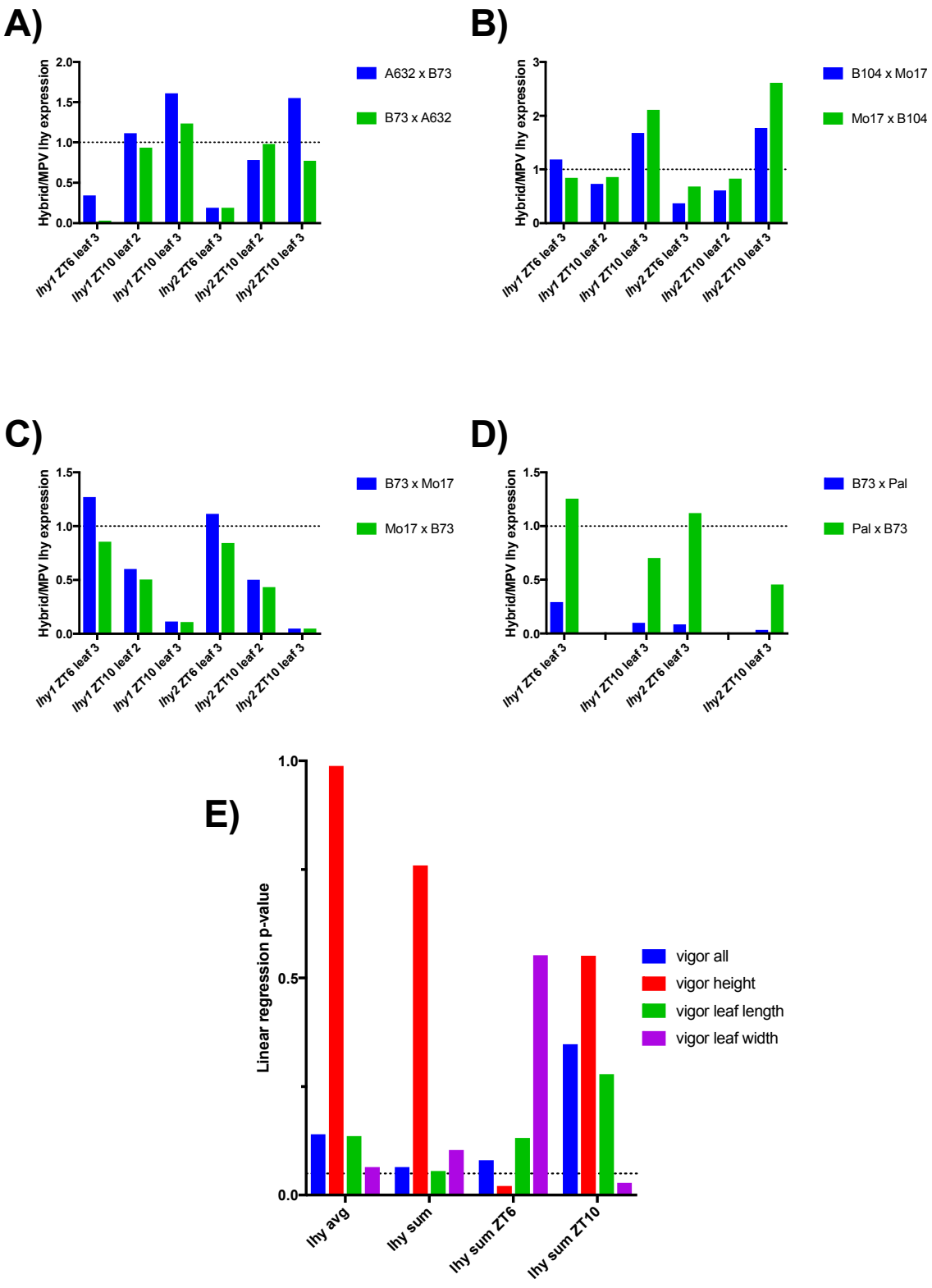


Figure 16. *lhy* expression in different hybrid combinations.

A)-D) Ratio of *lhy* expression in hybrid relative to MPV of inbred parents. Gene expression measured by RT-qPCR, normalized to the geometric mean of two reference genes, average of two technical replicates. Plants sampled at two time points, ZT6 and ZT10, in two tissues, leaf 2 and leaf 3.
E) Performed linear regression between 4 measures of *lhy* expression in hybrid/MPV (average of *lhy* expression at all time points, sum of all, or expression at only ZT6 or ZT10) and average vigor (hybrid value/MPV) for all 3 physical measurements, plus all physical measurements averaged. Graphed p-value for each regression (test for whether slope is significantly non-zero).

TABLES

Table 1. *lhy* expression in *ubq*:RNAi-*lhy* and *lhy*:RNAi-*lhy* from generation T0 to T1.

In left column, condition where plants were grown and number of transgenic line. Generations T0, 1, 2 listed on top. Lines with lowest lhy expression highlighted in green.

	ZT3 transg lhy / lowest WT		ZT3 transg / non-transg sib %			
	T0		T1		T2	
oxford	lhy1	lhy2	lhy1	lhy2	lhy1	lhy2
4 5-1	29	36	19	24	-	-
5 9-3	1	2	104	80	-	-
greenhouse						
4 2-3	32	41	13	14	18	22
4 8-5	44	61	26	29	-	-
5 2-2	39	38	66	62	-	-
5 2-3	61	60	67	53	-	-
5 6-3	30	38	75	89	-	-
5 7-1	24	16	57	61	-	-

Chapter 4: Hemizygous *lhy2*-OX insertion in putative maize glutamyl-tRNA reductase may cause environmentally-responsive interveinal chlorosis

ABSTRACT

Of all the maize lhy transgenic lines characterized in Chapter 3, one *lhy*-OX line 313 had by far the most dramatic growth phenotypes. In addition 313 had pronounced interveinal chlorosis, while only mild chlorosis appeared in other *lhy* misexpression lines. We discovered that the 313 transgene insertion was located in a putative maize glutamyl-tRNA reductase, which may be responsible for a crucial step in chlorophyll biosynthesis. Whether the 313 growth phenotypes were stronger than other *lhy* misexpression lines because of this insertional mutation, or simply because of greater *lhy* overexpression, they were highly environmentally variable. In some growth conditions, we saw complete reversion of the 313 height and chlorosis phenotypes. In order to see what conditions affected 313 growth, we varied temperature, light intensity, and nutrient availability independently. Both increased temperature and greater iron availability partially rescued 313 chlorosis and height phenotypes. We saw no consistent relationship between *lhy* expression and 313 phenotype severity across multiple conditions. 313 growth phenotypes could be due to *lhy* overexpression, or the insertional mutation that potentially affects chlorophyll biosynthesis, or both.

INTRODUCTION

As shown in Chapter 3, one *lhy*-OX line had the highest *lhy* expression by far, in addition to a phase advance in *toc1a* and *toc1b* expression. This one *lhy*-OX line, 313, was the only *lhy* misexpression line with large effects on plant height and leaf chlorophyll content, and was not discussed in depth in Chapter 3. 313 transgenics had one additional phenotype that never appeared in any other line – interveinal chlorosis. Though interveinal chlorosis appeared only in 313, some other *lhy2*-OX lines intermittently showed slightly decreased chlorophyll (Chapter 3), so we thought interveinal chlorosis might be the stronger version of this phenotype. Eventually, after doing all the experiments described in the results below, we discovered the *lhy2*-OX 313 transgene is in a gene that encodes a putative maize glutamyl-tRNA reductase.

Glutamyl-tRNA reductase (GluTR) is the enzyme responsible for the first committed common step of tetrapyrrole biosynthesis. Tetrapyrrole biosynthesis, which includes chlorophyll, heme, and siroheme, has nine steps that make common precursors to all three proteins (the siroheme branch separates earlier), then synthesis pathways for each branch off from these common steps (Tanaka and Tanaka, 2007). ALA is an intermediate in the common steps, and ALA synthesis is the rate-limiting step in tetrapyrrole biosynthesis, which is regulated by GluTR (Tanaka et al., 2011). If chlorophyll biosynthesis is interrupted, singlet oxygen is generated by free tetrapyrroles exposed to light, which causes oxidative damage (op den camp et al., 2013)

Based on the role of glutamyl-tRNA reductase in chlorophyll biosynthesis, the insertional mutation in this gene in the 313 transgenic line could plausibly be responsible for the interveinal chlorosis phenotype. In Arabidopsis, 3 *HEMA* genes (1, *2*, and 3) encode glutamyl-tRNA reductase. *HEMA1* is light-inducible and expressed in photosynthetic tissues, while *HEMA2* and *3* are not light inducible and are expressed in most tissues (Tanaka et al., 1996). A homozygous Arabidopsis *hema1* mutant is small and light green and a double *hema1;hema2* mutant is even smaller and lighter green (Nagai et al., 2007), while *hema2* looks normal (Nagai et al., 2007; Apitz et al., 2014). Since mutations in Arabidopsis *HEMA* genes must be homozygous to show a phenotype (and a homozygous *hema2* mutant has no phenotype), we might not expect a heterozygous *hema* mutation in maize to show any phenotype.

The original motivation for the experiments included in Chapter 4 was to gain an understanding of why 313 transgenic phenotypes were so variable, thinking they were caused by *lhy2* overexpression. After discovering that the location of the transgene itself may disrupt the function of an important gene, it was not longer clear that 313 phenotypes could be explained by *lhy2* overexpression. The Results section explains the rationale for each set of experiments based on knowledge available at the time and the Discussion interprets these findings from both historic and current perspectives.

RESULTS

lhy2-OX 313 has dramatically decreased growth

As a reminder from Chapter 3, 313 transgenics were much shorter than any other *lhy*-OX line (Figure 1A). 313 also had a growth phenotype that never presented in other lines, namely pronounced interveinal chlorosis (Figure 1B). The 313 interveinal chlorosis phenotype was quantified by measuring leaf chlorophyll content. Chlorosis in 313 lines consistently presented as a decrease in both chlorophyll A, chlorophyll B, and total chlorophyll (Figure 1C). It is important to emphasize that the phenotypes for the *lhy2*-OX 313 line were present in plants with the transgene in a hemizygous state, representing the first generation progeny of hemizygous transgenic plants and the B104 inbred. The transgene was maintained in this state to avoid silencing of *lhy2* expression.

lhy2-OX 313 height and chlorosis are environmentally responsive

In order to study 313 growth phenotypes, we grew a large number of plants in the field during the summer. This grow out was separated into two plantings, planted 1 week apart but otherwise treated identically. Unexpectedly, height data for this summer showed a huge difference between the two plantings – planting #2 had heights almost identical to WT (Figure 2B,D) and planting #1 was much shorter than WT as we saw in all previous grow outs (Figure 2A,C). This suggested that the 313 growth phenotypes were affected by environmental conditions, so we sought to characterize the effects of these conditions individually.

In this grow out *lhy2* overexpression in 313 was overall much weaker than measured previously, with expression below WT from ZT0-6 spanning the normal peak, then above WT from ZT9-21 (Figure 3B). Despite this weakened *lhy2* overexpression, 313 growth phenotypes were still present in planting #1 (Figure 2A,C). The only difference in *lhy* expression in transgenic plants between the two plantings was a brief decrease in both *lhy1* and *lhy2* expression at ZT3 when the WT *lhy* peak usually occurs; in both *lhy* paralogs this caused a slight peak in expression at ZT6, shifting peak *lhy* expression 3 hours later in the day (Figure 3A,B). In order to ask whether this small difference in *lhy* expression might matter biologically, we measured expression of 3 *lhy*-target genes, *toc1a*, *toc1b*, and *cab* (Figure 3C,D,E). Expression of *toc1* in both plantings was not decreased or phase-shifted (Figure 3C,D) as expected from *lhy2*-OX 313 in Chapter 3. Expression of *toc1a* and *cab* was lower for timepoints surrounding the peak in 313 transgenics in planting #2 than in planting #1 or WT in either planting (Figure 3C,E). However, this is the opposite of the expected relationship, as *lhy* represses *toc1* and *cab*, so we would expect their expression to increase in 313 planting #2 transgenics, because *lhy* expression is decreased.

