GENOMIC APPLICATIONS TO PLANT BIOLOGY

By

Genevieve Hoopes

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Cell and Molecular Biology – Doctor of Philosophy

2021

ABSTRACT

GENOMIC APPLICATIONS TO PLANT BIOLOGY

By

Genevieve Hoopes

The study of the total nuclear DNA content of an organism, i.e., the genome, is a relatively new field and has evolved as sequencing technology and its output has changed. A shift from model species to ecological and crop species occurred as sequencing costs decreased and the technology became more broadly accessible, enabling new discoveries in genome biology as increasingly diverse species and populations were profiled. Here, a genome assembly and several transcriptional studies in multiple non-model plant species provided new knowledge of molecular pathways and gene content. Over 157 Mb of the genome of the medicinal plant species *Calotropis gigantea* (L.) W.T.Aiton was sequenced, *de novo* assembled and annotated using Next Generation Sequencing technologies. The resulting assembly represents 92% of the genic space and provides a resource for discovery of the enzymes involved in biosynthesis of the anticancer metabolite, cardenolide. An updated gene expression atlas for 79 developmental maize (*Zea mays* L., 1753) tissues and five abiotic/biotic stress treatments was developed, revealing 4,154 organ-specific and 7,704 stress-induced differentially expressed (DE) genes. Presence-absence variants (PAVs) were enriched for organ-specific and stress-induced DE genes, tended to be lowly expressed, and had few co-expression network connections, suggesting that PAVs function in environmental adaptation and are on an evolutionary path to pseudogenization. The Maize Genomics Resource (http://maize.plantbiology.msu.edu/) was developed to view and data-mine these resources. Through profiling global gene expression over time in potato (*Solanum tuberosum* L.) leaf and tuber tissue, the first circadian rhythmic gene expression profiles of the below-ground

heterotrophic tuber tissue were generated. The tuber displayed a longer circadian period, a delayed phase, and a lower amplitude compared to leaf tissue. Over 500 genes were differentially phased between the leaf and tuber, and many carbohydrate metabolism enzymes are under both diurnal and circadian regulation, reflecting the importance of the circadian clock for tuber bulking. Most core circadian clock genes do not display circadian rhythmic gene expression in the leaf or tuber, yet robust transcriptional and gene expression circadian rhythms are present.

ACKNOWLEDGEMENTS

I am grateful for the support and advice my primary mentor Dr. Robin Buell has given me during my time at Michigan State University, providing training not only in the plant genomics field, but also in scientific communication, leadership development, career preparation, and more. Dr. Buell has continually challenged me to pursue excellence and rigor in my research, and I am exceedingly thankful for the significant role she has played in my development into an independent scientist. Drs. Eva Farré and Dave Douches have also significantly contributed to my development through their advice and training they have provided in circadian biology, molecular biology, genetics, and breeding. My other committee members Drs. Jiming Jiang and David Arnosti have provided valuable feedback on my scientific communication and other advice.

In addition to my mentors and guidance committee, members of both the Buell and Douches Labs have imparted significant knowledge and training in genomics and potato biology. The trainers and participants in the Plant Biotechnology for Health and Sustainability program have advanced my training in science communication, leadership development, and career preparation through the yearly symposia and other events. My course instructors and fellow classmates also passed on in depth knowledge and critical thinking skills.

I am grateful for the support and love my family and friends have shown me during my time in graduate school. My parents and siblings have always supported me in my goals and helped me in any way they could. I am also exceedingly thankful for the Skidmore family and many others at Red Cedar Church as they have shown me love and generosity far beyond expectation. Finally, I want to acknowledge the ever present help and comfort the LORD has given me, and the central role He has played in shaping me into who I am.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

bp	Base Pair
cDNA	Complementary Deoxyribonucleic Acid
DNA	Deoxyribonucleic Acid
LD/HC	Cycling Light/Dark Hot/Cold
LL	Constant Light
mRNA	Messenger Ribonucleic Acid
nt	Nucleotide
UTR	Untranslated Region

CHAPTER 1

INTRODUCTION

OVERVIEW

Plant biology encompasses biochemistry, molecular/cellular biology, physiology, ecology, evolution, systematics, genetics, and genomics. Holistically, these subdisciplines span the molecule to ecosystem of the Viridiplantae. Here, the field of genomics is expounded upon, with a brief history of the field of plant biology, followed by changes in bioinformatic approaches to adapt to emerging technology platforms, and finally applications to plant biology.

A HISTORY OF SEQUENCING TECHNOLOGIES

Genomic DNA sequencing:

The study of the complete DNA content of an organism arose as a new scientific field in the mid-1990's with the advent of automated sequencing technologies and advanced computing hardware and software. One of the initial DNA sequencing technologies employed chain termination with 2,3-dideoxynucleoside triphosphates (ddNTP), also known as Sanger sequencing, which was later automated by Applied Biosystems using a fluorescence-labeled ddNTPs (Giani *et al.*, 2020) (Figure 1.1). Whole genome shotgun (WGS) sequencing, in which random genome fragments are sequenced, was employed to generate the genome sequence of *Haemophilus influenzae* in 1995, the first non-viral free-living organism with a complete genome sequence (Fleischmann *et al.*, 1995); Sanger sequencing of plasmids were used in the *H. influenzae* genome project. The larger and more complex genome of the model plant *Arabidopsis thaliana* (L.) Heynh. was sequenced and assembled from bacterial artificial chromosome clones that spanned the genome (The Arabidopsis Genome Initiative, 2000). In 2002, rice (*Oryza sativa* L.) became the

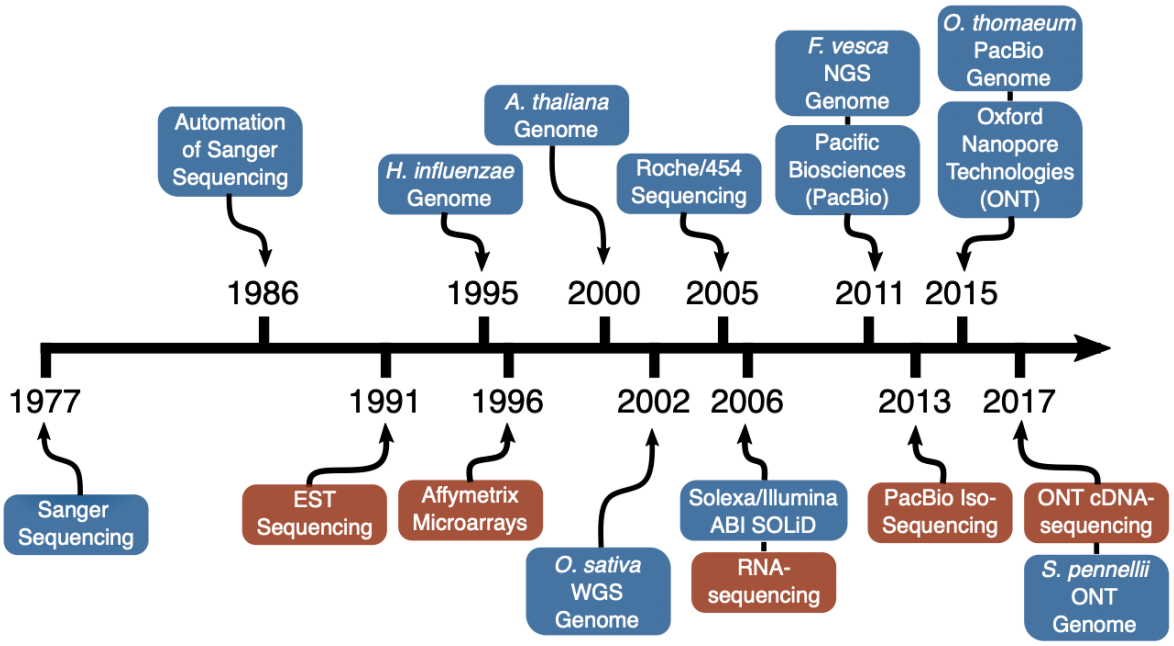

Figure 1.1: Timeline of Sequencing Technologies
The year each technology was released to the public is indicated in the timeline along with related milestones in the plant genomics field. DNA sequencing technologies and released genome assemblies are indicated in blue, while RNA sequencing technologies are indicated in brown. 'EST' represents 'Expressed Sequence Tags'.

first plant genome sequenced and assembled using WGS sequencing approaches (Goff *et al.*, 2002; Yu *et al.*, 2002); it was also sequenced using a bacterial artificial chromosome clone approach (International Rice Genome Sequencing Project and Saki, 2005) that is considered to be significantly more accurate than the WGS assembly released in 2002.

Advancements in sequencing chemistries, miniaturization, and computational technologies during the early 2000's enabled the development next generation sequencing (NGS) approaches (Giani *et al.*, 2020). 454 Life Sciences released the first NGS technology in 2005 in which emulsion PCR is followed by pyrosequencing-by-synthesis (SBS) in which a light signal is released and captured (Figure 1.1). Solexa, which was later acquired by Illumina, also developed a SBS technology that utilized bridge PCR whereas Applied Biosystems developed sequencing by oligonucleotide ligation and detection (SOLiD), which used emulsion PCR followed by sequencing-by-ligation (Glenn, 2011; Giani *et al.*, 2020) (Table 1.1). The 454 technology produced

NGS read lengths of up to 700 nt but had high errors in homopolymer stretches (Glenn, 2011). Initially, both the Illumina and SOLiD technologies produced similar read lengths, but through continual improvements in the Illumina platform, read lengths of up to 300 nt are now possible. Ultra-high throughput combined with improved accuracy and longer read lengths has propelled Illumina to the predominant NGS technology still widely in use today.

In 2007, grape (*Vitis vinifera* L.) was the first plant genome to incorporate NGS technologies with WGS Sanger sequencing (Velasco *et al.*, 2007) while in 2011, strawberry (*Fragaria vesca* L.) was the first plant genome assembled solely with NGS technology (Shulaev *et al.*, 2011; Michael and Jackson, 2013) (Figure 1.1). Genome assemblies based on NGS technologies are often highly fragmented due to the inability to resolve low-complexity and/or

Technology	Sequencing Method	Read Length	Read Accuracy	Strengths	Weaknesses
Sanger	Dideoxy Chain Termination	650 nt[a]	99.90%	High accuracy	Low throughput, expensive per read costs
Roche/ 454	Pyrosequencing	700 nt	99%	Long read lengths	Inaccurate in homopolymer stretches
Illumina	Sequencing-by-synthesis	50-300 nt	99.90%	Inexpensive, ultra-high throughput	Short read lengths
SOLiD	Sequencing-by-ligation	50-75 nt	99.99%	High accuracy	Technology abandoned
PacBio	SMRT	8k-30k nt[a]	99-85%	Long reads, high accuracy	High platform acquisition costs, limited platform availability
ONT	Nanopore	10k-30k nt[a]	98-91%	Longest reads, accessible platform	Low accuracy

Table 1.1: Strengths and Weakness of Sequencing Technologies
Sequencing technologies are listed with their sequencing method, read lengths, estimated read accuracies, strengths of the technology, and weaknesses of the technology. Read lengths are reported as number of nucleotides, where 'k' refers to nucleotides in the 1,000s. 'SOLiD' stands for 'Sequencing by Oligonucleotide Ligation and Detection'; 'PacBio' stands for 'Pacific Biosciences'; 'SMRT' stands for 'single molecule real time sequencing'; 'ONT' stands for 'Oxford Nanopore Technologies'. [a]average read lengths reported.

repetitive regions; several NGS-based approaches have been developed to aid in resolving these challenges in genome assembly. A few notable methods include 10x Genomics linked-read sequencing in which barcodes are used to assign reads from the same high molecular weight (HMW) DNA molecule (Zheng *et al.*, 2016). Another technology is that of high-throughput chromatin conformation capture (HiC) in which chromatin interactions are captured by cross-linking DNA in chromatin, digesting the DNA with restriction enzymes, adding adaptors, reversing the cross-links, constructing an Illumina-compatible library, and sequencing the library (Lieberman-aiden *et al.*, 2009).

Despite the use of these NGS-based approaches, chromosome-scale assembly still remains a challenge. Third generation sequencing (TGS) technologies bypass the limitations of short reads. In 2011, Pacific Biosciences (PacBio) released one of the first TGS platforms which utilized single molecule real-time sequencing through an SBS-like process within Zero-Mode Waveguide (ZMW) nanowell arrays (Giani *et al.*, 2020) (Figure 1.1). Initially, the error rate was significantly higher compared to Illumina-based sequencing, but through continual improvements in the technology and the use of circular consensus sequences in which a single DNA molecule is sequenced multiple times (termed 'HiFi sequencing'), the error rates have been lowered. However, due to sequencing length limitations of the DNA polymerase in the ZMWs, the error rate is correlated with DNA molecule length with longer DNA molecules having higher error rates (Amarasinghe *et al.*, 2020). A major limitation to the PacBio platform is the high acquisition costs and limited platform availability among genome sequencing centers (Table 1.1).

Oxford Nanopore Technologies (ONT) is an emerging TGS technology first released in 2015 (Figure 1.1). Instead of using an SBS approach, changes in a membrane current as the DNA molecule passes through a nanopore are used to detect the nucleotide sequence (Giani *et al.*, 2020);

it also can detect cytosine methylation states (Jain *et al.*, 2016). Similar to PacBio, read error rates are higher than Illumina sequences, but are lowering as the technology improves. A wide range of read lengths are possible with ONT, with the longest read generated thus far over four million nucleotides in length (https://nanoporetech.com/learn-more). One of the major strengths of the ONT technology is the high accessibility of their platform as it is essentially a free platform upon purchase of a flow cell and the ability to sequence using a laptop or smart phone (Table 1.1).