In addition to this disappearance of the 313 transgenic phenotypes in planting #2 in Oxford field, another grow out in a field at UC Davis had an unusual absence of height (not measured) and chlorophyll phenotypes; in this Davis field growout, 313 transgenics had no less chlorophyll than non-transgenic siblings (Figure 4A). Therefore, some conditions in the Davis field caused the chlorosis and height phenotypes to dramatically lessen. The Davis field was hotter and sunnier, received more fertilizer, and had higher soil iron content than the Oxford field. Therefore, we varied these conditions and measured the effect on 313 height and leaf chlorophyll content.

lhy2-OX 313 growth increases with row cover application

In order to study the effect of temperature on 313 plant growth in the field, we used row cover application to increase the temperature around the plants in the first few weeks of growth. We planted 313 in the Oxford field during the summer again, this time in four different groups, all planted at the same time but with varying weeks protected by row cover: 0, 1, 2, and 3. We covered all the plants with netting to protect them from birds when they had no row cover without affecting light or temperature.

Overall, more weeks under row cover correlated with higher chlorophyll content, especially in weeks 5-7 (Figure 4B). This effect increased from 0 to 1 week, and from 1 to 2 weeks, but decreased from 2 to 3 weeks, indicating that 2 weeks is the optimal time under row cover to maximize chlorophyll content in 313 transgenics (Figure 4C). The increase in leaf chlorophyll content with reemay treatment still didn't bring 313 plants anywhere near WT chlorophyll levels in weeks 7-9, and all 313 plants were still shorter than WT (Figure 5A). However, final 313 transgenic plant height did increase with 1-3 weeks of row cover treatment compared to 0 weeks, while row cover did not affect WT plant height.

These two phenotypes, chlorosis and reduced plant height, could easily be related if chlorosis causes decreased growth – if 313 plants have less chlorophyll than WT then they have a lower photosynthetic capacity and less resources for growth. The height difference between 313 transgenic and non-transgenic siblings appeared later than the difference in chlorophyll content (about 68 days or 9 weeks, and 4-5 weeks after planting, respectively) (Figure 2A, 4B), which suggests it could be the cause. Therefore, we focused on finding the origin of the 313 interveinal chlorosis.

When comparing final plant height across all growouts, growing plants in the field with 1-3 weeks of row cover increased plant height compared to growing them in the field with no row cover, and growing plants in the greenhouse during the summer increased height a little more, but most plants were still much shorter than WT (Figure 5B). However, the two times average 313 transgenic plant height approached that of WT were when grown in the field with 3 weeks of row cover (Figure 5B) and in the Davis field (not measured). Temperature was higher with row cover, greenhouse, and Davis conditions, and light intensity was lower with row cover and greenhouse conditions, and higher in Davis. Based on the plant heights shown in Figure 5B, light intensity and/or temperature seemingly mattered and temperature was more consistent across conditions. However, while row cover and greenhouse conditions consistently increased transgenic plant height over field conditions without row cover, there must have been additional factors in the Davis and Oxford 2015 3 week row cover growouts that caused transgenic plant height to almost equal that of WT.

Though not included in Figure 5B, plants in the Davis field growout reverted to near non-transgenic sibling height (not measured), and the Davis field was hotter than any of the other conditions and also received more fertilizer. Additionally, it is possible that plants in planting #2 were in a part of the Oxford field with different nutrient availabilities than planting #1. Since we hypothesized that the 313 height

phenotype is caused by the interveinal chlorosis, and interveinal chlorosis is usually caused by a nutrient deficiency or imbalance, we next investigated the effects of nutrient availability on 313 interveinal chlorosis.

Phosphorous:Iron ratio is highest in chlorotic transgenic 313 plants

Interveinal chlorosis is usually associated with nutrient deficiencies in the micronutrients magnesium, manganese, sulfur, copper, iron, and boron, or nutrient imbalances in the ratio of phosphorous:iron or calcium:potassium. After seeing 313 interveinal chlorosis almost disappear in the Davis planting, we had a transgenic non-chlorotic control for nutrient analysis. we sampled six different 313 leaves, two from Oxford with strong interveinal chlorosis, and one from Davis with weak interveinal chlorosis, and matching non-transgenic controls for all.

All three transgenic samples consistently had more phosphorous and potassium and less calcium than non-transgenic controls, regardless of chlorosis (Table 1). For both transgenics and non-transgenics, there was 3-4x more iron in the Davis sample than in either Oxford sample (Table 1). All this affected the P:Fe and Ca:K ratios, which can cause interveinal chlorosis if they're too high or low, respectively. Ca:K was 2-10x lower in transgenics, and was closest to non-transgenics in the least chlorotic sample (Table 1). In the two chlorotic leaf samples, P:Fe was higher than non-transgenics, but close to non-transgenic in the less chlorotic leaf sample (Table 1). Based on this data, either P:Fe or Ca:K could be responsible for chlorosis, as both are closest to non-transgenic controls in the least chlorotic sample. In order to more clearly see the effects of row cover and fertilizer on interveinal chlorosis and nutrient content, we varied each independently.

Row cover decreases iron accumulation

As discussed above, 313 transgenic height and chlorophyll content were higher in plants with 1-3 weeks of row cover compared to 0 weeks (height is highest with three weeks, chlorophyll with two). We used tissue from this same experiment to measure leaf nutrient content and see if it correlated with row cover treatment.

Regardless of genotype, 0 weeks of row cover caused significantly more iron accumulation than 3 weeks (Table 2). In non-transgenics, both boron and phosphorous:iron were affected by reemay; in transgenics, no nutrient accumulation was affected by reemay (Table 2).

Transgenics generally accumulated more phosphorous than non-transgenics, but phosphorous accumulation was unaffected by reemay in both genotypes; transgenics also accumulated more magnesium and nitrogen than non-transgenics, and had a higher P:Fe (Table 2). This repeats our finding from the first leaf nutrient content analysis, which showed that transgenics had consistently higher P:Fe ratios than non-transgenics (Table 1). It also suggests that the increase in leaf chlorophyll content and plant height with row cover treatment was not through an increase in the P:Fe ratio.

Row cover effect on 313 transgenic height and chlorophyll is not through *lhy*

I also sampled plants from 0wk and 3wk row cover treatments and measured gene expression for *lhy* and its target *toc1*. Previously in Oxford 2015 planting #2

we saw a small change in *lhy* expression at ZT3 that correlated with height and chlorophyll increase. Here, we saw no changes in *lhy* or *toc1* expression in transgenic plants between row cover treatments, but we did see a change in *toc1* in non-transgenic plants (Figure 6). Non-transgenic 313 plants with 0wk row cover had 1.5-3x higher *toc1a* and *toc1b* expression than non-transgenic 313 plants with 3wk row cover, or transgenic plants with either treatment (Figure 6). Since *toc1* expression in transgenic plants matched non-transgenic plants with 3wk row cover (when the plants were closer to normal), but not with 0wk row cover (when the mutant phenotypes were more pronounced), *toc1* expression could be a mechanism for improved 313 transgenic phenotypes with row cover treatment. In this situation, *lhy2* overexpression could make 313 transgenic *toc1* expression less sensitive to the effects of row cover.

Iron-containing fertilizer application may decrease 313 transgenic chlorosis

Next we wanted to see if we could change leaf nutrient and chlorophyll content by applying different fertilizers to plants in the more controlled greenhouse environment. Group 1 received NPKCaMg 3 days/week and group 3 received NPKCaMgFe fertilizer 1 day/week. We measured leaf chlorophyll content once a week, then sampled ear leaf tissue for nutrient analysis in mature plants after silking.

In the beginning, leaf chlorophyll content was similar between 313 transgenic and non-transgenic siblings, but they diverged around 57 days after planting (Figure 7A). There were no significant differences between chlorophyll content in transgenics and non-transgenics at either the beginning or end of fertilizer treatment in any group (Figure 7B,C). At the end, 313 transgenic chlorophyll was highest with fertilizer 3 (NPKCaMgFe M), while non-transgenic chlorophyll was highest with fertilizer 1 (NPKCaMg MWF) (Figure 7C). This suggests that a component of fertilizer 3 is more important for transgenics than for non-transgenics, likely the iron.

Taking into account only significant differences in nutrient content (Student's t-test, $p<0.05$), there are a few interesting patterns. Non-transgenic plants had lower iron content when watered with iron-containing fertilizer (group 3) than without (group 1); transgenics had higher iron content than non-transgenics in both fertilizer groups, but were not different between fertilizer groups (Table 3). This indicates that application of iron-containing fertilizer actually decreased non-transgenic iron accumulation, while transgenic iron accumulation was insensitive to iron fertilizer application; transgenics seem to constitutively accumulate more iron than non-transgenics.

Although Fe in transgenics was not significantly different between fertilizer groups, group 3 transgenics had a lower P:Fe ratio than group 1 transgenics, and were closest to the non-transgenic ratios in any group (Table 3). The Ca:K ratio was unchanged in transgenics between groups 1 and 3, and was therefore unlikely responsible (Table 3). Therefore, higher P:Fe ratio in transgenic plants could cause interveinal chlorosis, and application of Fe-containing fertilizer could decrease the P:Fe ratio in transgenics and make them less chlorotic. Accordingly, it is possible

that application of iron-containing fertilizer in Davis, or greater availability of iron in the soil, could decrease chlorosis.

In terms of other nutrients, transgenics overall accumulated less calcium than non-transgenics, but more phosphorous, potassium, zinc, copper, iron, and nitrogen. Accumulation of some nutrients in transgenic plants was unaffected by the differences between the two fertilizer regimes, but was affected by fertilizer in non-transgenic plants: magnesium, copper, iron, boron, nitrogen; while calcium and magnesium were affected in transgenics but not non-transgenics. Therefore, transgenics may constitutively over-accumulate copper, iron, and nitrogen whereas they over-accumulate phosphorous and zinc but do respond to external availability.

lhy2-OX 313 transgene insertion is in a gene encoding a putative glytamyl-tRNA reductase

One possibility was that the unique phenotypes of the *lhy2-OX* 313 line resulted from the transgene interrupting a gene important from metabolism and growth. Therefore, we used hiTAIL-PCR to amplify the genomic sequence adjacent to the transgene position. The 357-nucleotide sequence recovered from hiTAIL-PCR (Data File 1) was used as a query sequence for BLAST against the Zm-B73-REFERENCE-GRAMENE-4.0 genome (Jiao et al., 2017). A BLAST job at MaizeGDB (Portwood et al., 2018) returned three high quality alignments: hit #1 covered nucleotides 1-356 of the query sequence (99.4% ID, e-value 0), hit #2 covered nucleotides 1-358 (84% ID, e-value 4e-87), and hit #3 covered nucleotides 1-177 and 242-356 (92% ID, e-value 3e-84). The remaining alignments covered only 84-110bp of the query sequence with 76-88% identity, and corresponded to either no gene model or uncharacterized genes. Hit #1 was 100% identical to a contiguous nucleotide sequence when aligned with BLAST against the B104 pseudomolecule assembly (Zm-B104 ISU_USDA 0.1 assembly). This was consistent with the 313 transgenic line having been generated from the B104 inbred. The gene identified by hit #1 was Zm00001d032532 (assembly version Zm-B73-REFERENCE-GRAMENE-4.0; AC213521.3_FG005 in assembly B73 RefGen_v3), which is annotated as "glutamyl-tRNA reductase 1 chloroplastic". The sequence bordering the 313 transgene insertion matched the second exon, second intron, and third exon (Figure 8). The gene identified by hit #2 was Zm00001d026405 (GRMZM2G177412 in assembly B73 RefGen_v3), which is annotated as "glutamyl-tRNA reductase1". The third hit was Zm00001d013915 (GRMZM2G107402 in assembly B73 RefGen_v3) and this gene is not annotated.

Although Zm00001d032532 is annotated as a chloroplastic glutamyl-tRNA reductase 1, its expression is not specific to photosynthetic tissues according to RNA-seq data from a gene expression atlas made to represent maize development (Nuechterlein et al., 2016). Instead, this gene is broadly expressed at similar levels in many tissues (Figure 9), including tissues unlikely to be exposed to light like ear primordia and roots. This expression pattern is unlike the photosynthetic tissue-specific expression of Arabidopsis *HEMA1* (Tanaka et al., 1996). Expression of Zm00001d013915 (hit #3) largely mirrors that of Zm00001d032532 in the same RNA-seq dataset (Figure 9). This may be a reflection of these two genes being syntenic paralogs (Schnable et al., 2011). On the other hand, Zm00001d026405 is

expressed mostly in leaves (Figure 9), and its expression is over 100x higher than Zm00001d032532 and Zm00001d013915.