The first plant genomes assembled with PacBio and ONT sequencing methods were *Oropetium thomaeum* (L.f.) Trin., a desiccation tolerant grass (Vanburen *et al.*, 2015), and *Solanum pennellii* Correll, a wild tomato species (Schmidt *et al.*, 2017), respectively (Figure 1.1). Both of these assemblies highlight the increased contiguity through repetitive regions due to longer read lengths and highlight a promising future of chromosome-scale assemblies for plant species that are often plagued by long repetitive regions, heterozygosity, and polyploidy.

mRNA and cDNA Sequencing:

Sequencing of mRNAs is a well established gene discovery method and when mRNA sequences are aligned to a genome, they can reveal gene structure. Quantitative sequencing of mRNAs capture gene expression levels and is another dominant area of the genomics field. Starting in 1991, single pass Sanger sequencing of cDNA clones, known as expressed sequence tags (ESTs), were the predominant method for obtaining mRNA sequences, gene discovery, and characterizing gene structure (Adams *et al.*, 1991) (Figure 1.1). Due to the high cost of sequencing, microarrays were the preferred method for quantitation of gene expression. Fluorescence-based hybridization arrays developed in the 1990's and early 2000's used light-directed, spatially addressable chemical synthesis of probes (Pease *et al.*, 1994; Bumgarner, 2013), culminating in

the release in 1996 of the Affymetrix gene expression microarray technology (Blanchard *et al.*, 1996) (Figure 1.1). In the microarray technology, cDNAs are fluorescently labelled and hybridized to oligonucleotide probes affixed to the array surface, providing a quantitative measurement of the mRNA expression abundances (Blanchard *et al.*, 1996). While microarrays enable gene expression quantification simultaneously for large numbers of genes, the detection range is limited and expression abundances can only be estimated for genes included in the array.

With the advent of NGS technologies and their increased throughput coupled with reduced costs, both mRNA sequences and their relative expression abundances could be generated in a high-throughput manner. In 2006, the first mRNA-sequencing (mRNA-seq) study was conducted using 454 sequencing of cDNAs from a cancer cell line (Bainbridge *et al.*, 2006), closely followed by 454 sequencing of mRNAs in *Medicago truncatula* Gaertn. (Cheung *et al.*, 2006). Since that time, adaptations of the traditional mRNA-seq method have been made. In one method, the 3' end of a cDNA is selectively sequenced to quantitate gene expression abundances, termed 3' mRNA-seq or Tag-seq or Quant-seq, and has been used to lower sequencing costs as fewer reads are required for accurate estimation of expression levels (Moll *et al.*, 2014). Today, mRNA-seq remains the predominant method for globally profiling gene expression in organisms across the tree of life.

Use of TGS technologies to sequence cDNAs have enabled improved detection of alternative splice forms. Profiling the entire transcriptome via cDNA sequencing with PacBio technology, termed Iso-Seq, began in 2013 (Au *et al.*, 2013; Sharon *et al.*, 2013). In 2017, cDNA and direct mRNA sequencing were demonstrated with ONT, enabling characterization of both gene structure and base modifications (Garalde *et al.*, 2018). Since their release, RNA sequencing approaches with TGS technologies have been increasingly used for structural genome annotation

and are a promising application of state-of-the-art sequencing technologies to understanding RNA structure, expression, modification, and regulation.

EVOLUTION OF BIOINFORMATIC APPROACHES

Genome Assembly:

Bioinformatic approaches have evolved as sequencing technology and sequence data have changed. Initially, genome sequences were ordered through the use of physical maps or genome walking approaches (Bender *et al.*, 1983), as exemplified by the *A. thaliana* genome assembly (The Arabidopsis Genome Initiative, 2000). While these methods provided high quality chromosome-scale assemblies, they were slow, laborious, and expensive. WGS Sanger sequencing approaches required the use of a mathematical assembly model to order the reads and the overlap layout consensus (OLC) algorithm was used to solve this problem. In OLC, the overlaps between reads are first identified, followed by the generation of the overlap layout, and finished with inferring the consensus sequence (Li *et al.*, 2012). A popular OLC-based assembler was Celera Assembler (Myers *et al.*, 2000).

As NGS technologies produced higher sequencing depths and shorter read lengths, OLC methods became impractical. De Bruijn graph (DBG) algorithms became an alternative memory- and time-efficient approach to handle large amounts of NGS data where sequences of length k (termed k-mers) are used to form a DBG. The use of k-mers rather than the full read length enables increased computational efficiency, but comes at the loss of read contextual information and reduced ability to resolve repetitive regions (Li *et al.*, 2012). Commonly used DBG NGS assemblers include: SOAPdenovo (Luo *et al.*, 2012), ALLPATHS-LG (Butler *et al.*, 2008), and ABySS (Simpson *et al.*, 2009).

The high read error rates of TGS technologies complicate the assembly process and are a significant consideration for long read genome assemblers. Both OLC and DBG approaches are commonly used with differing success rates depending on the input data and genome complexity. In OLC methods, finding the overlap between error-prone reads presents challenges and as a consequence, exact short substring or seed alignments are used to allow mis-mismatches in the overlaps. Conversely, spurious k-mers often split the graph in DBG algorithms and k-mer filtering methods or approximate k-mer matches are used to overcome these challenges (Rizzi *et al.*, 2019). Popularly used OLC- and DBG-based long read assemblers are CANU (Koren *et al.*, 2017) and Flye (Kolmogorov *et al.*, 2019), respectively. New assembly methods are being developed for HiFi PacBio reads to capitalize on the reduced error rates; HiCanu (Nurk *et al.*, 2020) and hifiasm (Cheng *et al.*, 2021) are two OLC-based assemblers which have been released so far.

Transcriptome Assembly:

In addition to genome assembly, transcriptome assembly is a commonly used approach to study gene structure in lieu of a genome assembly; assembled transcripts are also powerful resources to annotate genome assemblies for gene content. Due to the single pass nature of Sanger-based ESTs, and the often truncated and redundant sequences, EST clustering and assembly became an important component of EST-based gene discovery and accurately characterizing gene structure. After quality filtering, the first step involves clustering ESTs based on pairwise-sequence similarity and then assembling the EST clusters using an OLC-approach (Nagaraj *et al.*, 2007). Several data analysis pipelines and databases were developed including the National Center for Biotechnology Information (NCBI) UniGene database (Sehuler, 1997), Sequence Tag Alignment

and Consensus Knowledge Base (STACK) (Miller *et al.*, 1999), and The Institute for Genomic Research (TIGR) Gene Indices (Lee *et al.*, 2005).

Assembling transcriptomes using NGS-based RNA-seq data evolved as assembly methods changed and more genome assemblies became available. Two main approaches were used: *de novo* and genome-guided transcript assembly (Martin and Wang, 2011). *De novo* assembly occurs much as it does for genome assembly where a DBG approach is applied to obtain assembled cDNAs. While this method is necessary when a genome assembly is not available, high cDNA coverage is needed to assemble full length transcripts and as a consequence, large computational resources are used during the assembly process. Genome-guided assembly starts by aligning the reads to a reference genome, followed by clustering overlapping reads and traversing the clustered reads to build the cDNA model. This method uses fewer computational resources and needs less cDNA coverage. However, the assembled cDNAs are highly dependent upon the mapping strategy employed and the genome sequence; as a consequence, errors can be propagated into the cDNA models when a high-quality genome assembly is not used. Combinations of these two methods have also been employed in an effort to get the best of both strategies, especially when confidence is lacking in a genome assembly (Martin and Wang, 2011). Commonly used *de novo* and genome-guided transcript assemblers are Trinity (Grabherr *et al.*, 2011) and Cufflinks (Trapnell *et al.*, 2010), respectively.

Identification of accurate splice isoforms are difficult to obtain with NGS-based RNA-seq data due to the use of short read lengths. While TGS-based sequencing aims to solve this problem through generation of full-length cDNA sequences, often these sequences are truncated, redundant, and error-prone. PacBio developed a proprietary pipeline to process their Iso-seq reads to obtain full length, error-corrected consensus transcripts. First, circular consensus sequences (CCS) are

built from the raw subreads, then full-length CCSs are identified by the presence of sequencing primers and clustered to obtain consensus sequences. The consensus sequences are then polished to remove sequencing errors (https://github.com/ben-lerch/IsoSeq-3.0). Full length cDNAs are identified solely on the basis of sequencing primers and can still be truncated if the input cDNA molecule was fragmented; as a consequence, full length cDNA sequences may not be detected for every gene. Emerging long read transcriptome assemblers generate true, full length cDNA models with reduced redundancy. StringTie2 is a popular genome-guided long read transcript assembler that is able to adjust the alignments of the reads to properly identify the intron-exon structure and has the functionality to merge transcripts identified from multiple data sources to generate a set of non-redundant isoforms (Kovaka *et al.*, 2019).

Gene Expression Normalization:

Quantification and normalization of gene expression abundances is a fundamental component of processing data from gene expression studies. Systematic experimental bias and technical variation present in gene expression profiling experiments are dependent on the platform being used and as a consequence, different normalization methods are employed for different platforms. In microarray experiments, differences in RNA quantities, fluorescent dye detection efficiencies, and hybridization efficiencies are among some of the systematic biases observed (Quackenbush, 2002). To correct for these biases, quantile (Bolstad *et al.*, 2003) and locally weighted scatterplot smoothing (LOWESS) (Berger *et al.*, 2004) normalizations were commonly used. In RNA-seq experiments, differences in sequencing depth, gene length, and GC content bias the data, and normalization methods are dependent upon library preparation methods and downstream applications (Dillies *et al.*, 2013; Abbas-Aghababazadeh *et al.*, 2018). Fragments Per

Kilobase per Million mapped reads (FPKM) (Trapnell *et al.*, 2010) and Transcripts Per Million (TPM) (Li *et al.*, 2010) are commonly used normalization methods to correct for both sequencing depth and gene length, but are not suitable for differential expression analyses as normalizing for gene length is known to bias results (Oshlack and Wakefield, 2009). A scaling factor in the DESeq2 (Love *et al.*, 2014) and EdgeR (Robinson *et al.*, 2010) programs is used to account for sequencing depth bias in differential expression analyses. When using library preparation methods in which the full gene length is not profiled, for example in 3' mRNA-seq, normalization methods not incorporating gene length are also required.

APPLICATIONS OF GENOMICS TO PLANT BIOLOGY

Availability of Genome Assemblies:

Since the first plant genome assembly in 2000 (The Arabidopsis Genome Sequencing Initiative, 2000), the genomics field has progressed from focusing on a single *A. thaliana* accession to over 1,000 (The 1001 Genomes Consortium, 2016), and from focusing on a diploid, homozygous model plant to polyploid, heterozygous crop and ecological species (https://www.plabipd.de/portal/web/guest/sequenced-plant-genomes). The volume of genomic sequence deposited to NCBI has shifted from Sanger sequence traces in the Trace Archive to NGS- and TGS-based sequences in the Sequence Read Archive (SRA) (Figure 1.2). Since the release of NGS technologies, the amount of sequence produced has increased exponentially (Figure 1.2), and over 500 diverse angiosperm species have a published genome assembly (https://www.plabipd.de/portal/web/guest/sequenced-plant-genomes), with numerous additional assemblies available for multiple accessions of a species. As sequencing costs have decreased and genome assemblies have become less cost prohibitive, applications of transcriptome assemblies

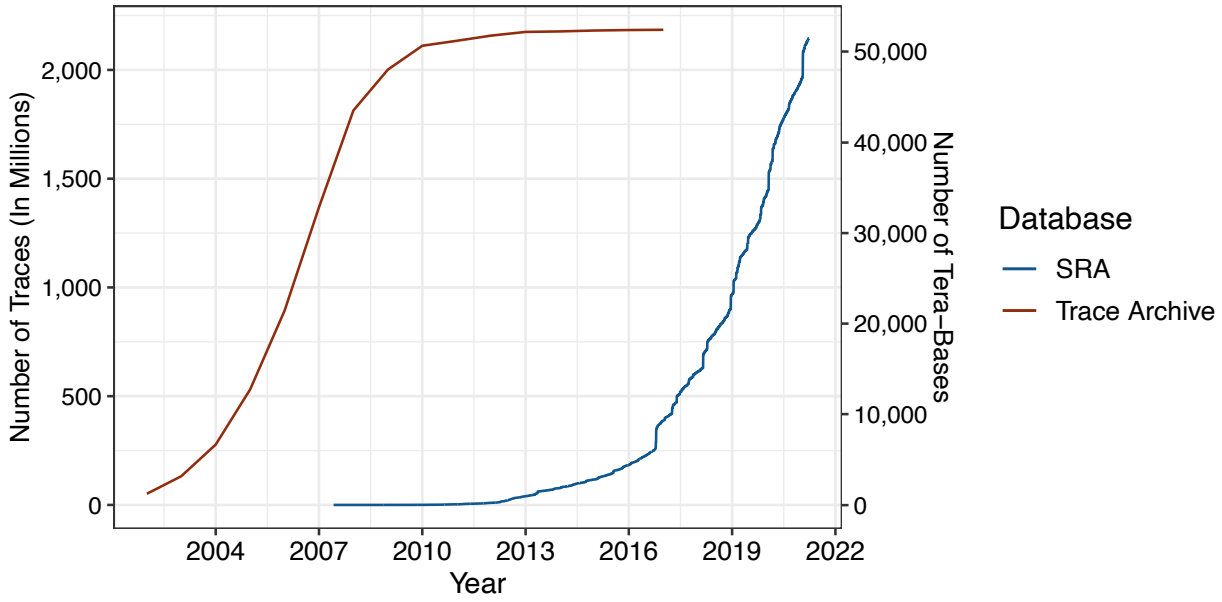

Figure 1.2: Frequency of Sequence Submissions to NCBI Databases
The number of Sanger sequencing traces submitted to the NCBI Trace Archive (left y-axis, red line) and the number of nucleotide bases present in the NCBI Sequence Read Archive (SRA) (right y-axis, blue line) over time.

have evolved from being an alternative to genome assemblies to primarily being used as evidence for genome annotation. Indeed, a new revolution is underway to generate accession-specific genome assemblies to remove the ambiguities often present when comparing accessions to a single reference genome assembly (Gan *et al.*, 2011) and to efficiently enable crop variety improvement in much the same way personalized human medicine seeks to do so for patients.

Gene Regulatory Networks:

The use of mRNA quantification methods has matured from identifying basal expression levels in tissues and organs to systems-level pathway analyses. Inferring gene function through a 'guilt-by-association' approach, which is known as 'co-expression' or 'co-regulation' (Rensink and Buell, 2005; Usadel *et al.*, 2009), involves clustering gene expression patterns across many samples (D'Haeseleer, 2005). Co-expression visualization in graph-based networks provides a powerful approach to assess gene function in relationship to other genes and to refine co-

expression patterns for specific pathways (Usadel *et al.*, 2009). Further integrating co-expression networks with other network types such as metabolic and protein networks provides support for gene relationships and a more complete view of gene regulatory networks, as demonstrated in tomato (Mounet *et al.*, 2009). Recently, increasingly refined patterns of gene expression are being explored through the use of single cell RNA-seq experiments, enabling localization of pathways within a tissue during cellular differentiation and removing potential biases present when profiling a whole tissue (Ryu *et al.*, 2019; Liu *et al.*, 2021).

Metabolic Gene Clusters:

One interesting discovery in plant genomics is the physical clustering of genes that encode specialized metabolic biosynthetic pathways. The first metabolic gene cluster identified was in maize (*Zea mays* L., 1753) using a mutant screening approach (Frey *et al.*, 1997) and in 2008, the first metabolic gene cluster was identified by mining a genome assembly (Field and Osbourn, 2008). Specialized metabolites with gene clusters encode terpenes, alkaloids, and cyanogenic glycosides, and have been identified in a broad range of species from the plant kingdom, including the Apocynaceae, Brassicaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, Papaveraceae, Poaceae, and Solanaceae families (Nützmann *et al.*, 2016). Plant natural products have been used by humans for centuries for uses ranging from medicinal compounds to fragrances, and genome mining for metabolic gene clusters has become an important first step to facilitate characterization of metabolite biosynthetic pathways.

Pan-Genomes:

The increasing depth of genomic sequence for a single species introduced the concept of 'pan-genomes' in 2005 (Tettelin *et al.*, 2005). Pan-genomes consist of a core subset of the genes present in all accessions of a species, and dispensable genes which are present in a subset of accessions. These dispensable genes or structural variants include copy number variants (CNVs) and presence-absence variants (PAVs). Pan-genomes have been characterized in many plant species using a variety of approaches as sequencing technologies improved and costs decreased, including array comparative hybridization (Springer *et al.*, 2009; Muñoz-Amatriaín *et al.*, 2013; Anderson *et al.*, 2014), RNA-seq (Hirsch *et al.*, 2014), WGS (Lai *et al.*, 2010; Tan *et al.*, 2012; Hardigan *et al.*, 2016; Montenegro *et al.*, 2017), and more recently whole genome assembly comparisons (Gordon *et al.*, 2017; Sun *et al.*, 2018; Walkowiak *et al.*, 2020). CNVs and PAVs have been found to function in environmental adaptation responses including stress responses (Hattori *et al.*, 2009; Gaines *et al.*, 2010; Cook *et al.*, 2012), secondary metabolism (Winzer *et al.*, 2012), and flowering time (Díaz *et al.*, 2012). Several studies have shown PAVs to be more lowly expressed, less conserved, and evolving more rapidly compared to core genes (Hirsch *et al.*, 2016; Gordon *et al.*, 2017), leading to hypotheses that PAVs are sub-population specific genes being lost due to genome fractionation and/or pseudogenization (Gordon *et al.*, 2017).

DISSERTATION PROJECTS AND SIGNIFICANCE

Here, a genome assembly and annotation of the medicinal plant *Calotropis gigantea* (L.) W.T.Aiton, differential and co-expression analyses in maize, and circadian clock regulated co-expression networks in potato provide rich genomics resources for understanding basic aspects of plant biology. The *C. gigantea* genome was assembled using NGS technologies with a near

complete representation of the genic space, enabling discovery of novel enzymes involved in the biosynthesis of the anti-cancer cardenolide compounds (Hoopes *et al.*, 2018). Through the development of an expanded gene atlas that profiles maize development and stress responses, tissue-specific and stress-induced gene expression were characterized using co-expression principles (Hoopes *et al.*, 2019). An online database was also developed to disseminate the improved gene functional annotation to the maize community (http://maize.plantbiology.msu.edu). Diurnal and circadian co-expression networks were also characterized in potato leaf and tuber tissue using time course experiments, elucidating circadian rhythms in a heterotrophic tissue and providing the means to explore pathways regulated by the circadian clock.

REFERENCES

REFERENCES

Abbas-Aghababazadeh, F., Li, Q. and Fridley, B.L. (2018) Comparison of normalization approaches for gene expression studies completed with highthroughput sequencing. *PLoS One*, **13**, e0206312.

Adams, M.D., Kelley, J.M., Gocayne, J.D., et al. (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science (80-.).*, **252**, 1651–1656.

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E. and Gouil, Q. (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.*, **21**, 30.

Anderson, J.E., Kantar, M.B., Kono, T.Y., et al. (2014) A roadmap for functional structural variants in the soybean genome. *G3 Genes, Genomes, Genet.*, **4**, 1307–1318.

Au, K.F., Sebastiano, V., Afshar, P.T., et al. (2013) Characterization of the human ESC transcriptome by hybrid sequencing. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, E4821–E4830.

Bainbridge, M.N., Warren, R.L., Hirst, M., et al. (2006) Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics*, **7**, 246.

Bender, W., Spierer, P., Hogness, D.S. and Chambon, P. (1983) Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in Drosophila melanogaster. *J. Mol. Biol.*, **168**, 17–33.

Berger, J.A., Hautaniemi, S., Järvinen, A.K., Edgren, H., Mitra, S.K. and Astola, J. (2004) Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC Bioinformatics*, **5**, 194.

Blanchard, A.P., Kaiser, R.J. and Hood, L.E. (1996) High-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.

Bolstad, B.M., Irizarry, R.A., Åstrand, M. and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Bumgarner, R. (2013) Overview of dna microarrays: Types, applications, and their future. *Curr. Protoc. Mol. Biol.*, **101**, 22.1.1-22.1.11.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C. and Jaffe, D.B. (2008) ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Res.*, **18**, 810–820.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. and Li, H. (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*, **18**, 170–175.

Cheung, F., Haas, B.J., Goldberg, S.M.D., May, G.D., Xiao, Y. and Town, C.D. (2006) Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. *BMC Genomics*, **7**, 272.

Cook, D.E., Lee, T.G., Guo, X., et al. (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science (80-.).*, **338**, 1206–1209.

D'Haeseleer, P. (2005) How does gene expression clustering work? *Nat. Biotechnol.*, **23**, 1499–1501.

Díaz, A., Zikhali, M., Turner, A.S., Isaac, P. and Laurie, D.A. (2012) Copy number variation affecting the photoperiod-B1 and vernalization-A1 genes is associated with altered flowering time in wheat (Triticum aestivum). *PLoS One*, **7**, e33234.

Dillies, M.A., Rau, A., Aubert, J., et al. (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.*, **14**, 671–683.

Field, B. and Osbourn, A.E. (2008) Metabolic Diversification—Independent Assembly of Operon-Like Gene Clusters in Different Plants. *Science (80-.).*, **320**, 543–547.

Fleischmann, R.D., Adams, M.D., White, O., et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (80-.).*, **269**, 496–512.

Frey, M., Chomet, P., Glawischnig, E., et al. (1997) Analysis of a chemical plant defense mechanism in grasses. *Science (80-.).*, **277**, 696–699.

Gaines, T.A., Zhang, W., Wang, D., et al. (2010) Gene amplification confers glyphosate resistance in Amaranthus palmeri. *Proc. Natl. Acad. Sci.*, **107**, 1029–1034.

Gan, X., Stegle, O., Behr, J., et al. (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature*, **477**, 419–423.

Garalde, D.R., Snell, E.A., Jachimowicz, D., et al. (2018) Highly parallel direct RN A sequencing on an array of nanopores. *Nat. Methods*, **15**, 201–206.

Giani, A.M., Gallo, G.R., Gianfranceschi, L. and Formenti, G. (2020) Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.*, **18**, 9–19.

Glenn, T.C. (2011) Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.*, **11**, 759–769.

Goff, S.A., Ricke, D., Lan, T.-H., et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science (80-.).*, **296**, 92–100.

Gordon, S.P., Contreras-Moreira, B., Woods, D.P., et al. (2017) Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure. *Nat. Commun.*, **8**, 2184.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644–52.

Hardigan, M.A., Crisovan, E., Hamilton, J.P., et al. (2016) Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated Solanum tuberosum. *Plant Cell*, **28**, 388–405.

Hattori, Y., Nagai, K., Furukawa, S., et al. (2009) The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature*, **460**, 1026–1030.

Hirsch, C.N., Foerster, J.M., Johnson, J.M., et al. (2014) Insights into the Maize Pan-Genome and Pan-Transcriptome. *Plant Cell*, **26**, 121–135.

Hirsch, C.N., Hirsch, C.D., Brohammer, A.B., et al. (2016) Draft Assembly of Elite Inbred Line PH207 Provides Insights into Genomic and Transcriptome Diversity in Maize. *Plant Cell*, **28**, 2700–2714.

Hoopes, G.M., Hamilton, J.P., Kim, J., Zhao, D., Wiegert-Rininger, K., Crisovan, E. and Buell, C.R. (2018) Genome assembly and annotation of the medicinal plant Calotropis gigantea, a producer of anticancer and antimalarial cardenolides. *G3 Genes, Genomes, Genet.*, **8**, 385–391.

Hoopes, G.M., Hamilton, J.P., Wood, J.C., Esteban, E., Pasha, A., Vaillancourt, B., Provart, N.J. and Buell, C.R. (2019) An updated gene atlas for maize reveals organ-specific and stress-induced genes. *Plant J.*, **97**, 1154–1167.

International Rice Genome Sequencing Project and Saki, T. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Jain, M., Olsen, H.E., Paten, B. and Akeson, M. (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.*, **17**, 239.

Kolmogorov, M., Yuan, J., Lin, Y. and Pevzner, P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.*, **37**, 540–546.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M. (2017) Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res.*, **27**, 722–736.

Kovaka, S., Zimin, A. V., Pertea, G.M., Razaghi, R., Salzberg, S.L. and Pertea, M. (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.*, **20**, 278.

Lai, J., Li, R., Xu, X., et al. (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.*, **42**, 1027–1030.

Lee, Y., Tsai, J., Sunkara, S., et al. (2005) The TIGR Gene Indices: Clustering and assembling EST and know genes and integration with eukaryotic genomes. *Nucleic Acids Res.*, **33**, D71–D74.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A. and Dewey, C.N. (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, **26**, 493–500.

Li, Z., Chen, Y., Mu, D., et al. (2012) Comparison of the two major classes of assembly algorithms: Overlap-layout-consensus and de-bruijn-graph. *Brief. Funct. Genomics*, **11**, 25–37.

Lieberman-aiden, E., Berkum, N.L. Van, Williams, L., et al. (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science (80-.).*, **326**, 289–294.

Liu, Q., Liang, Z., Feng, D., et al. (2021) Transcriptional landscape of rice roots at the single-cell resolution. *Mol. Plant*, **14**, 384–394.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

Luo, R., Liu, B., Xie, Y., et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, **1**, 18.

Martin, J.A. and Wang, Z. (2011) Next-generation transcriptome assembly. *Nat. Rev. Genet.*, **12**, 671–682.

Michael, T.P. and Jackson, S. (2013) The First 50 Plant Genomes. *Plant Genome*, **6**.

Miller, R.T., Christoffels, A.G., Gopalakrishnan, C., Burke, J., Ptitsyn, A.A., Broveak, T.R. and Hide, W.A. (1999) A comprehensive approach to clustering of expressed human gene sequence: The Sequence Tag Alignment and Consensus Knowledge base. *Genome Res.*, **9**, 1143–1155.

Moll, P., Ante, M., Seitz, A. and Reda, T. (2014) QuantSeq 3′ mRNA sequencing for RNA quantification. *Nat. Methods*, **11**, i–iii.

Montenegro, J.D., Golicz, A.A., Bayer, P.E., et al. (2017) The pangenome of hexaploid bread wheat. *Plant J.*, **90**, 1007–1013.

Mounet, F., Moing, A., Garcia, V., et al. (2009) Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. *Plant Physiol.*, **149**, 1505–1528.

Muñoz-Amatriaín, M., Eichten, S.R., Wicker, T., et al. (2013) Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biol.*, **14**, R58.

Myers, E.W., Sutton, G.G., Delcher, A.L., et al. (2000) A Whole-Genome Assembly of Drosophila. *Science (80-.).*, **287**, 2196–2204.

Nagaraj, S.H., Gasser, R.B. and Ranganathan, S. (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Brief. Bioinform.*, **8**, 6–21.

Nurk, S., Walenz, B.P., Rhie, A., et al. (2020) HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.*, **30**, 1291–1305.

Nützmann, H.W., Huang, A. and Osbourn, A. (2016) Plant metabolic clusters – from genetics to genomics. *New Phytol.*, **211**, 771–789.

Oshlack, A. and Wakefield, M.J. (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biol. Direct*, **4**, 14.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. and Fodor, S.P.A. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 5022–5026.

Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat. Genet.*, **32**, 496–501.

Rensink, W.A. and Buell, C.R. (2005) Microarray expression profiling resources for plant genomics. *Trends Plant Sci.*, **10**, 603–609.

Rizzi, R., Beretta, S., Patterson, M., Pirola, Y., Previtali, M., Vedova, G. Della and Bonizzoni, P. (2019) Overlap graphs and de Bruijn graphs: data structures for de novo genome assembly in the big data era. *Quant. Biol.*, **7**, 278–292.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Ryu, K.H., Huang, L., Kang, H.M. and Schiefelbein, J. (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.*, **179**, 1444–1456.

Schmidt, M.H.W., Vogel, A., Denton, A.K., et al. (2017) De novo assembly of a new Solanum pennellii accession using nanopore sequencing. *Plant Cell*, **29**, 2336–2348.

Sehuler, G.D. (1997) Pieces of use puzzle: Expressed sequence tags and the catalog of human genes. *J. Mol. Med.*, **75**, 694–698.

Sharon, D., Tilgner, H., Grubert, F. and Snyder, M. (2013) A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.*, **31**, 1009–1014.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., et al. (2011) The genome of woodland strawberry (Fragaria vesca). *Nat. Genet.*, **43**, 109–116.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M. and Birol, I. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Res.*, **19**, 1117–1123.

Springer, N.M., Ying, K., Fu, Y., et al. (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.*, **5**, e1000734.

Sun, S., Zhou, Y., Chen, J., et al. (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.*, **50**, 1289–1295.

Tan, S., Zhong, Y., Hou, H., Yang, S. and Tian, D. (2012) Variation of presence/absence genes among Arabidopsis populations. *BMC Evol. Biol.*, **12**, 86.

Tettelin, H., Masignani, V., Cieslewicz, M.J., et al. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." *Proc. Natl. Acad. Sci.*, **102**, 13950–13955.

The 1001 Genomes Consortium (2016) 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell*, **166**, 481–491.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Genet. Resour. Crop Evol.*, **408**, 796–815.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Baren, M.J. van, Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, **28**, 511–515.

Usadel, B., Obayashi, T., Mutwil, M., et al. (2009) Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. *Plant, Cell Environ.*, **32**, 1633–1651.