DISCUSSION

The *lhy2*-OX line 313 had the highest *lhy2* expression of the over 100 transgenic lines we characterized and, therefore, it appeared to be an excellent tool to study the role of *lhy* in the maize circadian clock. This transgene was always maintained in a hemizygous state, representing the first generation progeny of hemizygous transgenic plants and the B104 inbred, to avoid silencing of *lhy2* expression from the transgene. Thus, the interveinal chlorosis and reduction in both height and internode elongation represent dominant effects of the transgene, which is consistent with an overexpression phenotype. However, the 313 transgene sits in a gene encoding a putative glutamyl-tRNA synthetase, an enzyme that is predicted to be responsible for the second common step of tetrapyrrole biosynthesis. Therefore, it is possible that the chlorosis and growth phenotypes of this line arise from a mutation in glutamyl-tRNA synthetase; however, it is difficult to rationalize the dominant nature of these phenotypes with a knockout of gene function. Nevertheless, it is also difficult to confidently attribute the observed phenotypes solely to the consequences of *lhy2* overexpression.

The 313 insertion is in a putative maize glutamyl t-RNA reductase Zm00001d032532 which is expressed ubiquitously and at much lower levels than another putative maize gluTR Zm00001d026405, which is expressed highly and mostly in mature leaves. Based on these expression patterns, the 313 insertion is in the putative maize gluTR corresponding to Arabidopsis *HEMA2* or *HEMA3* rather than *HEMA1* which is responsible for most chlorophyll biosynthesis. If this is the case, then the 313 insertion in Zm00001d032532 is unlikely responsible for 313 chlorosis, as it may not be responsible for chlorophyll synthesis in photosynthetic tissues. However, without confirmation of these maize genes' functions, this inference is inconclusive. In order to settle this question, future experiments could test different maize mutants in gluTR and see if they have chlorosis similar to that of 313 transgenics.

The most complete 313 transgenic phenotype reversions were in the Davis field and planting #2 in the 2015 Oxford field. The two common variables between these conditions were temperature and nutrient availability. Application of row cover and growing in the Davis field both increased temperature compared to the Oxford field with no row cover. Application of iron-containing fertilizer in the greenhouse and the more intensive fertilization scheme and/or naturally occurring higher iron levels in Davis field soil increased nutrient availability compared to the Oxford field which received less fertilizer. It is likely that the Davis reversion was due to either the extreme heat, which was not replicated in any other growout, or to the combination of increased heat and iron-containing fertilizer application. However, the Oxford reversion is harder to explain – plantings #1 and #2 were in the same field, planted only a week apart, treated with two vs. three weeks of row cover. In the controlled row cover experiment, neither two nor three weeks of row cover came close to fully rescuing the height or chlorophyll phenotype. It is hard to

imagine that the two field locations had drastically different iron availabilities, but we did not test the soil to rule this out.

One aspect of these three conditions that partially or fully rescued 313 transgenic phenotypes was inconsistent – row cover seemed to decrease iron accumulation in leaves but increased leaf chlorophyll content, while application of iron-containing fertilizer increased iron accumulation and chlorophyll content. One possible explanation is that samples for nutrient analysis were taken at 4 weeks from mostly non-chlorotic leaves for the row cover experiment, and from the non-chlorotic ear leaf after silking for the fertilizer experiment (all nutrient content tests summarized in Table 4). Nutrient accumulation could certainly differ between these different developmental stages and tissues sampled. When confined to one developmental stage and tissue, clearer patterns emerge. Among the samples of ear leaves after silking, the P:Fe ratio and the amount of Fe do correlate with chlorophyll content – Davis transgenics had the most iron and chlorophyll, then Fe-fertilized plants in the greenhouse, then non-Fe fertilized plants in the greenhouse.

The potential mechanisms for 313 interveinal chlorosis differ based on the putative cause: *lhy2* expression and/or a heterozygous insertional mutation in a gene coding for a putative glutamyl-tRNA reductase. If the mutation in glutamyl-tRNA reductase is contributing to chlorosis, then plants are presumably suffering from a shortage of gluTR that decreases chlorophyll synthesis, and may be exacerbated by environmental conditions. There is precedent for iron levels affecting activity of glutamyl-tRNA reductase in Arabidopsis: a large decrease in HEMA1 activity has been shown in iron-deficient plants (Rodríguez-Celma et al., 2013). Therefore, higher plant iron content in the Davis growout and in the greenhouse with iron-containing fertilizer might upregulate the otherwise-reduced activity of gluTR and increase leaf chlorophyll content. In Arabidopsis, HEMA1 is light inducible, so increased light intensity in Davis could also help upregulate gluTR in maize 313 and decrease chlorosis.

We measured *lhy* and *toc1* expression in two situations where 313 transgenic phenotypes varied: Oxford 2015 plantings #1 and #2, and Oxford 2016 0wk and 3wk row cover. We saw no consistent pattern in expression of either *lhy* or *toc1* that might explain the change in phenotypes between groups. We did see one change in each experiment that could theoretically affect these phenotypes. In Oxford planting #2, *lhy1* and *2* decreased in transgenics at ZT3 compared to planting #1. As discussed in Chapter 3, increased *lhy* expression at ZT0 and 3 in maize, and decreased *cca1* expression in Arabidopsis at ZT6 and 9, both seem to increase hybrid vigor. Therefore, we might expect a decrease in *lhy* expression at ZT3 to do the opposite, to make 313 growth phenotypes worse. This decrease in *lhy* expression also didn't have the expected effect on toc1 expression, because it decreased rather than increased toc1 expression. we saw no change in *lhy* expression between the row cover treatments, but we did see an increase in peak expression of *toc1a* and *b* at ZT12 in the 0wk non-transgenic sample. This does not provide an explanation for increased transgenic plant height and chlorophyll content in the 3wk row cover treatment. If increased *lhy2* expression is causing chlorosis in 313 transgenics, this may be through increased repression of circadian clock output pathways involved in chlorophyll synthesis. Consistently, chlorophyll

biosynthesis genes, including *HEMA1* in Arabidopsis, are circadian-regulated (Matsumoto et al., 2004).

Notably, the interveinal chlorosis in 313 transgenic plants appeared around the time of the juvenile to adult developmental transition. This transition is marked by morphological changes including the appearance of epidermal hairs and a waxy leaf coating. Usually leaves up to 5 and 6 express only juvenile traits, and leaves 8 and up express only adult traits (timing varies between inbreds), and those in between can express both juvenile and adult traits in different areas of the leaf (Moose and Sisco, 1994). Apart from morphological phenotypes, another study showed large transcriptional differences between juvenile and adult leaves, including upregulation of photosynthesis and stress response in juvenile leaves (Beydler et al., 2016).

Upregulated photosynthesis and stress response in juvenile leaves would fit with either cause of 313 chlorosis. Plant sensitivity to decreased photosynthesis (due to *lhy2*-OX or gluTR mutation) could be buffered by upregulated photosynthesis genes in juveniles, then emerge in adults with subsequent downregulation of these genes. It is notable that application of row cover for 1-3 weeks at the beginning of plant development caused increased height in adult plants months later compared to no row cover. This underscores the importance of early development in juvenile maize plants, and the establishment of photosynthesis. Finally, it is possible that increased *lhy2* expression and heterozygous mutation of glutamyl-tRNA reductase together cause 313 chlorosis. Perhaps the less extreme *lhy* overexpression in other *lhy*-OX lines was also buffered two functional copies of glutamyl-tRNA reductase, resulting in only slight reductions in chlorophyll.

MATERIALS AND METHODS

All experiments were performed as described in Chapter 3, with the exception of those below.

hiTAIL PCR for transgene insertion location

As previously described (Liu and Chen, 2007).

Nutrient content analysis

Whole leaves were sampled as specified. Leaves were dried, then sent to the Soil and Plant Tissue Nutrient Laboratory at the University of Massachussets for plant tissue analysis without nitrogen (https://ag.umass.edu/services/soil-plant-nutrient-testing-laboratory).

Chapter 4 Figures

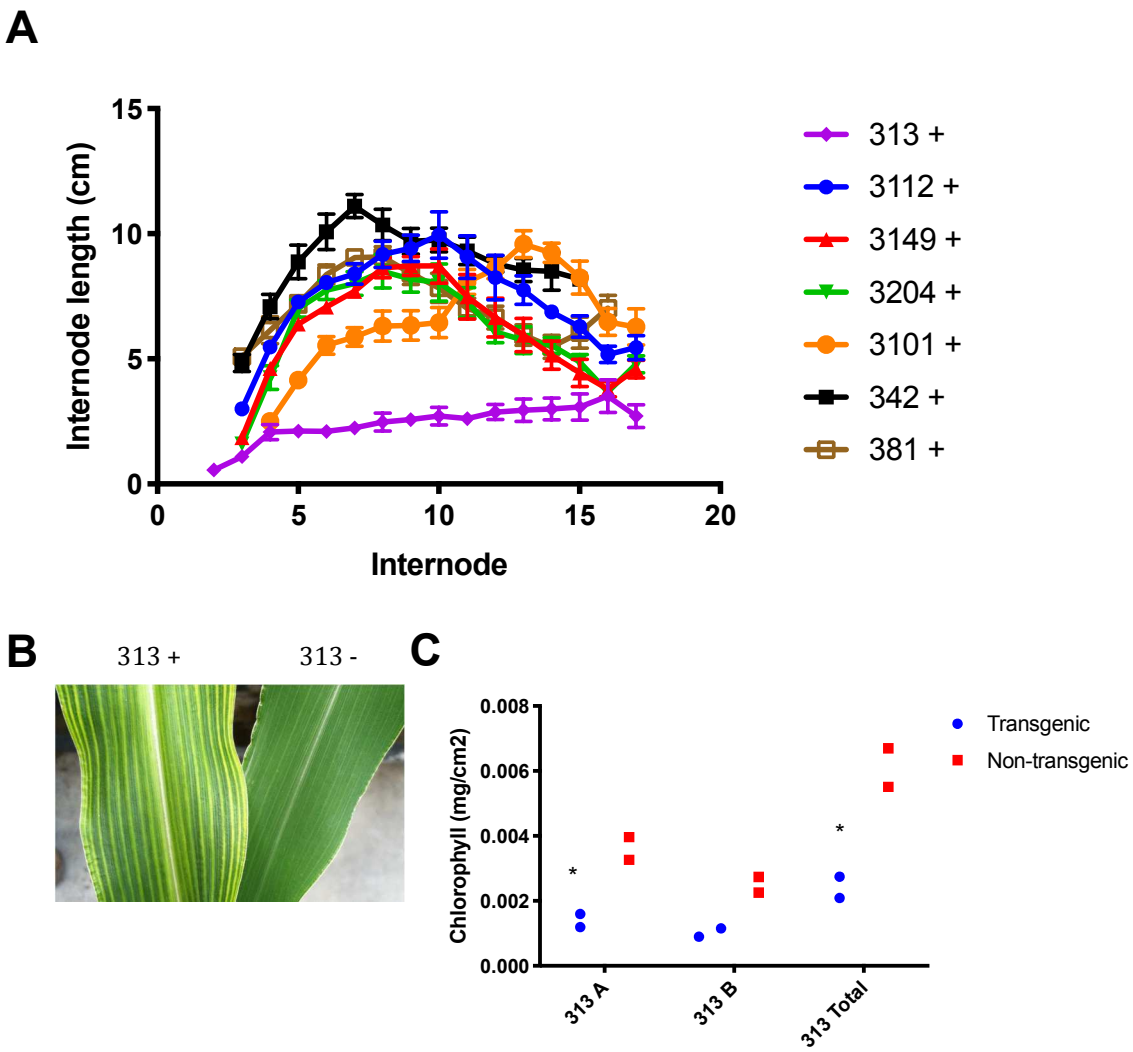


Figure 1. lhy2-OX 313 has decreased internode length and interveinal chlorosis.

A) Length of every internode for different *lhy2*-OX transgenic lines (independent transformation events) in T_1 generation. Each point is the mean of n=5 individual plants, error bars are ± SEM.

B) Interveinal chlorosis in 313 transgenic leaf (left) and normal coloration in non-transgenic sibling (right).

C) Chlorophyll A, chlorophyll B, and total chlorophyll in transgenic (+) and non-transgenic (-) 313 plants. Plants grown in Oxford summer 2015, generation T1. Genotype (transgenic or non-transgenic sibling), chlorophyll type, and interaction between the two are significant sources of variation by two-way ANOVA ($p<0.05$). Significant differences ($p<0.05$) between means for transgenic and non-transgenic plants by Sidak's multiple comparison test are shown with asterisks.