Vanburen, R., Bryant, D., Edger, P.P., et al. (2015) Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. *Nature*, **527**, 508–511.

Velasco, R., Zharkikh, A., Troggio, M., et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One*, **2**, e1326.

Walkowiak, S., Gao, L., Monat, C., et al. (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature*, **588**, 277–283.

Winzer, T., Gazda, V., He, Z., et al. (2012) A papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science (80-.).*, **336**, 1704–1708.

Yu, J., Hu, S., Wang, J., et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science (80-.).*, **296**, 79–92.

Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., et al. (2016) Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.*, **34**, 303–311.

CHAPTER 2

GENOME ASSEMBLY AND ANNOTATION OF THE MEDICINAL PLANT *CALOTROPIS GIGANTEA*, A PRODUCER OF ANTICANCER AND ANTIMALARIAL CARDENOLIDES

This chapter was published in the following manuscript:

Hoopes G.M., Hamilton J.P., Kim J., Zhao D., Wiegert-Rininger K, Crisovan E., Buell C.R. (2018) Genome assembly and annotation of the medicinal plant Calotropis gigantea, a producer of anti-cancer and anti-malarial cardenolides. G3 8:385-391. doi: 10.1534/g3.117.300331.

CHAPTER 3

AN UPDATED GENE ATLAS FOR MAIZE REVEALS ORGAN-SPECIFIC AND STRESS-INDUCED GENES

This chapter was published in the following manuscript:

Hoopes G.M., Hamilton J.P., Wood J.C., Esteban E., Pasha A. Vaillancourt B., Provart N.J., Buell C.R. (2019) An Updated Gene Atlas for Maize Reveals Organ-Specific and Stress-Induced Genes. Plant Journal 97: 1154-1167. doi: 10.1111/tpj.14184.

CHAPTER 4

KEEPING TIME IN THE DARK: POTATO DIEL AND CIRCADIAN RHYTHMIC GENE EXPRESSION REVEALS TISSUE-SPECIFIC CIRCADIAN CLOCKS

Genevieve M. Hoopes[1], Daniel Zarka[2], Kaitlyn Acheson[1], John P. Hamilton[1], David Douches[2], C. Robin Buell[1,3,4], Eva Farré[1]

[1]Department of Plant Biology, Michigan State University, East Lansing, MI 48824 USA.

[2]Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI 48824 USA.

[3]Michigan State University AgBioResearch, Michigan State University, East Lansing, MI 48824 USA.

[4]Plant Resilience Institute, Michigan State University, East Lansing, MI 48824 USA.

This chapter is in preparation for submission as an Original Research Article to New Phytologist.

SUMMARY

The circadian clock is an internal molecular oscillator and coordinates numerous physiological processes at specific times of day through regulation of molecular pathways. Tissue-specific clocks connected by mobile signals have previously been found to run at different speeds in above- and below-ground *Arabidopsis thaliana* tissues. However, the implications of tissue variation in circadian clocks in crop species are unknown as no study has profiled diel or circadian gene expression in the specialized below-ground heterotrophic tissues. In this study, leaf and tuber global gene expression in potato (*Solanum tuberosum* L.) under cycling and constant environmental conditions was profiled, revealing diel and circadian expression patterns among 16.4% and 5.5% of the expressed genes in the tuber. A longer period, delayed phase, and lower amplitude was observed in the tuber for genes cycling in both tissues, with over 500 genes displaying differential diel phases. Core plant pathways were enriched for rhythmic gene expression, with pervasive diel and circadian expression among genes encoding enzymes that function in carbohydrate metabolism, a key pathway for tuber formation and bulking. Intriguingly, few core circadian clock genes displayed circadian expression patterns in either leaf or tuber tissue, while all core clock genes were circadian rhythmic in closely related tomato (*Solanum lycopersicum* L.) leaves, suggesting alternative regulatory mechanisms and/or clock composition is present in potato. Furthermore, robust diel and circadian transcriptional rhythms were observed among detached tubers, indicating the presence of tissue-specific independent circadian clocks. Our results provide the first evidence of a functional circadian clock in below-ground storage organs and hold important implications for other storage root and tuberous crops.

INTRODUCTION

Daily environmental changes, such as light and temperature cycles, have led organisms to time molecular and physiological activities at specific times of day. For example, C3 plants open and close their stomata during the day and night, respectively, and solar tracking species such as sunflower follow the course of the sun during the day and return to an easterly direction at night (Vandenbrink *et al.*, 2014). Many organisms have adapted internal molecular oscillators or circadian clocks, composed of interlocked transcriptional and post-translational feedback loops with groups of genes expressed at specific times of day, to anticipate these environmental changes and attune their physiology to their environment (Greenham and McClung, 2015). In *Arabidopsis thaliana*, CIRCADIAN-CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) are partially redundant morning-phased MYB transcription factors, which transcriptionally repress *PSEUDO-RESPONSE REGULATORS* (*PRRs*) and members of the EVENING COMPLEX (EC), including *EARLY FLOWERING 4* (*ELF4*) and *LUX ARRHYTHMO* (*LUX*) (Nakamichi, 2020). REVEILLES (RVEs) and NIGHT LIGHT INDUCIBLE AND CLOCK-REGULATED proteins (LNKs) interact to transcriptionally activate *PRR5*, *Timing of CAB1* (*TOC1*), and other evening genes. In turn, *PRRs* are expressed throughout the day and repress the morning-phased *CCA1*, *LHY*, *RVEs*, and *LNKs*. *TOC1* is a PRR expressed at dusk and represses *ELF4* and *LUX*. The evening-phased ELF3, ELF4, and LUX form the EC and repress the *PRR7* and *PRR9* (Nakamichi, 2020). Numerous physiological processes are regulated by the circadian clock in plants including photosynthesis (Dodd *et al.*, 2014), carbohydrate metabolism (Graf *et al.*, 2010), defense responses (Butt *et al.*, 2020), flowering (Shim *et al.*, 2017). Diel and circadian transcriptomes have been profiled in *A. thaliana* seedlings and leaves (Harmer *et al.*, 2000; Smith *et al.*, 2004; Bläsing *et al.*, 2005; Edwards *et al.*, 2006; Covington and Harmer, 2007;

Michael *et al.*, 2008), and in leaves of poplar (Filichkin *et al.*, 2011), rice (Filichkin *et al.*, 2011), tomato (Müller *et al.*, 2016), lettuce (Higashi *et al.*, 2016), barley (Müller *et al.*, 2020), and other species.

The vast majority of circadian clock studies have been performed in *A. thaliana* seedlings without differentiation between above or below-ground tissues. Several studies have demonstrated a functional circadian clock in *A. thaliana* roots with many of the core circadian clock genes expressed in roots (Voß *et al.*, 2015; Bordage *et al.*, 2016). Differences in circadian period (the amount of time complete one cycle), phase (the timing of the cycle), and amplitude (the intensity of the cycle) have been observed between *A. thaliana* shoots and roots using luciferase reporter assays. Roots had a longer period, delayed phase, and lower amplitude (Bordage *et al.*, 2016; Greenwood *et al.*, 2019; Chen *et al.*, 2020). In addition to tissue-specific circadian clocks, some studies reported altered circadian rhythms in *A. thaliana* roots upon disruption of the clock in the shoot apical meristem (Takahashi *et al.*, 2015) with mobile signals passing from *A. thaliana* shoots to the roots to coordinate the pace of each clock (Chen *et al.*, 2020).

Cultivated potato (*Solanum tuberosum* L.) is a highly heterozygous, vegetatively propagated tetraploid ($2n = 4x = 48$) and the fourth most produced food crop globally (http://www.fao.org/faostat/en/#home). Potato and tomato (*Solanum lycopersicum* L.) are estimated to have diverged only 6 million years ago (Wang *et al.*, 2008) and share similar domestication histories (Lin *et al.*, 2014; Hardigan *et al.*, 2017). During domestication and improvement efforts, a lengthening of the circadian period was observed for tomato, while potato retained a short circadian period (Müller *et al.*, 2016; Hardigan *et al.*, 2017). In potato, several core circadian clock genes, including *RVE3/5*, *LNK3/4*, and *ELF4*, had signatures of selection during both domestication and improvement efforts (Hardigan *et al.*, 2017).

In potato, there is diel regulation of carbohydrate metabolism in both leaves and tubers. Starch is the most abundant component of the under-ground heterotrophic tuber after water, accounting for 16-24% of total weight (Hoover R., 2001). Transitory starch in the leaf accumulates during the day and is degraded at night, and the rate of starch accumulation in tubers is higher at the end of the day than at dawn (Geigenberger and Stitt, 2000). Moreover, analysis of growth ring formation in tuber starch granules indicates a role of the circadian clock in starch deposition (Pilling and Smith, 2003). Other diel and circadian rhythms have been observed in potato leaf tissue, including delayed fluorescence (Hardigan *et al.*, 2017), leaf movement (Yanovsky *et al.*, 2000), and gene expression (Morris *et al.*, 2014); however, no study of global circadian rhythmic gene expression has been conducted in potato. Indeed, little is known about the potato circadian clock and the pathways which are regulated by the clock. While studies have examined the circadian clock in *A. thaliana* (Voß *et al.*, 2015) and soybean (*Glycine max* (L.) Merr.) roots (Matsuda *et al.*, 2020), as well as other above-ground heterotrophic organs including maize (*Zea mays* L.) ears (Hayes *et al.*, 2010) and grape (*Vits vinifera* L.) berries (Rienth *et al.*, 2014), no study has investigated the circadian clock in other below-ground heterotrophic tissues such as storage roots or tubers.

Here, potato leaf and tuber tissues were sampled in time course experiments under both cycling and constant environmental conditions, and global gene expression was profiled via 3' mRNA-sequencing to characterize tissue-specific diel and circadian rhythmic gene expression. Differences in circadian period, phase, and amplitude between the tissues were examined and genes displaying differentially rhythmic expression patterns identified. Both carbohydrate metabolism and core circadian clock genes were investigated. Potato luciferase reporter lines were

developed using the promoter of a gene displaying circadian rhythmic gene expression, and transcriptional rhythms were examined in tubers and detached tubers.

METHODS

Plant Material and Growth Conditions:

The round white tetraploid chipping cultivar Atlantic was used for all studies and maintained through vegetative propagation. For the time course transcriptional profiling experiments, Atlantic was grown in Suremix soil (Michigan Grower Products, Galesburg, MI) in 4-inch pots in a BioChambers TPC-19 growth chamber under 340 $\mu mol \bullet m^{-2} \times s^{-1}$ light intensity under 12 hour light/dark 21°C/18°C (LD/HC) cycles for 52 days. Afterwards, replicated leaf and tuber tissues were harvested every four hours for 24 hours under the same cycling conditions (LD/HC samples) and were harvested every four hours for 48 hours under constant light and temperature (22°C) conditions (LL samples). For the luciferase reporter assays, transgenic Atlantic lines were grown in Green's Grade soil (Profile, Buffalo Grove, IL) in 4-inch pots in the same BioChambers TPC-19 growth chamber under 340 $\mu mol \times m^{-2} \times s^{-1}$ light intensity under 12 hour light/dark 21°C/18°C cycles for six weeks. A KB400 growth chamber (Binder, Germany) was used with 70 $\mu mol \times m^{-2} \times s^{-1}$ light intensity provided by Heliospectra RX-30 (Heliospetra, Chicago, IL) and the light spectrum set as described previously (Hardigan et al. 2017) for the luciferase assays. A 12 hour light/dark cycle and constant light conditions were used over the course of five days at 22°C during the assay.

Library Preparation and Sequencing:

Leaf and tuber samples harvested from the LD/HC and LL time course experiments were

processed as follows. For both leaves and tubers, three biological replicates were collected at each time point. For each replicate, a 1.27 cm diameter cork borer was used to collect two leaf punches from the penultimate leaflet on the third leaf from the top of the plant that were flash frozen in liquid nitrogen. For tubers, a 0.64 cm diameter cork borer was used to collect one core each from three tubers on each plant. The periderm of the tuber core was cut off prior to freezing in liquid nitrogen. RNA was isolated from the leaf and tuber samples using hot phenol as described previously (The Potato Genome Sequencing Consortium, 2011) and DNase-treated with TURBO DNase (Invitrogen, Waltham, MA). DNase-treated RNA samples were sent to the Genomics Facility at Cornell University (http://www.biotech.cornell.edu/brc/genomics-facility) where Lexogen QuantSeq 3' mRNA-sequencing FWD libraries (Vienna, Austria) were constructed and subsequently sequenced on an Illumina NextSeq 500 instrument (San Diego, CA, USA) to obtain 86 nucleotide (nt) single end reads.

Calculation of Expression Abundances and Preferential Tissue Expression:

FastQC (v0.11.8) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was used to assess the quality of the 3' mRNA-seq and Cutadapt (v1.16) (Martin, 2011) was used to trim adapter sequences and remove low quality bases (-q 20,20) and N-bases (-trim-n). For the 3' mRNA-seq reads, the first 12 nt of the read were removed (-u 12) and poly(A) tails were trimmed (-a "A{10}"$). The trimmed 3' mRNA-seq reads were aligned to the *S. tuberosum* Group Phureja Doubled Monoploid (DM 1-3 R44 genome (v6.1; hereafter DM) (Pham *et al.*, 2020) using HISAT2 (v2.2.1) (Kim *et al.*, 2019) with a maximum intron length of 20 kb. The 3' mRNA-seq reads were aligned in a stranded manner (--rna-strandness F) and the mis-match penalty was set as follows: --mp 1,1. The featureCounts program in the SUBREAD package (v2.0.1) (Liao *et al.*,

2014) was used to count reads with a minimum MAPQ score > 0 (-Q 1) by exon, which were grouped by gene. The 3' mRNA-seq alignments were counted in a stranded manner (-s 1) and the 5' end of the gene was extended 500 bp (--readExtension5 500) to account for un-annotated UTR regions. To compare potato leaf gene expression with tomato, raw mRNA-sequencing (mRNA-seq) reads generated previously (Müller *et al.*, 2016) were downloaded from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/). Reads were assessed for quality and cleaned as described above for potato and aligned using HISAT2 (v2.2.1) (Kim *et al.*, 2019) with a maximum intron length of 20 kb to the International Tomato Genome Sequencing Project (ITAG) *S. lycopersicum* genome (SL4.0) (Hosmani *et al.*, 2019); expression abundances were determined using the ITAG *S. lycopersicum* (ITAG4.1) annotation (Hosmani *et al.*, 2019).