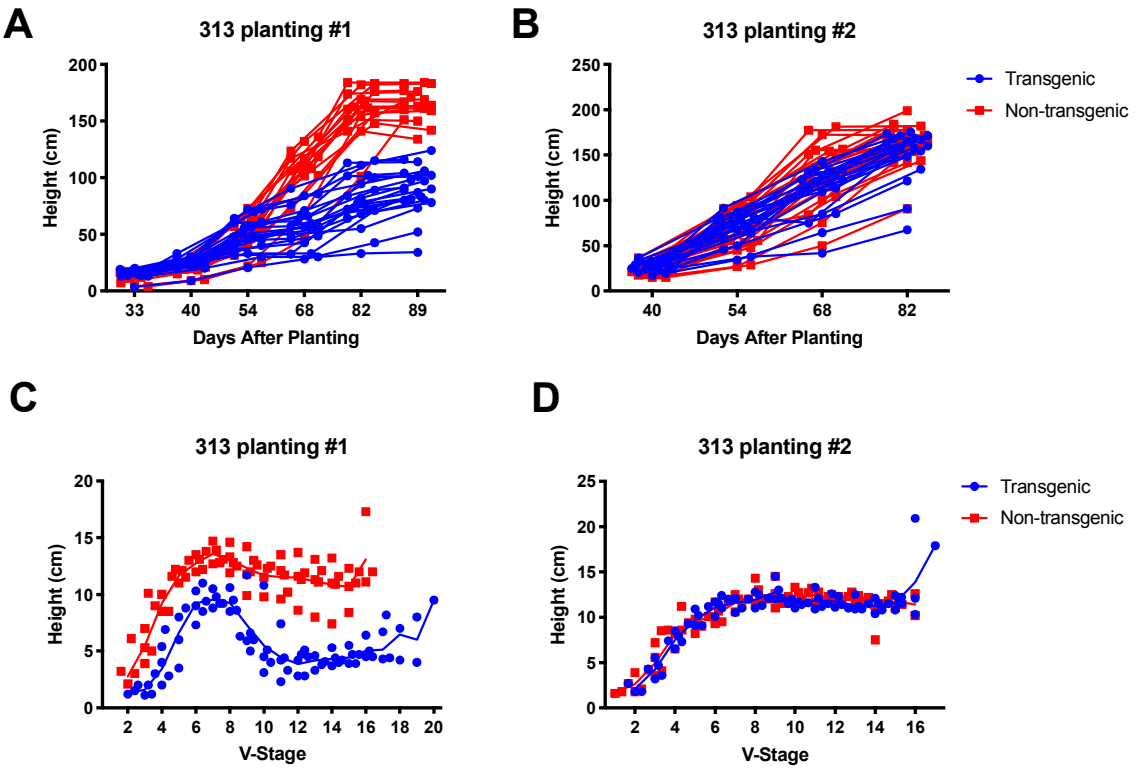


Figure 2. *lhy2*-OX 313 plants are usually short as in planting #1 but are rescued by environmental conditions in planting #2.

Height of 313 transgenic and non-transgenic sibling plants in plantings #1 (A,C) and #2 (B,D), in different locations in Oxford field summer 2015 and planted 1 week apart. T1 generation. Plant height graphed by days after planting (A,B) or plant developmental stage, "V-stage" (C,D).

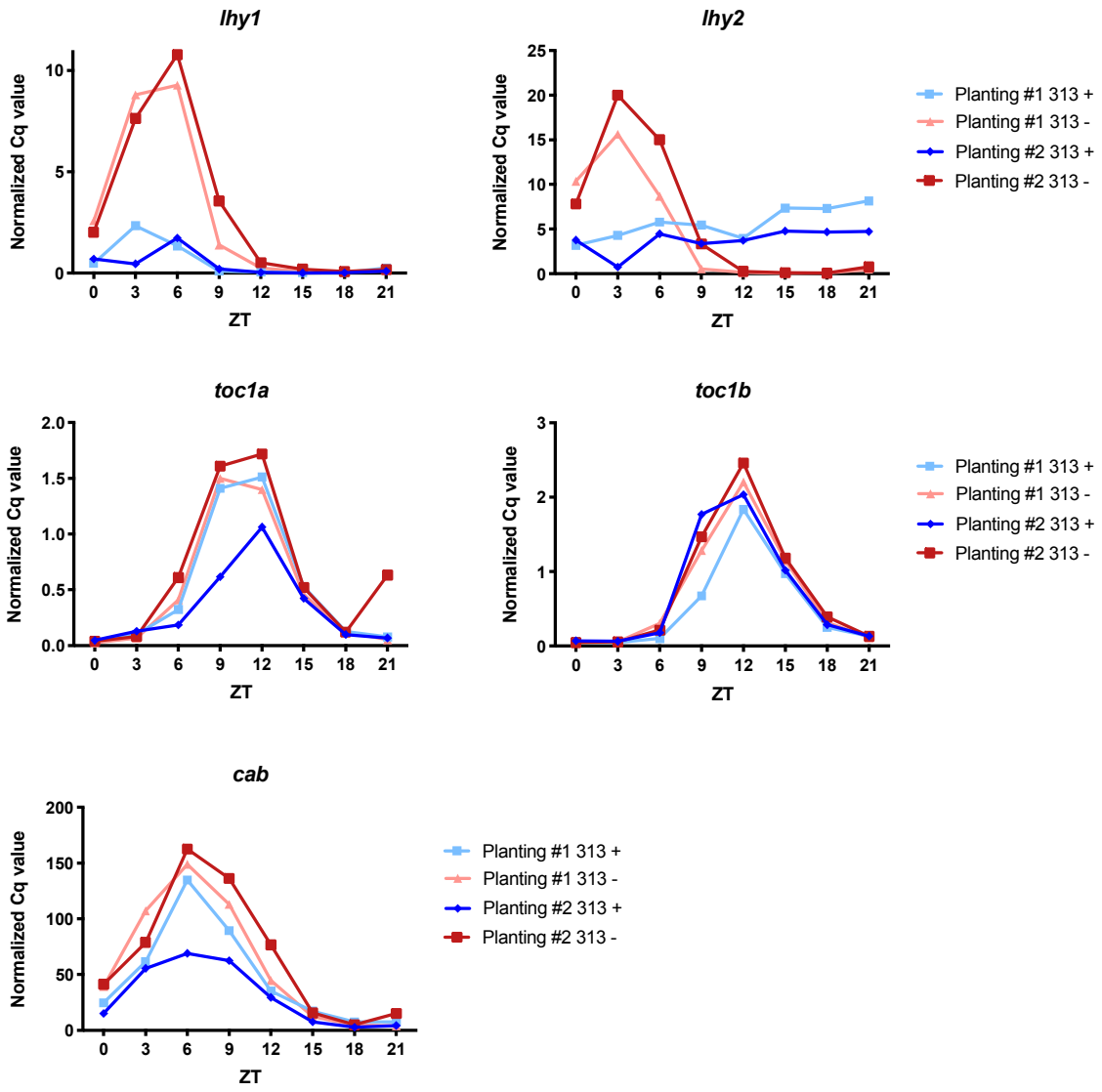


Figure 3. Gene expression is different in 313 transgenics in planting #2.

Expression of *lhy1*, *lhy2*, *toc1a*, *toc1b*, and *cab* genes in 313 transgenic (313+) and non-trangenic (313-) leaves from plantings #1 and #2 as indicated. These plantings were at different locations in the Oxford field and were planted 1 week apart during summer 2015. T1 generation. Tissue sampled from youngest fully-expanded leaf V7/V8 every three hours for a full day starting at dawn, 2 leaf punches per plant and 5 plants per genotype per sample. Gene expression measured by qRT-PCR, with values normalized to geometric mean of two reference genes, average of two technical replicates. + symbol for transgenic plants, - for non-transgenic siblings.

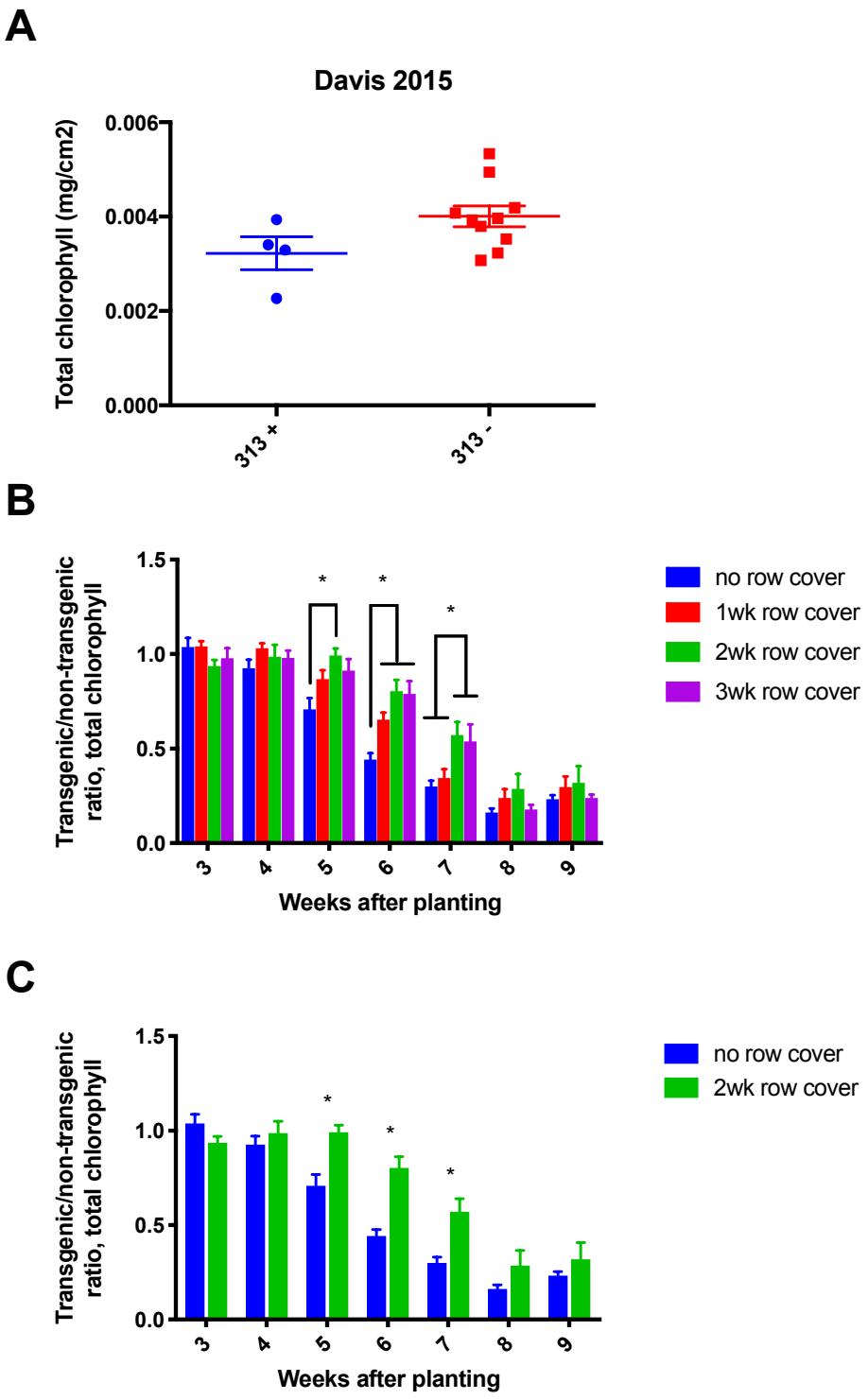


Figure 4. Environmental conditions and row cover affect 313 leaf chlorophyll content.

A) Total leaf chlorophyll content for plants grown in the Davis field and **B,C)** total chlorophyll for transgenic plants divided by total chlorophyll for non-transgenic sibling contols for plants grown in the Oxford field summer 2016 with 0-3 weeks row cover; T3 generation. Labels are + for transgenic plants and – for non-transgenic. All graphs show mean **A)** ±SEM or **B,C)** +SEM.