To test for tissue-specific expression and obtain normalized gene counts, DESeq2 (v1.22.2) (Love *et al.*, 2014) was used. Tomato leaf samples, potato leaf samples, and potato tuber samples were separately processed with the DESeq2 'rlog' function to obtain gene counts normalized for sequencing depth, RNA composition, and variance stabilization. To test for outliers in replicates, Pearson's Correlation Coefficients (PCC) were calculated using the R 'cor' function. One potato leaf sample harvested at 3 AM had PCC values below 0.9 with the other two biological replicates and was removed (Figure S4.1). Potato leaf and tuber samples were tested for preferential tissue expression using an alpha level of 0.01. Genes with an adjusted p-value < 0.01 and a log2-fold change > 2 or log2-fold change < -2 were considered differentially expressed.

Identification of Rhythmic Gene Expression:

Normalized gene counts for the potato samples were used with the JTK_CYCLE algorithm in MetaCycle (v1.2.0) (Wu *et al.*, 2016) and with ECHO (v4.0) (De Los Santos *et al.*, 2020) to

identify rhythmic genes. ECHO uses an extended harmonic oscillator to identify rhythms which have changes in amplitude over time (De Los Santos *et al.*, 2020). Default parameters were used in ECHO and the distribution of gene-wise period fits for significant genes (BH P-value < 0.01) (Figure S4.2) was used to define input period range limitations for MetaCycle. The gene-wise period range was limited to 12-28 hours for all LD/HC samples, while the gene-wise period range varied by tissue for LL samples (leaf 12-48 h and tuber 12-52 hours). Results between the JTK_CYCLE and ECHO algorithms were combined following the methods of MetaCycle where p-values are combined with Fisher's Method, period lengths are averaged, and the circular mean is calculated for the phases (Wu *et al.*, 2016). The relative amplitude was obtained from MetaCycle. This combined data set is referred to as 'JTK_ECHO' and genes with an FDR adjusted p-value < 0.01 were considered to display rhythmic gene expression.

Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008) was also used to identify rhythmic gene expression patterns independently of fitting a model to the data. Independently for LD/HC and LL, a coefficient of variance (COV) filter was used first to remove genes with unvarying gene expression patterns (Table S4.1) and WGCNA (v1.69) (Langfelder and Horvath, 2008) was then performed with the COV filtered gene sets (referred to as COV) (Table S4.1). COV gene co-expression modules displaying diel and circadian rhythmic gene expression were identified based on percent overlap with JTK_ECHO significant genes and manual curation of the eigengene patterns. Genes not grouped into a co-expression module are placed in a catch-all 'grey' module by WGCNA and any module displaying double the percentage of JTK_ECHO significant genes compared to the catch-all 'grey' module was considered a 'cycling module'. The catch-all 'grey' module was excluded from further analyses.

Species and Tissue Comparisons:

To compare the genes identified as displaying rhythmic gene expression between potato and tomato, orthologous groups were identified using OrthoFinder (v2.5.1) (Emms and Kelly, 2019) with the representative protein models of *A. thaliana* (TAIR10) (Lamesch *et al.*, 2012), *Solanum pennellii* (Bolger *et al.*, 2014), *S. lycopersicum* (ITAG4.1) (Hosmani *et al.*, 2019), *S. tuberosum* (DMv6.1) (Pham *et al.*, 2020). Orthologs of *A. thaliana* core circadian clock genes were assigned by examining the orthologous group gene tree from OrthoFinder. Orthologous groups were also determined between PGSCv4.03 (The Potato Genome Sequencing Consortium, 2011; Sharma *et al.*, 2013) and DMv6.1 (Pham *et al.*, 2020) using OrthoFinder (v2.5.1) (Emms and Kelly, 2019) to update carbohydrate metabolism genes from PGSC (Van Harsselaar *et al.*, 2017) to the DMv6.1 annotation using orthologous groupings.

To characterize time-dependent differences in expression patterns between leaf and tuber tissues, WGCNA (Langfelder and Horvath, 2008) was performed with genes containing a COV > 0.05 in both tissues (Table S4.2). DiPALM (v1.1) (Greenham *et al.*, 2020) was then used with the eigengene and gene expression values to test for significant differences in gene expression correlation with co-expression modules. Permutation tests were conducted by resampling the gene expression data 100,000 times and p-values were adjusted using the permutation test results, defining a gene as significant if it had an adjusted p-value < 0.01. All heatmaps were generated using 'pheatmap' (v1.0.12) and venn diagrams were generated with 'VennDiagram' (v1.6.20). Differences in the diel and circadian expression rhythms between shared leaf and tuber cycling genes were tested using paired Wilcoxon signed-rank tests as implemented by the 'wilcox.test' function in R (v3.6.2). Plots were generated using 'ggpubr' (v0.4.0), 'ggplot2' (v3.3.2) (Wickham, 2009), and 'RColorBrewer' (v1.1-2) R packages. Gene Ontology (GO) enrichment tests were

performed using 'topGO' (v2.36.0) and a Fisher's Exact test with the classic method and the DMv6.1 GO annotation (Pham *et al.*, 2020). P-values were adjusted using a false discovery rate (FDR) correction and GO terms with an adjusted p-value < 0.01 were retained.

Generation of Luciferase Reporter Constructs:

Modified pEarleyGate Gateway destination vector 301 (Earley *et al.*, 2006) was obtained in which the *HA* tag was replaced with Luciferase 2 (*LUC2*) (Jones *et al.*, 2015). The S*treptomyces hygroscopicus* herbicide-resistance gene *BAR*, which confers resistance to bialophos, was replaced with the *Escherichia coli NPTII* gene, which confers resistance to kanamycin. The *NPTII* gene, including the cauliflower mosaic virus derived 35S promoter and NOS terminator, were amplified from the pATMDomega vector (Welsh *et al.*, 2005) using the following primers 5' ATACAGGTACCAGCTTCCCGATCCTATCTG 3' and 5' CACACAGAGCTCGATCTAGTAACATAGATGAC 3' (Integrated DNA Technologies, Coralville, IA). After processing the amplified *NPTII* fragment and pEarleyGate 301-*LUC2* vector with the *Kpn*I-HF and *Sac*I-HF restriction enzymes (New England Biolabs, Ipswich, MA), the fragments were ligated together using the T4 DNA Ligase (New England Biolabs, Ipswich, MA). The *NPTII* insertion was sequenced via Sanger sequencing to verify no sequence variants were introduced; this destination vector was designated pEarleyGate 301-*LUC2*-kanR.

A total of 3 kb up-stream of the translational start site for the StGH3 (Soltu.DM.02G028070) gene was amplified from genomic DNA extracted from Atlantic using the following primers 5' CACCGATCATGTCACTAATACAAC 3' and 5' ATAAAGCAAGTAGAATAACTCTCAA 3' (Integrated DNA Technologies, Caralville, IA) and cloned into the pENTR/D-TOPO vector using the pENTR Directional TOPO cloning kit

(Invitrogen, Waltham, MA). After confirmation of successful integration of the promoter sequence into the vector, the Gateway LR reaction was performed with the pENTR/D-TOPO vector containing the promoter and the destination pEarleyGate 301-*LUC2*-kanR vector using the LR Clonase Enzyme Mix (Invitrogen, Waltham, MA) to obtain the final StGH3-LUC2 reporter construct.

Agrobacterium-mediated Transformation:

Atlantic plants were stably transformed with the *StGH3-LUC2* construct via agrobacterium-mediated transformation as previously described (Li *et al.*, 1999). Transgenic plants were first identified by rooting and growth on MS media containing the antibiotic kanamycin. DNA from putative transformants was tested for the presence of *NPTII* using primers *NPTII* Fwd: 5' TTTGTCAAGACCGACCTGTC 3' and *NPTII* Rev: 5' CCAACGCTATGTCCTGATAG 3' (Integrated DNA Technologies, Caralville, IA). A second confirmation used a forward primer from the *StGH3* fragment (GH3 Fwd: 5' CGTGACCGAACAAACTCACAAG 3') and the reverse from the *LUC2* gene (5' CGTCTTCGAGTGGGTAGAATGG 3') to amplify across the promoter-LUC2 junction.

Imaging Luciferase Luminescence:

With the exception of the leaf to be imaged, above-ground shoot tissue was wrapped in a screen material to prevent obstruction of the imaging plane during the assay (Figure S4.3). For tubers, on the day of the assay, tubers of *StGH3-LUC2* lines were removed from the soil and placed on top of the soil still attached to its stolon; detached tubers were removed from their stolon. The leaf and tubers to be imaged were then sprayed with 5 mM D-Luciferin (Gold Biotechnology, St.

Louis, MO) and an Andor iKon-M DU-934N-BV camera was used in the KB400 chamber to detect luciferase luminescence. Images were taken every two hours for four days using an exposure time of 20 minutes after a three minute delay. Images from the first 16 hours after spraying with D-luciferin were excluded from downstream analysis. To detrend the raw luminescence data from the initial high bioavailability of luciferase, a moving mean and standard deviation were calculated every 10 images using 'TTR' (v0.23-6). The raw luminescence values were then subtracted from the moving mean and then divided by the moving standard deviation. BioDare2 (Zielinski *et al.*, 2014) was used with the NFT NLLS algorithm and no detrending to calculate the period and phase values.

RESULTS AND DISCUSSION

Heterotrophic Tuber Tissue Displays Diel and Circadian Expression Patterns:

Little is known about the role of the circadian clock in heterotrophic, below-ground plant tissues compared to autotrophic leaf tissue, with studies to date mainly limited to *A. thaliana* roots. To determine if diel and circadian rhythmic gene expression was present in plants at the bulking developmental stage for the leaf and tuber, we profiled global gene expression in leaf and tuber tissue sampled every four hours under cycling light/dark and temperature (LD/HC) conditions, and constant light and temperature (LL) conditions for 24 and 48 hours, respectively. Genes were clustered based on expression patterns using Weighted Gene Co-expression Analysis (WGCNA) (Langfelder and Horvath, 2008) and co-expression modules were identified as displaying diel or circadian expression patterns based on percent overlap with JTK_ECHO results (see Materials and Methods) (Wu *et al.*, 2016; De Los Santos *et al.*, 2020). A total of 10,863 genes (47.3% of expressed genes or 33.0% of all genes) and 3,283 genes (16.4% of expressed genes or 10.0% of

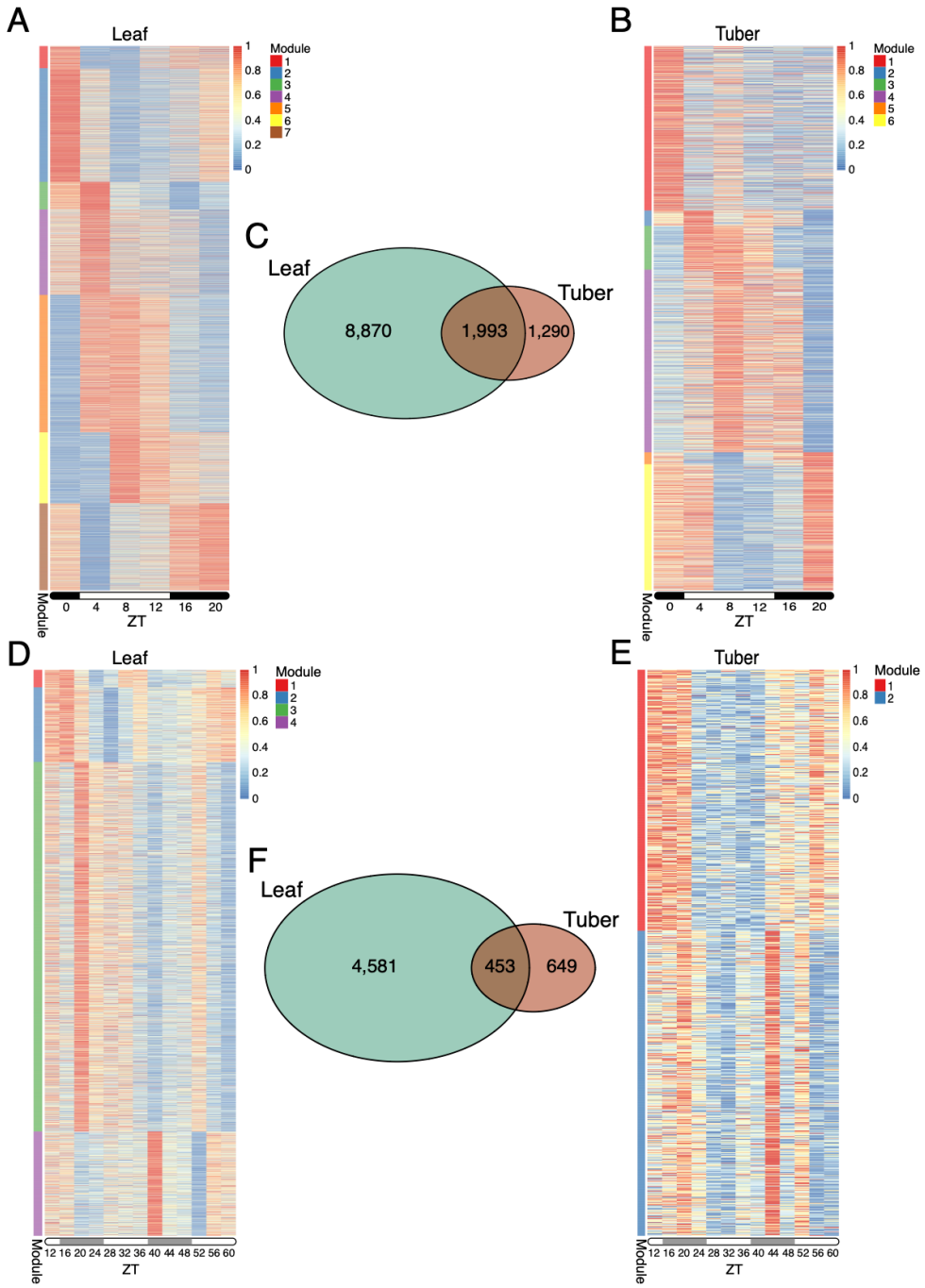

Figure 4.1: Leaf and Tuber Diurnal and Circadian Rhythmic Gene Expression
Heatmaps of the normalized gene expression values over time for the genes identified as displaying leaf diurnal (**A**), tuber diurnal (**B**), leaf circadian (**D**), or tuber circadian (**E**) expression patterns. Genes are ordered by co-expression module, which is indicated by the color along the left-side of the heatmap. The white and black bar along the bottom of each heatmap indicates the timing of day or night, respectively. The grey color in the circadian heatmaps indicates subjective night. Venn diagrams indicate how many rhythmic genes are shared between the leaf and tuber tissues for diurnal rhythms (**C**) and circadian rhythms (**F**).

all genes) displayed diel gene expression in the leaf and tuber, respectively (Figure 4.1A, 4.1B). A total of 5,034 genes (21.9% of expressed genes or 15.3% of all genes) and 1,102 genes (5.5% of expressed genes or 3.3% of all genes) displayed circadian gene expression in the leaf and tuber, respectively (Figure 4.1D, 4.1E). A total of 1,993 and 453 genes are diel and circadian rhythmic in both tissues, accounting for 60.7% and 41.1% of the cycling genes in the tuber, respectively (Figure 4.1C, 4.1F). Our approach of combining WGCNA with JTK_ECHO analyses enabled the identification of genes which had a poor JTK_ECHO model fit, but which displayed rhythmic expression patterns. Over 66% of the cycling genes in each tissue and condition were not defined as cycling by JTK_ECHO, yet displayed rhythmic expression patterns. The tuber had the highest percentage of genes uniquely characterized as cycling by WGCNA, accounting for 81.5% of the diel cycling genes. These results reflect the inability of model fitting programs to detect rhythmic gene expression patterns not conforming to model expectations.