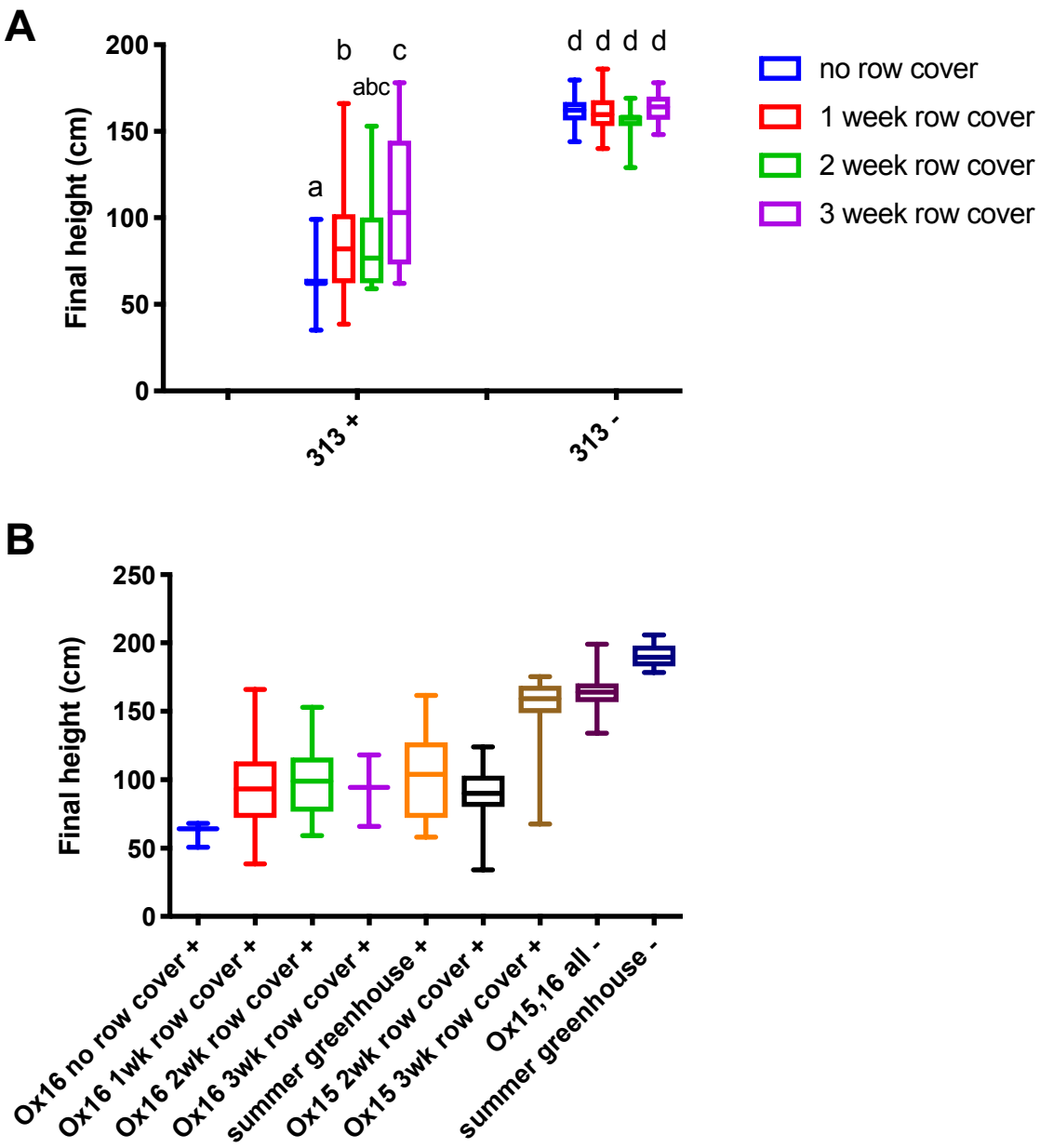


Figure 5. Environmental conditions and row cover affect 313 transgenic plant height.

A, B) Labels are + for transgenic plants and – for non-transgenic. Box plots are line for mean, box for SD, and error bars for min and max.
A) Final plant height from Oxford 2016 row cover experiment. Significant differences shown with letters above box plots: columns with different letters are significantly different by Sidak's multiple comparison test ($p<0.05$), columns with the same letter are not. **B)** Significant differences not shown.

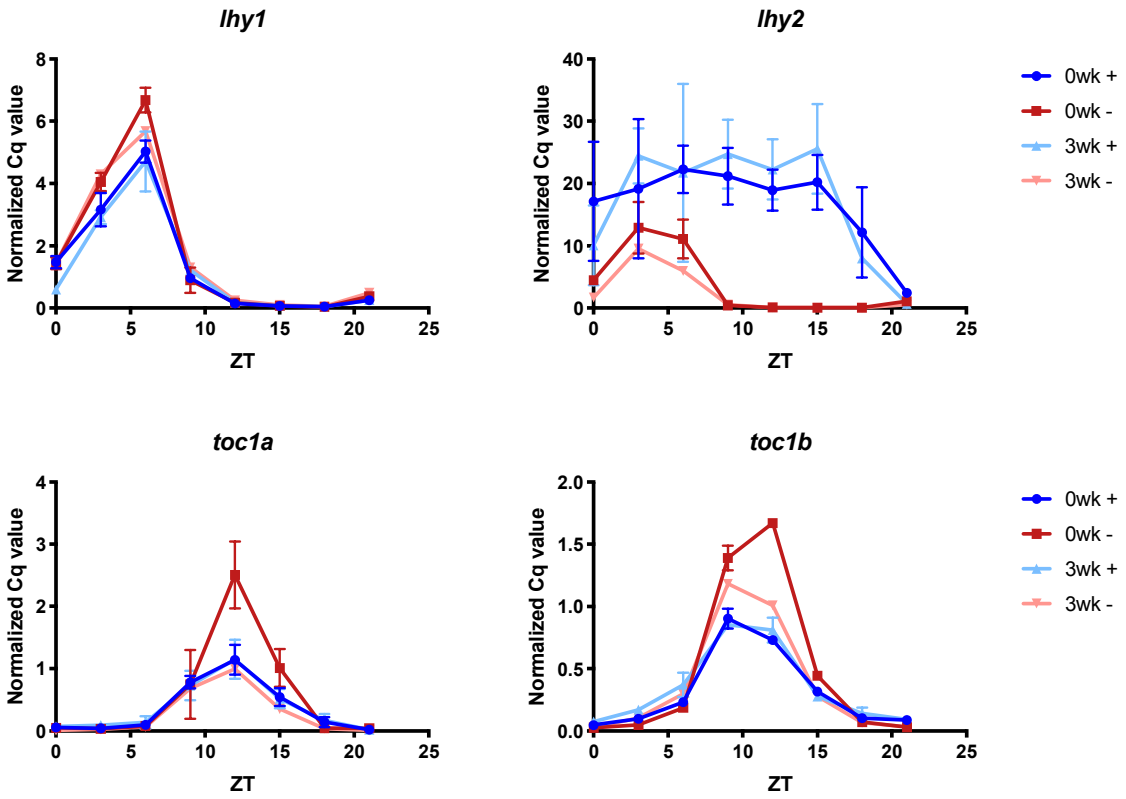


Figure 6. Row cover treatments alter gene expression in 313 transgenics and their non-transgenic siblings.

Expression of *lhy1*, *lhy2*, *toc1a*, and *toc1b* in leaves of the indicated genotypes determined by qRT-PCR. Points are mean with error bars ±SEM of two experiments with two technical replicates each. Blue and + symbol for transgenic plants, red and - for non-transgenic siblings. qRT-PCR determined expression values normalized to the geometric mean of two reference genes. Each sample is pooled tissue from 2 hole punches each from 3 individual plants.

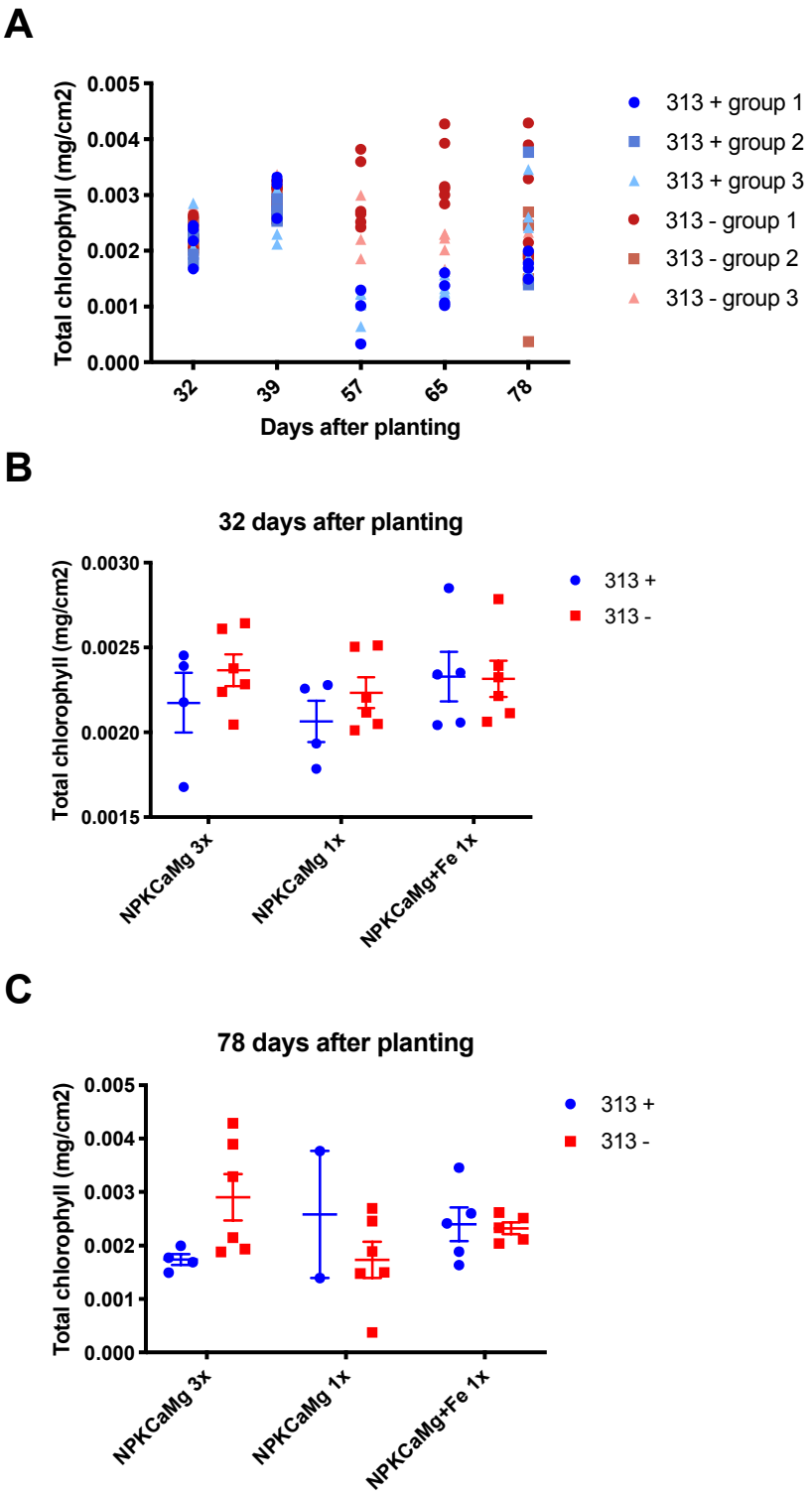


Figure 7. Fertilizer application can increase 313 transgenic leaf chlorophyll content.

A) Group 1 received NPKCalMg 3x/week, group 2 received NPKCaMg 1x/week, and group 3 received NPKCalMg+Fe 1x/week. Leaves were sampled for chlorophyll

measurement at time points during fertilizer application, with 32 days after planting being right before the first application, so **B)** is before fertilizer application and **C)** is after the last treatment.
A)-C) Labels are blue and + for transgenic plants and red and – for non-transgenic.
B), C) All data points shown with line for mean and error bars for ±SEM.

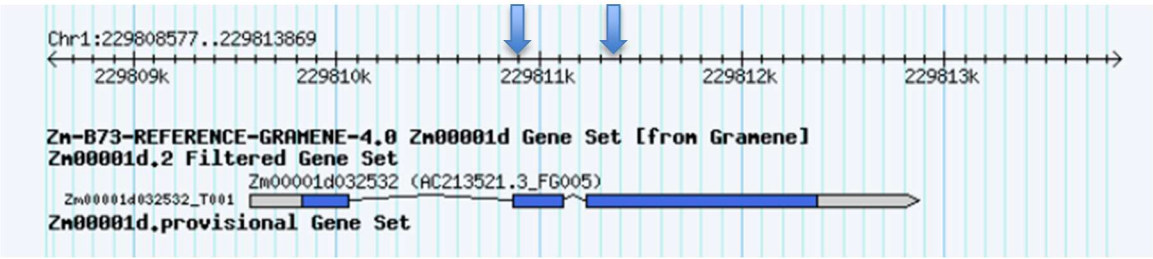
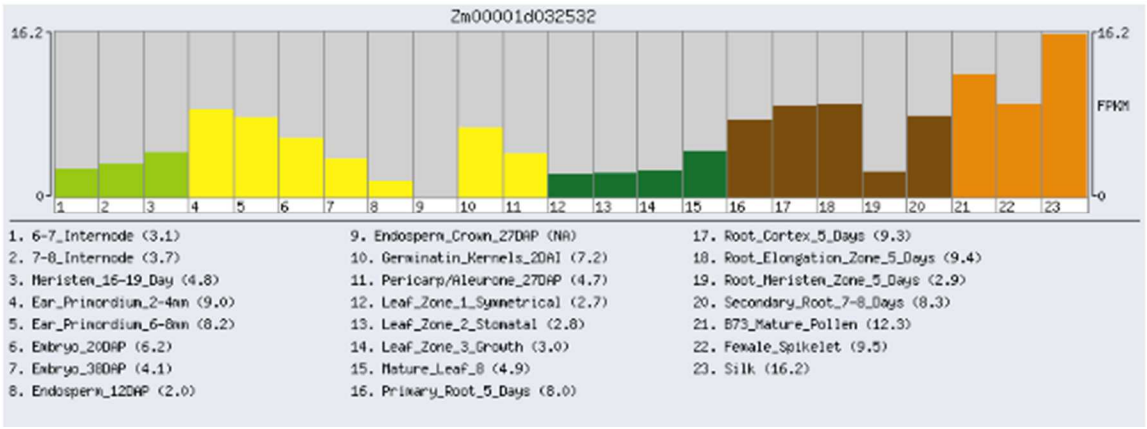


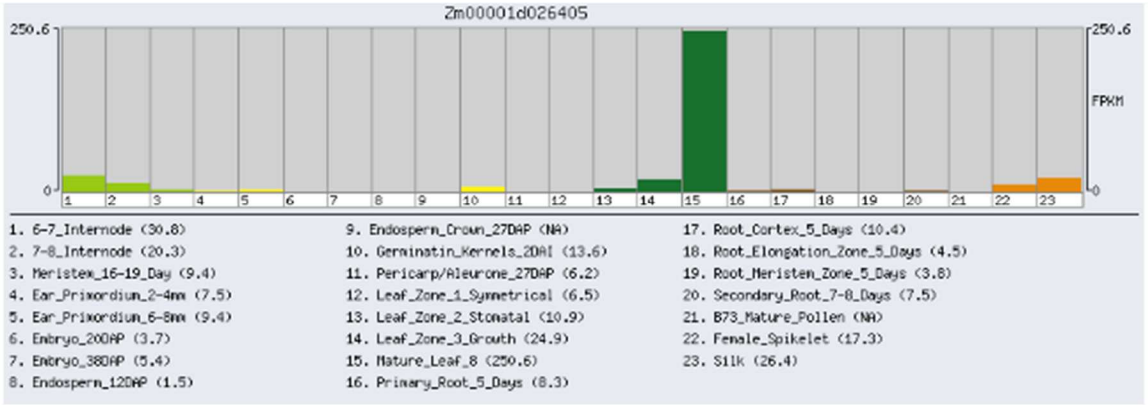
Figure 8. 313 transgene insertion is in maize *HEMA*-like gene Zm00001d032532.

Sequence bordering 313 transgene insertion (Data File 1) matches location 229810975 to 229811330 in B73 Reference Genome assembly B73RefGen_v4 (Jiao et al., 2017), marked with blue arrows. This is in Zm00001d032532, a putative maize *HEMA*-like gene.

A



B



C

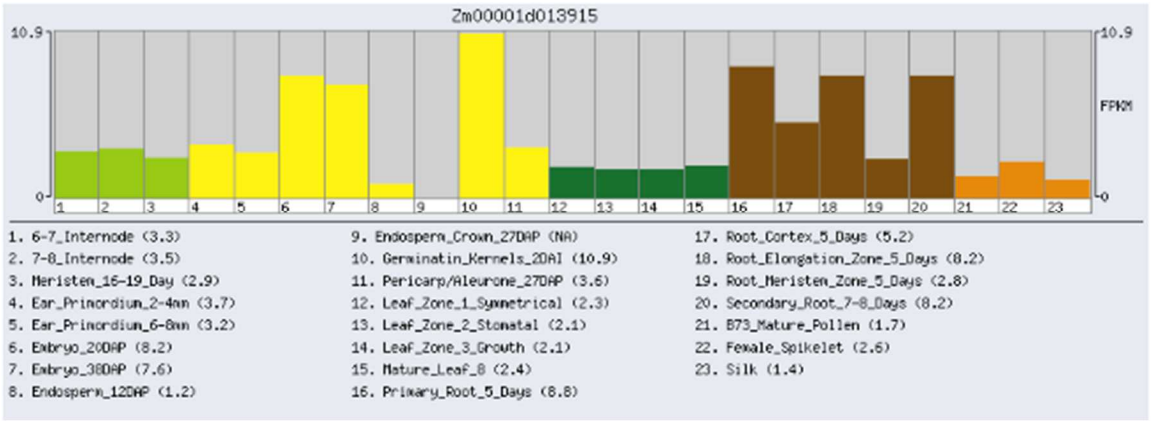


Figure 9. Gene expression patterns for putative maize *HEMA*-like genes.

Maize *HEMA*-like genes **A)** Zm00001d032532, **B)** Zm00001d026405, and **C)** Zm00001d013915. Graphs retrieved from a gene expression atlas made to represent maize development (Nuechterlein et al., 2016), available from MaizeGDB (Portwood et al., 2018).

Table 1. 313 transgenic and non-transgenic sibling leaf nutrient content from Oxford and Davis 2015 plants.

		Phosphorus (P)		Iron (Fe)		Potassium (K)		Calcium (Ca)		P/Fe		Ca/K	
	interveinal	0.3-0.5 %		21-251 ppm		1.7-2.5 %		0.2-1.0 %		N/A		N/A	
	chlorosis	+	-	+	-	+	-	+	-	+	-	+	-
2015 Oxford, ear leaf -4	strong	0.7	0.3	136	152	3.1	1.3	0.1	0.6	0.05	0.02	0.04	0.46
2015 Oxford, ear leaf -6	strong	0.6	0.3	125	90	2.9	1.5	0.2	0.3	0.04	0.03	0.07	0.23
2015 Davis, ear leaf	weak	0.6	0.4	387	399	2.6	1.6	0.3	0.3	0.01	0.01	0.09	0.17

Samples are from different leaves, either the leaf subtending the ear, or 4 or 6 leaves below this; all leaves sampled after silking. Values shown are a subset of nutrients measured. Acceptable range, as defined by testing center, is listed below each nutrient. Samples with nutrient levels above or below the acceptable range are highlighted in peach and blue, respectively.

Table 2. 313 transgenic and non-transgenic sibling leaf nutrient content from Oxford plantings with 0 or 3 weeks row cover.

Macro & micronutrients		Pct_P	Pct_K	Pct_Ca	Pct_Mg	Zn	Cu	Mn	Fe	B	Pct_N	P/Fe	Ca/K
"acceptable" range for maize		.25-.51	1.71-2.51	.21-1.01	.21-.61	20-71	N/A	20-151	21-251	N/A	N/A	N/A	N/A
3 wk row cover	1 +	0.51	4.34	0.76	0.30	118.70	10.60	53.80	99.00	6.30	4.70	0.005	0.18
	2 +	0.52	3.98	0.75	0.30	99.30	9.90	50.10	91.00	5.40	4.68	0.006	0.19
	3 +	0.50	4.14	0.98	0.34	187.50	12.00	75.40	102.00	3.90	4.90	0.005	0.24
	1 -	0.43	4.05	0.71	0.25	135.20	10.90	63.40	96.00	4.80	4.56	0.004	0.18
	2 -	0.45	3.98	0.89	0.29	137.80	15.70	69.80	99.00	5.00	4.66	0.005	0.22
	3 -	0.47	4.14	0.76	0.28	124.00	10.80	63.90	93.00	5.30	4.49	0.005	0.18
0 wk row cover	1 +	0.60	4.49	0.73	0.30	120.70	11.70	59.10	106.00	4.10	5.00	0.006	0.16
	2 +	0.54	4.51	0.87	0.33	143.30	12.80	76.90	131.00	4.90	4.90	0.004	0.19
	3 +	0.53	4.26	0.88	0.32	147.10	11.70	74.00	118.00	5.30	4.79	0.004	0.21
	1 -	0.44	4.15	0.83	0.29	141.10	11.50	83.30	125.00	3.30	4.73	0.004	0.20
	2 -	0.43	4.10	0.92	0.30	163.90	11.80	84.60	138.00	4.00	4.67	0.003	0.22
	3 -	0.43	4.16	0.68	0.25	123.40	10.60	70.30	111.00	4.20	4.71	0.004	0.16
Are + and - different?	+ vs. - 0wk	0.03	0.06	0.86	0.12	0.71	0.21	0.26	0.58	0.11	0.08	0.10	0.72
	+ vs. - 3wk	0.02	0.46	0.67	0.09	0.93	0.43	0.53	0.74	0.84	0.10	0.14	0.82
	0&3wk + vs. 0&3wk -	0.001	0.07	0.61	0.01	0.92	0.64	0.24	0.80	0.27	0.01	0.04	0.93
Are + or - affected by row cover?	+ 0wk vs. + 3wk	0.16	0.12	0.97	0.85	0.95	0.18	0.35	0.08	0.62	0.22	0.41	0.61
	- 0wk vs. - 3wk	0.28	0.22	0.81	0.75	0.47	0.55	0.08	0.06	0.03	0.10	0.01	0.94
Does row cover have an effect independent of genotype?	0wk +- vs. 3wk +-	0.66	0.09	0.86	0.77	0.67	0.97	0.05	0.003	0.09	0.11	0.07	0.72

All samples are two leaves pooled: fully-expanded leaves 7 and 8 counted with the cotyledon numbering 1, sampled at 4 weeks after planting. On 313 transgenic plants, leaf 7 looked the same as transgenic plants and leaf 8 had slight interveinal chlorosis on the outer edges.

Acceptable range, as defined by testing center, is listed below each nutrient. Samples with nutrient levels above or below the acceptable range are highlighted in peach and blue, respectively. The bottom three boxes show results of t-tests for various combinations of groups, to answer the questions on the left. Cells are highlighted green if $p<0.05$.

Table 3. 313 transgenic and non-transgenic sibling leaf nutrient content is affected by fertilizer.