Similar percentages of genes displaying rhythmic expression were reported in *A. thaliana* seedlings, where 27.5% and 14.9% of all genes cycled under LD/HC or LL after LD/HC entrainment, respectively (Mockler *et al.*, 2007). Diel expression patterns from 5 week old *A. thaliana* leaf tissue was also profiled, with 25.8% of all genes displaying diel expression rhythms (Mockler *et al.*, 2007). Remarkably, diel expression patterns were extensive in the tuber despite residing in the soil with little to no exposure to light conditions. In other reproductive heterotrophic tissues only 4.1% and 0.39% of all genes where identified as having diel expression patterns in grape (*Vitis vinifera* L.) berries (Rienth *et al.*, 2014) and maize (*Zea mays* L.) ears (Hayes *et al.*, 2010). A total of 4.9% of all genes displayed circadian expression patterns in *A. thaliana* lateral root primordia (Voß *et al.*, 2015), which is similar to the percentage of circadian rhythmic genes identified in the tuber.

Differential Rhythmic Gene Expression Patterns Between Leaf and Tuber Tissue:

To test for period, phase, and/or amplitude differences between leaf and tuber tissues, paired Wilcoxon and Watson-Wheeler tests were performed between genes with rhythmic expression patterns in both tissues. The tuber had a median free running period 5.4 hours longer than the leaf under LL conditions, and a lower relative amplitude in both the LD/HC and LL conditions (Figure 4.2A, 4.2B). In LD/HC conditions, genes were predominantly phased during the day in the tuber, with a median phase of 3.36 hours in LD/HC conditions (Figure 4.2C), and were delayed in their expression in the tuber compared to the leaf, with a median delayed phase of

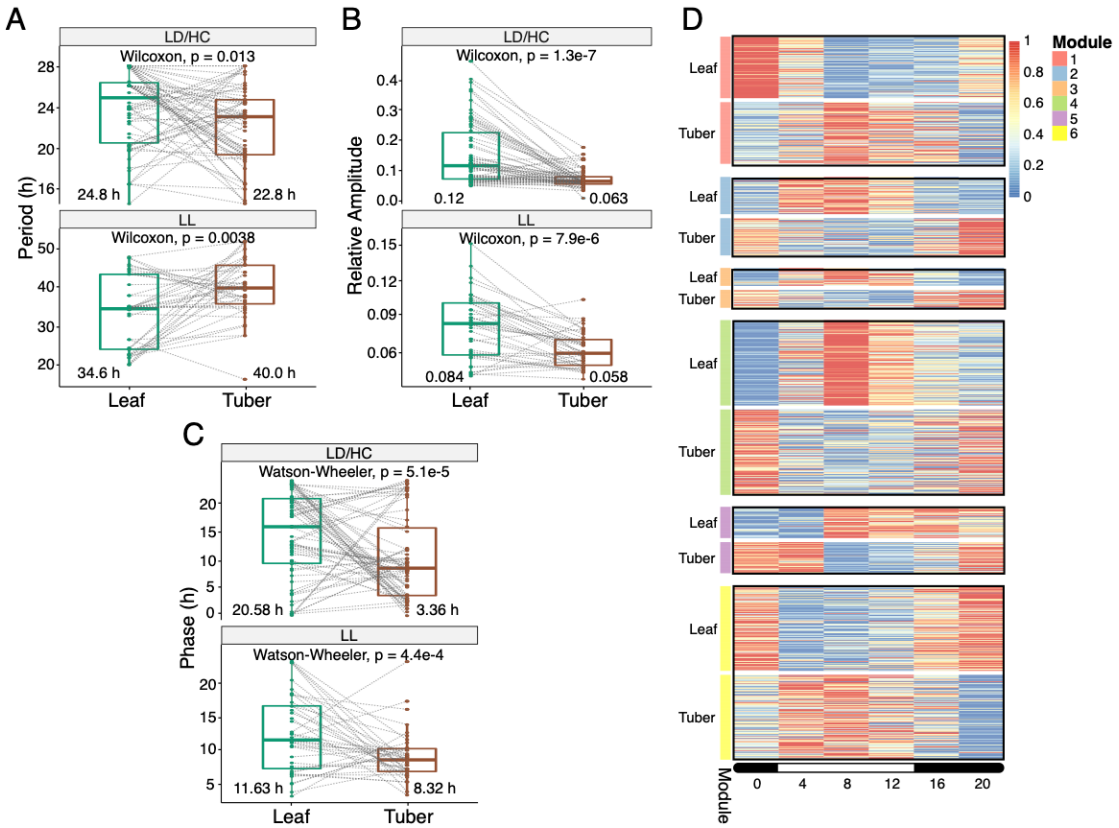

Figure 4.2: Tissue-specific Differences in Rhythmic Gene Expression Patterns
The circadian period (**A**), relative amplitude (**B**), and phase (**C**) are reported in box plots with lines connecting the data values by tissue for genes with diurnal (LC/HC) and circadian (LL) rhythms in both tissues. P-values from paired Wilcoxon and Watson-Wheeler tests are reported for each comparison and the median value is reported below the plot for each tissue. A total of 66 and 41 genes in LD/HC and LL conditions, respectively, were compared. A heatmap of the normalized gene expression abundances for the genes identified as displaying significantly different diurnal cycling patterns is provided in (**D**), with the co-expression module and tissue indicated in the colored bar on the left side and the light/dark periods indicated in the black and white bar at the bottom.

2.04 hours in LD/HC conditions. Similarly to potato tubers, *A. thaliana* roots had a longer period, shifted phase, and lower amplitude compared to shoot tissue (Bordage *et al.*, 2016; Greenwood *et al.*, 2019). These similar shifts of the circadian clock pace, timing, and intensity between two distantly related plant species and two different below-ground tissue types suggest a conserved mechanism may confer an adaptive advantage by slowing down the circadian clock in below-ground tissues.

To identify specific genes and pathways that displayed time-dependent differences between the leaf and tuber, DiPALM (Greenham *et al.*, 2020) was used to detect changes in correlation of a gene's expression pattern between co-expression modules from the two tissues. Out of the 7,487 genes with a COV> 0.05 in LD/HC conditions for both the leaf and tuber, a total of 568 genes displayed different expression patterns (Figure 4.2D), accounting for 8.6% of the tuber diel gene expression. A total of 107 genes displayed different gene expression patterns between the tissues for the LL conditions out of 7,025 genes with a COV > 0.05, accounting for 1.5% of the tuber circadian gene expression. Genes with different expression patterns in LD/HC between leaves and tubers had striking tissue-specific differences in phase (Figure 4.2D). In the leaf, these genes were predominantly phased at dusk, while in the tuber they were predominantly phased at dawn (Figure S4.4A). Indeed, a median phase difference of 3.9 hours was observed between leaf and tuber tissues (Figure S4.4B). Genes in the leaf also had significantly higher amplitude compared to the tuber in LD/HC conditions (Figure S4.4C, S4.4E). Among genes with different expression patterns under LL conditions, the leaf had a median period 7.2 hours longer compared to the tuber (Figure S4.4). No difference in amplitude was observed (Figure S4.4E, S4.4F). In LD/HC conditions, genes identified by DiPALM had a phase difference between leaf and tuber tissues almost two hours longer than the phase difference identified among genes with

diel expression patterns in both tissues, indicating DiPALM only confidently identifies changes in gene expression patterns when sizable differences are present and that many more genes are likely differentially phased and/or have period differences (Figure 4.2). Regardless, the identification of genes with differences in phases of expression between the leaf and tuber under LD/HC conditions suggest each tissue is operating shared pathways at different times of day as resource allocation is changing.

Pathways Displaying Rhythmic Gene Expression Patterns:

Several primary metabolic pathways and processes were enriched for diel and circadian rhythmic gene expression among leaf and tuber tissue. Genes functioning in photosynthesis (GO:0015979; adj. p-value < 1E-14) and related light harvesting pathways (GO:0009768; GO:0009765; GO:0019684) were highly enriched (adjusted p-value < 1E-18) among the shared genes for both diel and circadian rhythmic genes. All genes annotated with these GO terms were preferentially expressed in the leaf tissue, with an average log2-fold change (lfc) > 6 (Table S4.3), indicating that while these genes display rhythmic gene expression patterns in the tuber, they are lowly expressed. Phragmoplast-associated kinesin proteins and other ATP-binding microtubule motor proteins were enriched among shared diel rhythmic genes, suggesting mitosis and cytokinesis are, in part, regulated by environmental stimuli in both above- and below-ground tissues (Vavrdová *et al.*, 2019). Shared circadian rhythmic genes were enriched in DNA-directed RNA polymerase family proteins, including RNA polymerase beta' subunit-1 (RPOC1), RPOC2, and chloroplast RNA polymerase subunit beta (RPOB). Mitochondrial ribosome proteins were also enriched among shared circadian rhythmic genes, suggesting circadian regulation of chloroplast transcription and mitochondrial translation in both tissues. Multiple organellar RNA

editing factors and tetratricopeptide repeat proteins were enriched among leaf-specific diel rhythmic genes, and receptor-like kinases were enriched among leaf-specific circadian rhythmic genes. Diel and circadian rhythmic genes specific to the tuber were enriched for genes involved in photosynthetic light reactions in the electron transport chain and were lowly expressed in the tuber relative to the leaf leaf (lfc > 6) (Table S4.3). Previous studies in rice, lettuce, and barley have similarly found genes displaying diel and circadian rhythmic expression to be functioning in photosynthesis, transcription, translation, signal transduction, and cytoskeletal organization, among other pathways (Filichkin *et al.*, 2011; Higashi *et al.*, 2016; Müller *et al.*, 2020).

Tissue-Specific Diel Differences in Water Transport:

Genes displaying differences in phase between leaf and tuber tissue under LD/HC conditions were enriched for water transport (GO:0006833; adj. p-value = 8.93E-05). A total of four plasma membrane intrinsic proteins, which are a family of aquaporins located in the plasma membrane (Higuchi *et al.*, 1998), were present in Module 1 (Figure 4.2D). The genes in the leaf are phased at dawn while in the tuber they are phased at dusk. Over 250 genes were connected to the plasma membrane intrinsic proteins in the leaf diel co-expression network, with six genes connected to all four aquaporins. Furthermore, several dawn phased *A. thaliana* core circadian clock orthologs, including *LHY* and *RVE4/8*, were connected to three of the four aquaporins in the leaf diel co-expression network. The aquaporins were preferentially expressed in the leaf (lfc > 2.7) and while they were in a tuber diel cycling co-expression module, they did not meet the correlation threshold to be included in the tuber diel co-expression network. The tuber is composed of 70-80% water (Hoover R., 2001) and changes in water transport were thought to produce diel rhythms in tuber weight and volume in earlier studies (Baker and Moorby, 1969; Schnieders *et al.*,

1988). Recently, however, rhythms were not consistently observed in tuber volume in undisturbed plants using high resolution X-ray computed tomography (Pérez-Torres *et al.*, 2015). In dicots, circadian rhythms not accounted for by changes in carbohydrate metabolism have also been observed in leaf expansion (Poiré *et al.*, 2010). Our data suggest that the circadian clock is regulating the diel expression of aquaporins in the leaf for increased leaf expansion at dawn, and that the same aquaporins may be phased at dusk in the tuber at lower expression levels to aid in water transport.