Macro & micronutrients		Pct_P	Pct_K	Pct_Ca	Pct_Mg	Zn	Cu	Mn	Fe	B	Pct_N	P/Fe	Ca/K
"normal" range for maize		.25-.51	1.71-2.51	.21-1.01	.21-.61	20-71	NA	20-151	21-251	NA	NA	NA	NA
NPKCaMg MWF	313 group 1 -	0.30	1.85	0.32	0.16	18.40	6.30	24.10	89.00	62.30	2.37	0.0034	0.1730
	313 group 1 -	0.32	1.89	0.36	0.14	19.00	6.50	35.40	88.00	44.30	2.45	0.0036	0.1905
	313 group 1 -	0.38	2.21	0.46	0.13	21.90	8.50	87.20	92.00	21.80	2.61	0.0041	0.2081
NPKCaMgFe M	313 group 3 -	0.23	1.70	0.29	0.20	13.10	3.70	21.40	60.00	102.50	1.54	0.0038	0.1706
	313 group 3 -	0.25	1.80	0.25	0.20	14.30	3.50	16.00	68.00	100.00	1.54	0.0037	0.1389
	313 group 3 -	0.22	1.95	0.33	0.18	12.80	3.90	33.50	66.00	66.00	1.37	0.0033	0.1692
NPKCaMg MWF	313 group 1 +	0.56	3.25	0.23	0.18	31.50	9.20	30.20	93.00	47.40	3.21	0.0060	0.0708
	313 group 1 +	0.54	2.73	0.24	0.16	30.00	8.40	30.00	81.00	58.70	3.14	0.0067	0.0879
	313 group 1 +	0.54	3.13	0.24	0.17	33.40	8.60	34.80	90.00	33.00	3.21	0.0060	0.0767
NPKCaMgFe M	313 group 3 +	0.39	2.61	0.19	0.15	21.70	6.40	26.40	84.00	52.10	2.74	0.0046	0.0728
	313 group 3 +	0.39	2.34	0.22	0.18	19.20	7.40	24.50	81.00	60.00	2.89	0.0048	0.0940
	313 group 3 +	0.46	2.87	0.20	0.19	25.90	9.40	28.30	95.00	51.90	3.29	0.0048	0.0697
+ and - different?	group 1+ vs. 1-	0.009	0.007	0.074	0.075	0.001	0.135	0.469	0.697	0.812	0.005	0.001	0.002
	group 3+ vs. 3-	0.009	0.021	0.050	0.238	0.039	0.043	0.649	0.018	0.091	0.007	0.008	0.004
+ affected by fertilizer?	group 1+ vs. 3+	0.022	0.122	0.049	0.819	0.024	0.375	0.058	0.823	0.384	0.323	0.016	0.969
- affected by fertilizer?	group 1- vs. 3-	0.040	0.296	0.151	0.013	0.016	0.037	0.322	0.003	0.048	0.0005	0.735	0.100

Plants grown in the greenhouse and fertilized with either NPKCaMg every Monday, Wednesday, and Friday, or NPKCaMgFe every Monday. All samples are two leaves pooled: leaf subtending the ear (for top ear if multiple ears) and the leaf above. All plants sampled 1-2 weeks after silking.

Acceptable range, as defined by testing center, is listed below each nutrient. Samples with nutrient levels above or below the acceptable range are highlighted in peach and blue, respectively. The bottom three boxes show results of student's t-tests for various combinations of groups, to answer the questions on the left. Cells are highlighted green if $p<0.05$.

Table 4. Summary of Tables 1-3.

		Phosphorus (P)		Iron (Fe)		Potassium (K)		Calcium (Ca)		P/Fe		Ca/K	
	interveinal	0.3-0.5 %		21-251 ppm		1.7-2.5 %		0.2-1.0 %		N/A		N/A	
	chlorosis	+	-	+	-	+	-	+	-	+	-	+	-
2015 Oxford, ear leaf -4	strong	0.7	0.3	136	152	3.1	1.3	0.1	0.6	0.05	0.02	0.04	0.46
2015 Oxford, ear leaf -6	strong	0.6	0.3	125	90	2.9	1.5	0.2	0.3	0.04	0.03	0.07	0.23
2015 Davis, ear leaf	weak	0.6	0.4	387	399	2.6	1.6	0.3	0.3	0.01	0.01	0.09	0.17
2016 Oxford, 0wk row cover leaves 7,8	strong	0.6	0.4	118	125	4.4	4.1	0.8	0.8	0.005	0.003	0.19	0.20
2016 Oxford, 3wk row cover leaves 7,8	strong	0.51	0.45	97	96	4.2	4.1	0.8	0.8	0.01	0.005	0.20	0.19
2016 gh NPKCaMg MWF ear leaf	strong	0.6	0.3	88	90	3.0	2.0	0.2	0.4	0.01	0.004	0.08	0.19
2016 gh NPKCaMgFe M ear leaf	strong	0.4	0.2	87	65	2.6	1.8	0.2	0.3	0.005	0.004	0.08	0.16

Leaf nutrient content for 313 transgenic and non-transgenic sibling plants from Oxford 2015, Davis 2015, Oxford 2016 row cover experiment, Greenhouse 2016 fertilizer experiment.

Table 5. Primers used for experiments throughout Chapters 1-4.

Target	Primer name	Primer number	Sequence, 5'-3'
Arabidopsis primers			
IPP2	IPP2-QF	12	GTATGAGTTGCTTCTCCAGCAAAG
	IPP2-QR	13	GAGGATGGCTGCAACAAGTGT
PP2A	PP2A-QF_1741	2022	ACGTGGCCAAAATGATGCAA
	PP2A-QR_1803	2023	TCATGTTCTCCACAACCGCT
LHY	LHY_3'F	2426	ACGAAACAGGTAAGTGGCGACATT
	LHY_3'R	2427	TGGGAACATCTTGAACCGCGTT
LHY intron 1	LHY_I1F	2323	GGCTACTCTCAAGGGTATAACAGTT
	LHY_I1RqR	2522	CAGCAGCCAAACAGAGATCTTAGTAT
CCA1	CCA1_3'F	2432	TCTGTGTCTGACGAGGGTCGAATT
	CCA1_3'R	2433	ACTTTGCGGCAATACCTCTCTGG
LHYI1 lariat	AT1G01060int1p1	2901	TCTCAGCAGCCAAACAGAG
	AT1G01060int1p2	2902	GATACGAATCTGGTGATTGATTACTGA
CCA1I4 lariat	AT2G46830int4p1	2908	ACGATTACATCTCGTCATTTCTCCA
	AT2G46830int4p2	2909	TCCACAAACATGTACATAGATTCTTCA
RVE4I1 lariat	AT5G02840int1p1	2884	TGAGAAGAAACATTCATCAGCGA
	AT5G02840int1p2	2885	AGTGGTATTTTGCAGAAGTCGT
Maize primers			
cab1	qZm CAB1 F1	799	TCATGGTACAAACCGAAAACCA
	qZm CAB1 R1	800	GCCGCCGGTGACCTTTA
rnrp	qZmRnRP2F	839	CGAGACCAGGAGAACCAAAAG
	qZmRnRP2R	840	TGAGGAGTTTGTCCATAACCAC
luc	oNOSLUC_2238_F	1422	TGAGTAAAAATGTGTCAAATCGTG
	oNOSLUC_2538_R	1423	GGCAAGTCATAAAATGCATTAAAAA
toc1a	qZmTOC1chr53UTRF	1450	GACGAAGATGACGATCCAAC
	qZmTOC1chr53UTRR	1451	CAAGAGGACCTCTCCACTAC
toc1b	qZmTOC1Chr43UTRF	1452	CCCTTTTTGCTCACACACAG
	qZmTOC1Chr43UTRR	1453	AACCAGGGTAACAGCCACAT
lhy2	zmlhy2q_2F	1805	TGTGGAAGCGAAGGAAAAC
	zmlhy2q_2R	1806	ACTGCATTGCAAGATCTGAAG
lhy1	zmlhy1q_2F	1809	AGGACTTAAAGACCTTGCAATG
	zmlhy1q_2R	1810	TGCTTGGTGCTTAAATCAGAG

Data File 1. 357bp of sequence bordering 313 insertion, located using hiTAIL-PCR.

CCCCGTGCTATTCAAGAACTTACTAGTCTAAACCATATTGAAGAGGCTGCTGTTCTTAGT
ACTTGTAATAGAATGGAAATTTATGTGGTAGCCCTATCATGGAACCGAGGTATCAGAGA
AGTAGTTGACTGGATGTCAAAGGTGAGAACCAATCGATCAACTCTTTCTGCTCAGTCTTC
CTGCCGCAGTTTCTTGGCTGCTCAAAGAAATATTAGGAAAGATTCCTTAATGAGTGTTTC
TTTTTTGCGATTTCACAGAAAAGTGGTATTCCTGCTTCTGAGCTTAAGGAGCACCTATTC
ATGCTGCGTGACAGTGATGCTACACGCCATCTGTTCGAGGTATCAGCAGGGTTGGACA

REFERENCES

Ahmad, A., Dong, Y., Cao, X. (2011). Characterization of the PRMT gene family in rice reveals conservation of arginine methylation. PLoS One **6**:

Alabadí, D., Oyama, T., Yanovsky, M.J., Harmon, F.G., Más, P., Kay, S. a. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. Science (80-.). **293**:880–3.

Alabadí, D., Yanovsky, M.J., Más, P., Harmer, S.L., Kay, S. a. (2002). Critical role for CCA1 and LHY in maintaining circadian rhythmicity in Arabidopsis. Curr. Biol. **12**:757–761.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. Science (80-.). **301**:653–657.

Apitz, J., Schmied, J., Lehmann, M.J., Hedtke, B., Grimm, B. (2014). GluTR2 complements a hema1 mutant lacking glutamyl-tRNA reductase 1, but is differently regulated at the post-translational level. Plant Cell Physiol. **55**:645–657.

Arenas, J., Abelson, J. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. Proc. Natl. Acad. Sci. U. S. A. **94**:11798–802.

Bedford, M.T., Richard, S. (2005). Arginine Methylation: An Emerging Regulator of Protein Function. Mol. Cell **18**:263–272.

Bendix, C. (2015). The time has come: of GIGANTEA paralogs and grass circadian clocks. Dissertation. doi: 10.1017/CBO9781107415324.004

Beydler, B.D., Osadchuk, K., Cheng, C.-L., Manak, J.R., Irish, E.E. (2016). The juvenile phase of maize sees upregulation of stress-response genes and is extended by exogenous JA. Plant Physiol. **171**:pp.01707.2015.

Bitton, D.A., Rallis, C., Jeffares, D.C., Chen, Y.C., Codlin, S., Marguerat, S., Bähler, J. (2014). LaSSO , a strategy for genome-wide mapping of intronic lariats and branch-points using RNA-seq LaSSO , a strategy for genome-wide mapping of intronic lariats and branch-points using RNA-seq. 1169–1179.

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res. **24**:1774–1786.

Brenner, T.J., Guthrie, C. (2005). Genetic Analysis Reveals a Role for the C Terminus of the Saccharomyces cerevisiae GTPase Snu114 During Spliceosome Activation. **1080**:1063–1080.

Burgess, S.M., Guthrie, C. (1993). A Mechanism to Enhance mRNA Splicing Fidelity : The RNA-Dependent ATPase Prpl6 Governs Usage of a Discard Pathway for Aberrant Lariat Intermediates. Cell **73**:1377–1391.

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N.L., Lewis, S.E., Mungall, C.J., Basu, S., Chisholm, R.L., Dodson, R.J., et al. (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. **47**:D330–D338.

Cheng, D., Côté, J., Shaaban, S., Bedford, M.T. (2007). The Arginine Methyltransferase CARM1 Regulates the Coupling of Transcription and mRNA

Processing. Mol. Cell **25**:71–83.
Cheng, J., Zhang, Y., Li, Z., Wang, T., Zhang, X., Zheng, B. (2018). A lariat-derived circular RNA is required for plant development in Arabidopsis. Sci. China Life Sci. **61**:204–213.
Clontech. (2009). Yeast Protocols Handbook.
Collier, R., Dasgupta, K., Xing, Y.-P., Hernandez, B.T., Shao, M., Rohozinski, D., Kovak, E., Lin, J., de Oliveira, M.L.P., Stover, E., McCue, K.F., Harmon, F.G., Blechl, A., Thomson, J.G., Thilmony, R. (2017). Accurate measurement of transgene copy number in crop plants using droplet digital PCR. Plant J.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. 1188–1190.
Cui, X., Lu, F., Li, Y., Xue, Y., Kang, Y., Zhang, S., Qiu, Q., Cui, X., Zheng, S., Liu, B., Xu, X., Cao, X. (2013). Ubiquitin-Specific Proteases UBP12 and UBP13 Act in Circadian Clock and Photoperiodic Flowering Regulation in Arabidopsis. Plant Physiol. **162**:897–906.
Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., Pikaard, C.. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. 616–629.
Frisch, M., Thiemann, A., Fu, J., Schrag, T. a., Scholten, S., Melchinger, A.E. (2010). Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor. Appl. Genet. **120**:441–450.
Ganpudi, A.L., Schroeder, D.F. (2011). UV Damaged DNA Repair & Tolerance in Plants. Sel Top DNA Repair. doi: http://dx.doi.org/10.5772/57353
Green, R.M., Tobin, E.M. (1999). Loss of the circadian clock-associated protein 1 in Arabidopsis results in altered clock-regulated gene expression. Proc. Natl. Acad. Sci. U. S. A. **96**:4176–4179.
Gutiérrez, R. a, Stokes, T.L., Thum, K., Xu, X., Obertello, M., Katari, M.S., Tanurdzic, M., Dean, A., Nero, D.C., McClung, C.R., Coruzzi, G.M. (2008). Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. Proc. Natl. Acad. Sci. U. S. A. **105**:4939–4944.
Han, B., Park, H.K., Ching, T., Panneerselvam, J., Wang, H., Shen, Y., Zhang, J., Li, L., Che, R., Garmire, L., Fei, P. (2017). Human DBR1 modulates the recycling of snRNPs to affect alternative RNA splicing and contributes to the suppression of cancer development. 5382–5391.
Hernando, C.E., Sanchez, S.E., Mancini, E., Yanovsky, M.J. (2015). Genome wide comparative analysis of the effects of PRMT5 and PRMT4/CARM1 arginine methyltransferases on the Arabidopsis thaliana transcriptome. BMC Genomics **16**:192.
Hogg, R., Mcgrail, J.C., Keefe, R.T.O. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. Biochem. Soc. Transa **38**:1110–1115.
Hoskins, A., Moore, M.J. (2012). The spliceosome: a flexible, reversible macromolecular machine. **37**:179–188.
Huang, H., Alvarez, S., Bindbeutel, R., Shen, Z., Naldrett, M.J., Evans, B.S., Briggs,

S.P., Hicks, L.M., Kay, S.A., Nusinow, D.A., Elf, F., Lux, L.U.X.A. (2016). Identification of Evening Complex Associated Proteins in Arabidopsis by Affinity Purification and Mass Spectrometry * □. 201–217.

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B., Campbell, M.S., Stein, J.C., Wei, X., Chin, C.S., et al. (2017). Improved maize reference genome with single-molecule technologies. Nature **546**:524–527.

Karampelias, M., Neyt, P., Groeve, S. De, Aesaert, S., Coussens, G., Rol, J. (2016). ROTUNDA3 function in plant development by phosphatase 2A-mediated regulation of auxin transporter recycling. Proc. Natl. Acad. Sci. **113**:2768–2773.

Ko, D.K., Rohozinski, D., Song, Q., Taylor, S.H., Juenger, T.E., Harmon, F.G., Chen, Z.J. (2016). Temporal Shift of Circadian-Mediated Gene Expression and Carbon Fixation Contributes to Biomass Heterosis in Maize Hybrids. PLoS Genet. **12**:1–31.

Kuhn, P., Chumanov, R., Wang, Y., Ge, Y., Burgess, R.R., Xu, W. (2011). Automethylation of CARM1 allows coupling of transcription and mRNA splicing. **39**:2717–2726.

Lamond, A.I., Spector, D.L. (2003). Nuclear speckles: A model for nuclear organelles. Nat. Rev. Mol. Cell Biol. **4**:605–612.

Li, Z., Wang, S., Cheng, J., Su, C., Zhong, S., Liu, Q., Fang, Y., Yu, Y., Lv, H., Zheng, Y., Zheng, B. (2016). Intron Lariat RNA Inhibits MicroRNA Biogenesis by Sequestering the Dicing Complex in Arabidopsis. PLoS Genet. **12**:1–25.

Liu, Y.G., Chen, Y. (2007). High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. Biotechniques **43**:649–656.

Lou, P., Wu, J., Cheng, F., Cressman, L.G., Wang, X., McClung, C.R. (2012). Preferential Retention of Circadian Clock Genes during Diploidization following Whole Genome Triplication in Brassica rapa. Plant Cell **24**:2415–2426.

Luo, C., Cai, X.T., Du, J., Zhao, T.L., Wang, P.F., Zhao, P.X., Liu, R., Xie, Q., Cao, X.F., Xiang, C. Bin. (2016). PARAQUAT TOLERANCE3 Is an E3 Ligase That Switches off Activated Oxidative Response by Targeting Histone-Modifying PROTEIN METHYLTRANSFERASE4b. PLoS Genet. **12**:1–34.

Mancini, E., Iserte, J., Yanovsky, M., Chernomoretz, A. (2019). ASpli: Analysis of alternative splicing using RNA-Seq.

Manuelli, R., Seshadri, A. (2014). Frictionless Technology Diffusion : The Case of Tractors Author (s): Rodolfo E . Manuelli and Ananth Seshadri Source : The American Economic Review , Vol . 104 , No . 4 (APRIL 2014), pp . 1368-1391 Published by : American. Am. Econ. Rev. **104**:1368–1391.

Marquez, Y., Brown, J.W.S., Simpson, C., Barta, A., Kalyna, M. (2014). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res. **22**:1184–1195.

Marshall, C.M. (2017). Control of Alternative Splicing by SICKLE/WARP2 is Required for Adaptation of the Plant Circadian Clock to Cool Temperatures. University of California, Berkeley

Marshall, C.M., Tartaglio, V., Duarte, M., Harmon, F.G. (2016). The Arabidopsis sickle Mutant Exhibits Altered Circadian Clock Responses to Cool Temperatures and Temperature-Dependent Alternative Splicing. Plant Cell **28**:2560–2575.

Masaki, S., Yoshimoto, R., Kaida, D., Hata, A., Satoh, T., Ohno, M., Kataoka, N.

(2015). Identification of the specific interactors of the human lariat RNA debranching enzyme 1 protein. Int. J. Mol. Sci. **16**:3705–3721.

Matera, A. gregory, Wang, Z. (2014). A day in the life of the spliceosome. Nat Rev Mol Cell Biol. **4**:108–121.

Matos, D. a., Cole, B.J., Whitney, I.P., MacKinnon, K.J.M., Kay, S. a., Hazen, S.P. (2014). Daily changes in temperature, not the circadian clock, regulate growth rate in Brachypodium distachyon. PLoS One **9**:1–9.

Matsumoto, F., Obayashi, T., Sasaki-Sekimoto, Y., Ohta, H., Takamiya, K., Masuda, T. (2004). Gene Expression Profiling of the Tetrapyrrole Metabolic Pathway in Arabidopsis with a Mini-Array System. Plant Physiol. **135**:2379–2391.

McBride, A.E., Silver, P.A. (2001). State of the Arg: Protein Methylation at Arginine Comes of Age. Cell **106**:5–8.

Meyer, K., Koester, T., Staiger, D. (2015). Pre-mRNA Splicing in Plants: In Vivo Functions of RNA-Binding Proteins Implicated in the Splicing Process. Biomolecules **5**:1717–1740.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., Thomas, P.D. (2017). PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. **45**:D183–D189.

Mizoguchi, T., Wheatley, K., Hanzawa, Y., Wright, L., Mizoguchi, M., Song, H.R., Carré, I. a., Coupland, G. (2002). LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in Arabidopsis. Dev. Cell **2**:629–641.

Moll, R.H., Lonnquist, J.H., Fortuno, J. V, Johnson, E.C. (1965). The Relationship of Heterosis and Genetic Divergence in Maize. Genetics **52**:139–144.

Moose, S., Sisco, P. (1994). Glossy15 Controls the Epidermal Juvenile-to-Adult Phase Transition in Maize. Plant Cell **6**:1343–1355.

Nagai, S., Koide, M., Takahashi, S., Kikuta, A., Aono, M., Sasaki-Sekimoto, Y., Ohta, H., Takamiya, K., Masuda, T. (2007). Induction of isoforms of tetrapyrrole biosynthetic enzymes, AtHEMA2 and AtFC1, under stress conditions and their physiological functions in Arabidopsis. Plant Physiol. **144**:1039–51.

Nagel, D.H., Doherty, C.J., Pruneda-Paz, J.L., Schmitz, R.J., Ecker, J.R., Kay, S.A. (2015). Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **112**:E4802-10.

Ni, Z., Kim, E.-D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., Chen, Z.J. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature **457**:327–331.

Niu, L., Zhang, Y., Pei, Y., Liu, C., Cao, X. (2008). Redundant requirement for a pair of PROTEIN ARGININE METHYLTRANSFERASE4 homologs for the proper regulation of Arabidopsis flowering time. Plant Physiol. **148**:490–503.

Nuechterlein, G.L., Buitron, D., Bixby, J.L., Capranica, R.R., Whitaker, A., Wiebe, M.O., Hellewell, T.B., Rubolini, D., Romano, M., Martinelli, R., et al. (2016). Integration of omic networks in a developmental atlas of maize. Science (80-.). **353**:0–5.

Ohi, M.D., Ren, L., Wall, J.S., Gould, K.L., Walz, T. (2007). Structural

characterization of the fission yeast U5 . U2 / U6 spliceosome complex. **104**:2–7.

Oliveros, J.C. Venny. An interactive tool for comparing lists with Venn's diagrams.

Ooi, S.L., Dann, C., Nam, K., Leahy, D.J., Damha, M.J., Boeke, J.D. (2001). RNA lariat debranching enzyme. Methods Enzymol. doi: 10.1016/S0076-6879(01)42548-1

op den camp, R.G.., Przybyla, D., Ochsenbein, C., Laloi, C., Kim, C., Danon, A., Wagner, D., Hideg, É., Göbel, C., Feussner, I., Nater, M., Apel, K. (2013). Rapid Induction of Singlet Oxygen in Arabidopsis. Plant Cell **15**:2320–2332.

Poiré, R., Wiese-Klinkenberg, A., Parent, B., Mielewczik, M., Schurr, U., Tardieu, F., Walter, A. (2010). Diel time-courses of leaf growth in monocot and dicot species: Endogenous rhythms and temperature effects. J. Exp. Bot. **61**:1751–1759.

Portwood, J.I., Woodhouse, M., Cannon, E., Gardiner, J., Harper, L., Schaeffer, M., Walsh, J., Sen, T., Cho, K., Schott, D., et al. (2018). MaizeGDB 2018: the maize multi-genome genetics and genomics database. Nucleic Acids Res

Reddy, A.S.N., Marquez, Y., Kalyna, M., Barta, A. (2013). Complexity of the Alternative Splicing Landscape in Plants. Plant Cell **25**:3657–3683.

Richardson, A.D., Duigan, S.P., Berlyn, G.P. (2002). An evaluation of noninvasive methods to estimate foliar chlorophyll content. New Phytol. **153**:185–194.

Rodríguez-Celma, J., Pan, I.C., Li, W., Lan, P., Buckhout, T.J., Schmidt, W. (2013). The transcriptional response of Arabidopsis leaves to Fe deficiency. Front. Plant Sci. **4**:276.

Sahi, C., Lee, T., Inada, M., Pleiss, J.A., Craig, E.A. (2010). Cwc23, an Essential J Protein Critical for Pre-mRNA Splicing with a Dispensable J Domain. Mol. Cell. Biol. **30**:33–42.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carré, I. a., Coupland, G. (1998). The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell **93**:1219–1229.

Schnable, J.C., Springer, N.M., Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. U. S. A. **108**:4069–4074.

Schneider, C.A., Rasband, W.S., Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 671–675.

Shi, Y., Reddy, B., Manley, J.L. (2006). PP1/PP2A Phosphatases Are Required for the Second Step of Pre-mRNA Splicing and Target Specific snRNP Proteins. Mol. Cell **23**:819–829.

Talhouarne, G.J.S., Gall, J.G. (2018). Lariat intronic RNAs in the cytoplasm of vertebrate cells.

Tanaka, R., Tanaka, A. (2007). Tetrapyrrole Biosynthesis in Higher Plants. Annu. Rev. Plant Biol **58**:321–46.

Tanaka, R., Yoshida, K., Nakayashiki, T., Masuda, T., Tsuji, H., Inokuchi, H., Tanaka, a. (1996). Differential expression of two hemA mRNAs encoding glutamyl-tRNA reductase proteins in greening cucumber seedlings. Plant Physiol. **110**:1223–1230.

The Gene Ontology Consortium, Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., et al. (2011). Gene Ontology : tool for the unification of biology. Nat. Genet. **25**:25–29.

Wahl, M.C., Will, C.L., Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell **136**:701–718.

Wang, H., Hill, K., Perry, S.E. (2004). An Arabidopsis RNA Lariat Debranching Enzyme Is Essential for Embryogenesis. J. Biol. Chem. **279**:1468–1473.

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

Yoshimoto, R., Kataoka, N., Okawa, K., Ohno, M. (2009). Isolation and characterization of post-splicing lariat-intron complexes. Nucleic Acids Res. **37**:891–902.

Zhan, X., Wang, B., Li, H., Liu, R., Kalia, R.K., Zhu, J.-K.J.-K., Chinnusamy, V. (2012). Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. Proc. Natl. Acad. Sci. U. S. A. **109**:18198–203.