Diel and Circadian Regulation of Leaf and Tuber Carbohydrate Metabolism:

Light stimulus is widely recognized as a predominant regulatory mechanism in carbohydrate metabolism in leaf tissue (Skryhan *et al.*, 2018) and similarly to other plant species, starch biosynthesis and degradation occur during the day and night, respectively, in potato leaf tissue (Geigenberger and Stitt, 2000). Circadian rhythms in maltose levels, a key product from starch degradation, have also been found in *A. thaliana* leaves, along with diel and circadian expression patterns among starch degradation enzymes (Lu *et al.*, 2005). Using previously identified orthologs of *A. thaliana* enzymes involved in carbohydrate metabolism (Van Harsselaar *et al.*, 2017), diel expression patterns were identified in all parts of the starch biosynthetic and degradation pathway in the leaf, with at least one paralog displaying diel rhythms (Figure 4.3A, 4.3B; S4.5). Circadian expression patterns were also identified among genes encoding enzymes involved in both starch biosynthesis and degradation, including in *HEXOKINASE* (*HK*), *UDP-GLUCOSE PYROPHOSPHORYLASE* (*UGPase*), *PHOSPHOGLUCOISOMERASE* (*PGI*), *PHOSPHOGLUCOMUTASE* (*PGM*), *ADP-GLUCOSE PYROPHOSPHORYLASE* (*AGPase*),

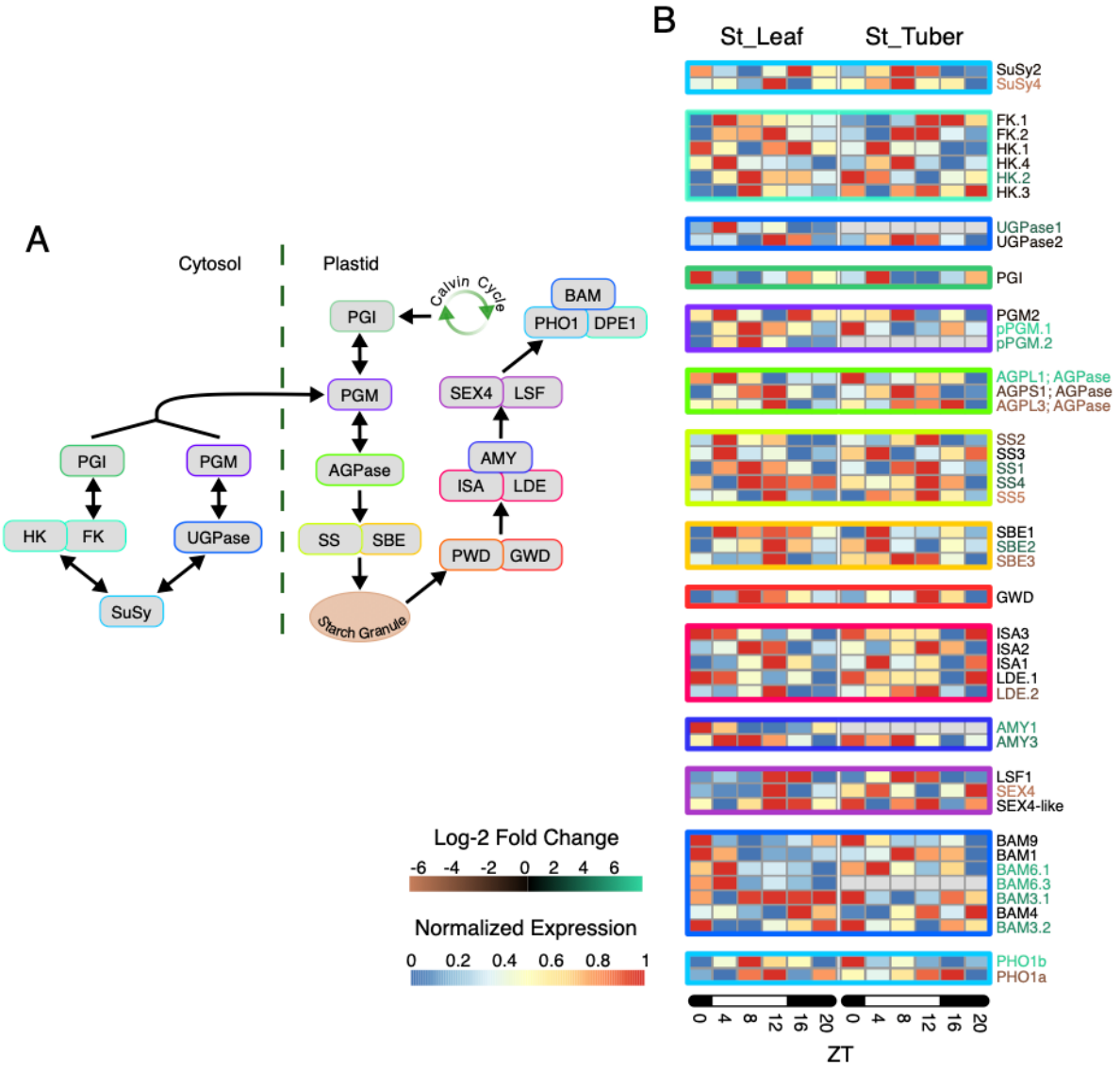

Figure 4.3: Diurnal and Circadian Rhythms in Carbohydrate Metabolism
(**A**). A model of the carbohydrate metabolism, in which boxes correspond to the enzymes in the metabolic pathway and the arrows indicate the flux of carbon. '*SuSy*': *SUCROSE SYNTHASE*; '*HK*': *HEXOKINASE*; '*FK*': *FRUCTOKINASE*; '*UGPase*': *UDP-GLUCOSE PYROPHOSPHORYLASE*; '*PGI*': *PHOSPHOGLUCOISOMERASE*; '*PGM*': *PHOSPHOGLUCOMUTASE*; '*AGPase*': *ADP-GLUCOSE PYROPHOSPHORYLASE*; '*SS*': *STARCH SYNTHASE*; '*SBE*': *STARCH BRANCHING ENZYME*; '*PWD*': *PHOSPHO-GLUCAN, WATER DIKINASE*; '*GWD*': *GLUCAN, WATER DIKINASE*; '*ISA*': *ISOAMYLASE*; '*LDE*': *LIMIT DEXTRINASE*; '*AMY*': *ALPHA-AMYLASE*; '*SEX4*': *STARCH EXCESS 4*; '*LSF*': *LIKE-STARCH EXCESS 4*; '*BAM*': *BETA-AMYLASE*; '*PHO1*': *ALPHA-GLUCAN PHOSPHORYLASE 1*; '*DPE1*': *DISPROPORTIONATING ENZYME 1*. (**B**). Heatmap of normalized expression abundances for genes encoding the enzymes involved in carbohydrate metabolism from the diurnal time course experiments. The black and white bar along the bottom indicates the timing of night and day, respectively. Colored boxes correspond to the gene color in the model. Text color corresponds to the log-2 fold change value for leaf gene expression compared to tuber gene expression.

PHOSPHO-GLUCAN, WATER DIKINASE (*PWD*), *LIMIT DEXTRINASE* (*LDE*), *BETA-AMYLASE* (*BAM*), and *ALPHA-GLUCAN PHOSPHORYLASE 1* (*PHO1*) (Figure 4.3A; S4.6). Preferential expression of carbohydrate enzymes between leaf and tuber tissue has been observed previously (Van Harsselaar *et al.*, 2017), and similar patterns were identified here. Gene families often had enzymes preferentially expressed in either the leaf or the tuber, with few enzymes displaying equal expression levels in both tissues (Figure 4.3B, S4.5, S4.6). Enzymes involved in starch biosynthesis which were not preferentially expressed in the tuber (lfc > -2) were phased throughout the day in the leaf, as expected since starch is known to accumulate during the day in the leaf. Accordingly, genes encoding enzymes involved in starch degradation not preferentially expressed in the tuber (lfc > -2) were phased around dusk in the leaves with the exception of *ISOAMYLASE 3* (*ISA3*), *LDE1*, *ALPHA-AMYLASE 1* (*AMY1*), and many of the *BAMs*. All of these gene families had members phased around dusk and redundancy in enzyme function may have enabled neo-functionalization among the members no phased at dusk. Indeed, sugars are known to act as signaling molecules in plants and enzymes involved in carbohydrate metabolism have been found to function in signaling pathways (Paulina Aguilera-Alvarado and Sanchez-Nieto, 2017; Janse van Rensburg and Van den Ende, 2018).

Carbohydrate metabolism is critical for tuber formation and bulking. As a heterotrophic tissue, the tuber relies on sugars transported from autotrophic leaf tissue, and carbon flux towards starch increases the end of the day, which correlated with an increase in expression level of *SUCROSE SYNTHASE* (*SuSy*) and *AGPase* (Geigenberger and Stitt, 2000). We confirmed these diel expression patterns and observed diel oscillations in genes involved in sucrose degradation, starch biosynthesis and degradation (Figure 4.3B, S4.5). Some of these genes, such as *SuSy*, *HK*, *AGPase*, *ISA*, *AMY*, *LIKE-STARCH EXCESS 4* (*LSF*), and *PHO1* also displayed circadian

expression patterns (Figure S4.6). Exceptions were genes encoding *PGI*, *PGM*, and *AMY* enzymes, which did not cycle (Figure 4.3B, S4.5). Starch biosynthetic enzymes not preferentially expressed in the leaf (lfc < 2) are predominantly phased in the afternoon in the tuber (Figure 4.3B, S4.5), supporting previous observations of higher expression in the evening compared to the morning (Geigenberger and Stitt, 2000). Enzymes involved in starch degradation, such as *GLUCAN, WATER DIKINASE* (*GWD)*, *ISA*, *LDE*, *LSF*, and *PHO1* were expressed in the tuber and predominantly phased close to dusk (Figure 4.3B, S4.5), suggesting starch may be degraded at night in bulking tubers in the absence of other sugars transported from the leaf.

Through identification of the time-of-day enzymes in carbohydrate metabolism are expressed in both leaf and tuber tissue, we have improved the functional characterization of these genes. The pervasive circadian expression patterns among both starch biosynthesis and degradation suggest carbohydrate metabolism is a key pathway regulated by the circadian clock. Transcriptional regulation of carbohydrate metabolism provides mid- to long-term coordination of starch turnover while post-transcriptional regulation is a key regulatory mechanism for enzyme activity (Kötting *et al.*, 2010; Skryhan *et al.*, 2018). Future studies are needed to determine if diel and circadian rhythms are present in protein abundance or other post-translational modifications in potato.

Circadian Clock Genes Do Not Display Circadian Rhythmic Expression:

Core circadian clock orthologs display circadian rhythmic expression patterns in leaf tissue among a diverse range of plant species, including *A. thaliana* (Bläsing *et al.*, 2005), rice (Filichkin *et al.*, 2011), poplar (Filichkin *et al.*, 2011), lettuce (Higashi *et al.*, 2016), barley (Müller *et al.*, 2020), and others. Diel expression patterns among core clock genes have also been observed in

heterotrophic tissues such as soybean roots (Matsuda *et al.*, 2020) and maize ears (Hayes *et al.*, 2010), and circadian expression patterns have been observed in *A. thaliana* roots (Voß *et al.*, 2015). To determine if potato leaf and tuber tissues have diel and circadian rhythmic expression of core circadian clock genes, orthologs of *A. thaliana* clock genes were identified in potato and the number of paralogs present for each gene were the same as previously identified (Bombarely *et al.*, 2016) except for *PRR3* in which two copies were found (Table S4.4). *CCA1* is not present in the Solanaceae family (Bombarely *et al.*, 2016), and a total of six *RVEs* and four *LNKs* were

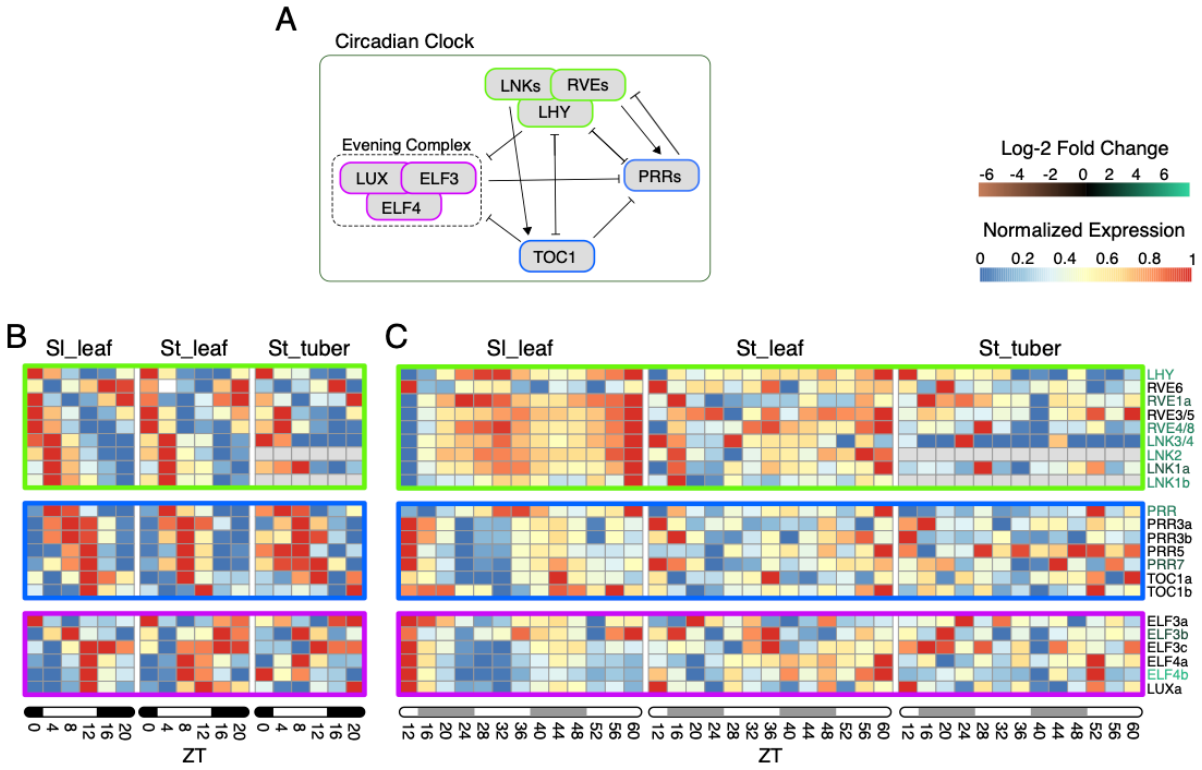

Figure 4.4: Diurnal and Circadian Rhythms Among the Core Circadian Clock Genes
(**A**). A model of the *Arabidopsis thaliana* circadian clock containing the core circadian clock genes. Arrows and bars indicate positive and negative transcriptional regulation, respectively. '*LHY*': *LATE ELONGATED HYPOCOTYL*; '*RVE*': *REVEILLES*; '*LNK*': *NIGHT LIGHT INDUCIBLE AND CLOCK-REGULATED*; '*PRR*': *PSEUDO-RESPONSE REGULATOR*; '*TOC1*': *TIMING OF CAB1*; '*ELF3*': *EARLY FLOWERING 3*; '*ELF4*': *EARLY FLOWERING 4*; '*LUX*': *LUX ARRHYTHMO*. (**B**), (**C**). Heatmaps of normalized gene expression abundances for the core circadian clock genes from the diurnal (**B**) and circadian (**C**) time course experiments for tomato leaf tissue (Sl_leaf), potato leaf (St_leaf), and potato tuber (St_tuber). The black, grey, and white bar along the bottom indicates the timing of night, subjective night, and day, respectively. Colored boxes correspond to the gene color in the model. Text color corresponds to the log-2 fold change value for leaf gene expression compared to tuber gene expression.

identified (Figure 4.4A, Table S4.4). All core clock genes were more highly expressed in leaf tissue than in tubers (Figure S4.7), with many significantly preferentially expressed in leaf tissue, including all *LHY* and *LNK* orthologs (Table S4.4). Diel expression patterns were found among all core clock orthologs in the leaf tissue (Figure 4.4B, S4.7), with the exception of a few paralogs of *RVE1*, *ELF4*, and *LUX*. These diel expression patterns are similar to the one of a previous study (Morris *et al.*, 2014). With the exception of *RVE3/5*, *PRR3a*, *PRR5*, *TOC1b*, *ELF4a*, and *LUXa*, most clock genes had very low expression level in the tubers, and only a subset of clock genes had diel expression patterns in this tissue, including *LHY*, *RVE1*, *RVE3/5*, *LNK3/4*, *PRR3*, *PRR5*, *TOCI*, *ELF3*, and *ELF4* (Figure 4.4B, S4.7). In contrast to the strong diel rhythms found in potato leaves, few clock-associated genes displayed circadian rhythmic expression in either the leaf or tuber (Figure 4.3E, S4.8). *RVE4/8*, *LNK3/4*, *PRR3*, *TOC1*, and *ELF3* displayed circadian rhythmic gene expression in the leaf, while only *RVE1a*, *RVE4/8*, *LNK1a* and *PRR3b* displayed circadian rhythmic gene expression in the tuber (Figure 4.4C, S4.8).

A previous study determined diel and circadian expression of clock associated genes in tomato leaves (Müller *et al.*, 2016). For a better direct comparison with our potato results, the tomato leaf RNA-seq data was similarly processed and analyzed. In tomato leaves, all core clock genes were both diel and circadian rhythmic (Figure 4.4B, 4.4C, S4.7, S4.8). Tomato and potato leaf tissues displayed very similar phases in LD/HC conditions, with at most a four hour difference between phases for *RVE3/5*, *PRR5*, *PRR7*, *TOC1*, *ELF3*, and *ELF4* (Figure 4.4B, S4.7). For the clock genes which displayed circadian expression patterns in potato, the respective tomato genes had a longer period and larger amplitude compared to potato under LL conditions (Figure 4.4C, S4.8), matching observations of period differences in delayed fluorescence rhythms in the leaf tissue (Hardigan *et al.*, 2017). It has been hypothesized that the tomato circadian clock was

weakened during domestication and improvement efforts (Müller *et al.*, 2016), and it is possible that a more extreme weakening of the circadian clock was selected for in potato. Indeed, *RVE3/5*, *LNK3/4*, *ELF4a*, and *ELF4b* had signatures of selection during both potato domestication and improvement (Hardigan *et al.*, 2017). All three of these genes displayed diel expression patterns in both tissues, and *LNK3/4* had circadian rhythms in leaf tissue. As potato exhibits a significant number of circadian rhythmic expression patterns despite lack of such patterns among core clock genes, it is possible that potato has an altered circadian clock such that alternative regulatory mechanisms (e.g. post-translational regulation) and/or a modified set of core clock genes are utilized to drive gene expression among downstream genes. Indeed, in *A. thaliana* morning phased clock genes were preferentially expressed in the mesophyll while evening phased clock genes were preferentially expressed in the vasculature (Endo *et al.*, 2014), and LUX mutants are arrhythmic in *A. thaliana* leaves, but circadian rhythms are present in the shoot apical meristem (Takahashi *et al.*, 2015). Future studies investigating mutants in core circadian clock genes and investigating mapping populations segregating for circadian rhythm robustness are needed to determine genes which are critical for clock function in potato.

Detached Tubers Display Robust Transcriptional Circadian Rhythms:

Only a few putative clock genes displayed significant expression levels in tubers, which might indicate the presence of only a weak or slave oscillator that might be dependent on input from autotrophic tissues (Figure S4.7). To test if circadian rhythms in the tuber are independent from above-ground tissues, we investigated attached- and detached-tuber rhythms using a luciferase reporter line. Soltu.DM.02G8028070 (hereafter referred to as *StGH3*) is annotated as a

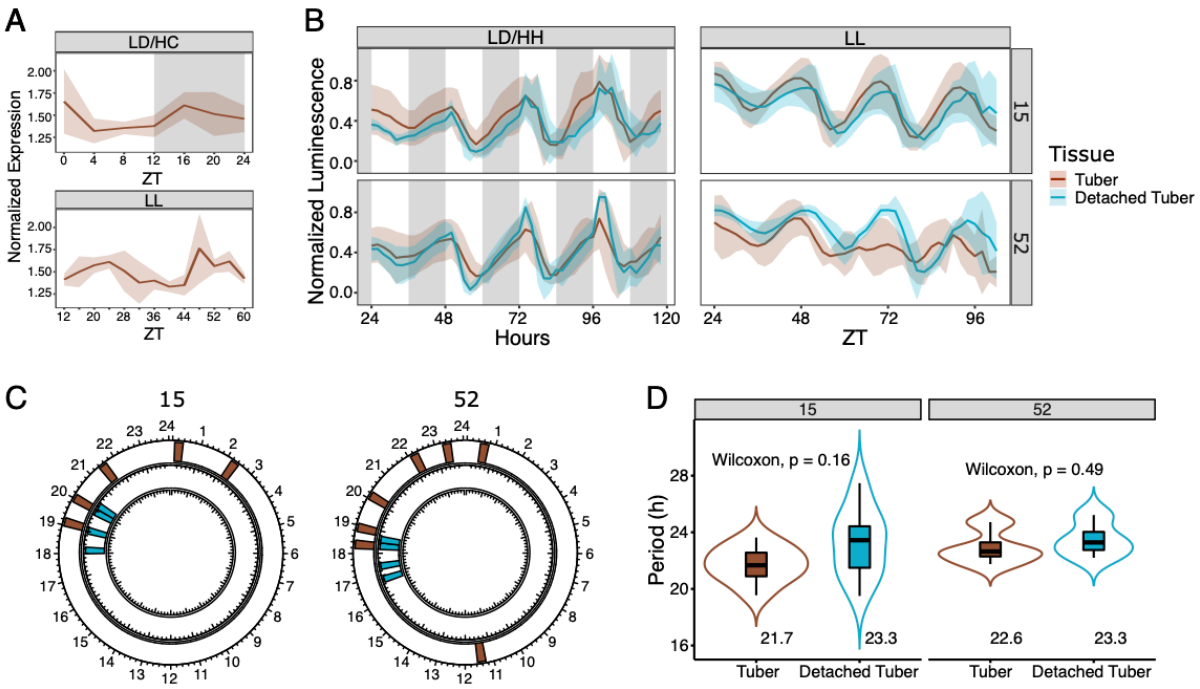

Figure 4.5: Transcriptional Diurnal and Circadian Rhythms in the Tuber
(**A**). DESeq2 rlog normalized gene expression abundances are plotted against Zeitgeber time (ZT) for both the LD/HC and LL time course experiments for *StGH3* (Soltu.DM.02G028070). The ribbon corresponds to the range of one standard deviation among the replicates. (**B**). Moving mean and standard deviation normalized luminescence from two independently transformed lines (#15 and #52) is plotted against ZT for tubers and detached tubers. Plants were imaged in cycling light/dark (LD/HH) and constant light (LL) conditions. The grey boxes indicate dark periods for the LD experiments. The ribbon corresponds to the range of one standard deviation among the replicates. (**C**). The density of phase in hours from the LD experiments is plotted in a circular manner for tubers and detached tubers separately for each line. The brown and blue colors indicate tuber and detached tuber, respectively. A total of 5 and 4 replicates measured for #15 tuber and detached tuber, respectively. A total of 7 and 4 replicates were measured each for #52 tuber and detached tuber, respectively. (**D**). The free-running period (in hours) is plotted for the tubers and detached tubers by the line. The median period for each tissue and line combination is indicated in the bottom of the plot and the p-value from Wilcoxon tests comparing the tuber and detached tuber in each line are provided. A total of 9 and 10 replicates were measured for #15 tuber and detached tuber, respectively. A total of 4 replicates were measured each for #52 tuber and detached tuber.

Gretchen Hagen 3 (GH3) auxin-responsive gene, which is a ubiquitous gene family rapidly induced upon plant treatment with auxin (Hagen and Guilfoyle, 2002). Diel and circadian rhythmic gene expression were observed in the tuber (Figure 4.5A); rhythmic RNA levels in the leaf were not observed (Figure S4.9). The promoter of *StGH3* was cloned into a firefly luciferase vector and luminescence was measured in two independently stably transformed six week old lines to test for diel and circadian transcriptional rhythms in attached and detached tubers. Both attached and detached tuber displayed robust transcriptional rhythms after four days under LD/HH and LL

conditions, with some noise present in the tuber circadian rhythms for line #52 (Figure 4.5B). Under LD/HH cycles, the median circular phase for the attached and detached tubers was 20.91 and 18.55 hours, respectively, (Figure 4.5C). This phase of the bioluminescence rhythms is almost exactly the same as reported by JTK_ECHO under LD/HC conditions for the gene expression, which was 19.80 hours, indicating all primary cis-regulatory components were captured in the promoter sequence. Under constant light, the median period was 22.10 and 23.33 hours for the attached and detached tuber, respectively (Figure 4.5D). Our results provide evidence for an internal molecular oscillator driving transcriptional rhythms in the tuber apart from any support from other plant tissues. Similar results have been found *A. thaliana* roots (Bordage *et al.*, 2016; Greenwood *et al.*, 2019) and may signify a more universal phenomena of tissue-specific circadian clocks in plants. Furthermore, these reporter lines present opportunities to profile rhythmic transcription in other tissues and developmental stages in future studies.

CONCLUSION

Through profiling global gene expression over time in potato leaf and tuber tissue, we report on the first circadian rhythmic gene expression study in the below-ground heterotrophic tuber tissue. The tuber displays a longer circadian period, has a delayed phase, and a lower amplitude compared to leaf tissue. Over 500 genes are differentially phased between the leaf and tuber, including several aquaporins. Many carbohydrate metabolism enzymes are under both diel and circadian regulation in both tissues, reflecting the importance of the circadian clock for tuber bulking. Most core circadian clock genes do not display circadian rhythmic gene expression in the leaf or tuber, yet robust transcriptional and gene expression circadian rhythms are present. Furthermore, detached tubers maintain circadian rhythms after four days, indicating tubers contain

an internal oscillator able to function independently from other plant tissues. These results hold exciting implications for other tuber and fibrous root crops, and further studies are needed to investigate how the core circadian clock is altered in potato and what changes occur in other developmental stages.

ACKNOWLEDGEMENTS

Funding for this project was provided by funds from the Michigan State University Foundation to C.R.B., Michigan State University Project GREEEN (GR19-078) to E.M.F, C.R.B, and D.S.D., and the NSF Plant Genome Research Project (IOS- 1950376) to E.M.F, C.R.B, and D.S.D.. G.M.H was supported by fellowships from the Michigan State University Plant Sciences Fellowship Program, Michigan State University Plant Biotechnology for Health and Sustainability NIH Training Program, and the USDA NIFA Predoctoral fellowship (#2020-67034-31731).

APPENDIX

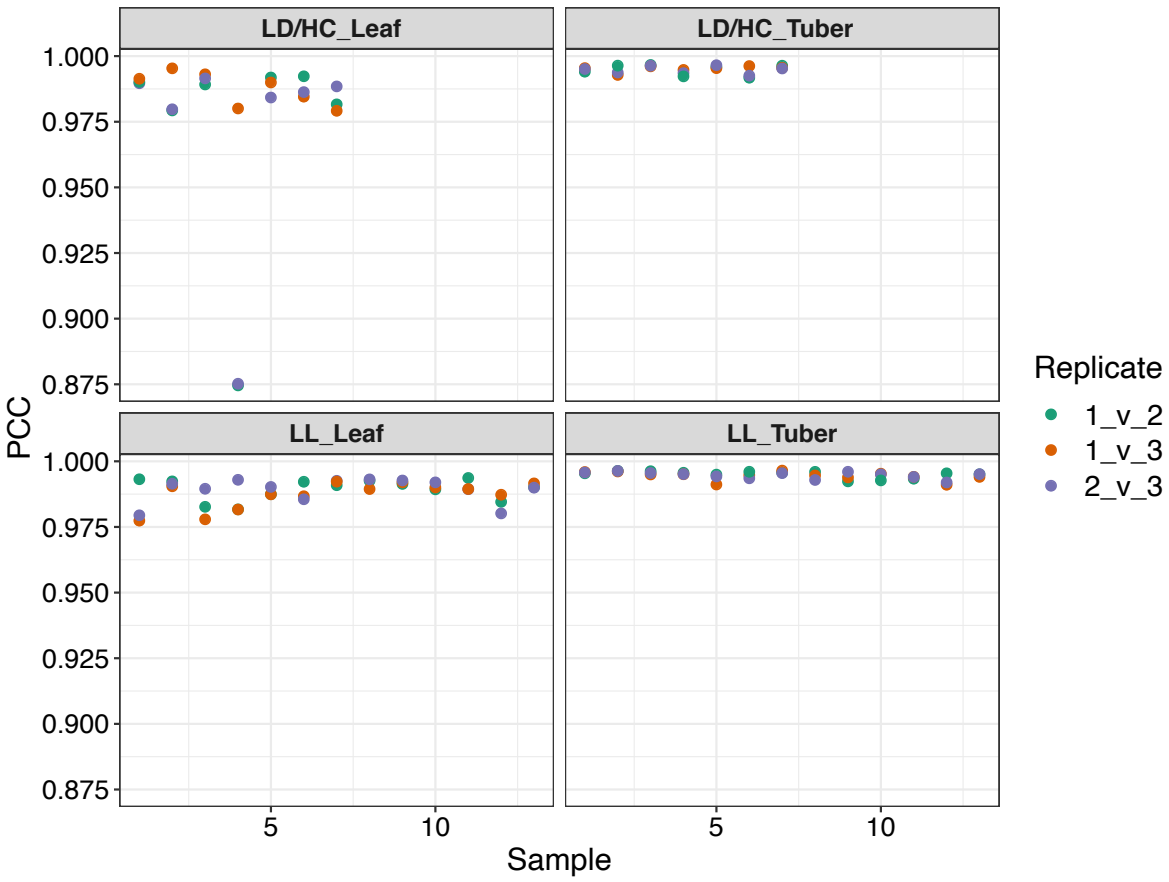

Figure S4.1: Biological Replicate Pearson's Correlation Coefficients
Pearson's Correlation Coefficients were calculated for potato leaf and tuber tissues among the sample biological replicates. Samples are separated based on tissue and experiment where LD/HC refers to cycling light and temperature conditions and LL refers to constant light and temperature conditions. Colors refer to which replicates are being compared. Sample #4 (3AM) in the LD/HC_Leaf panel has a PCC value < 0.9 for replicate 2 and was removed from further analyses.

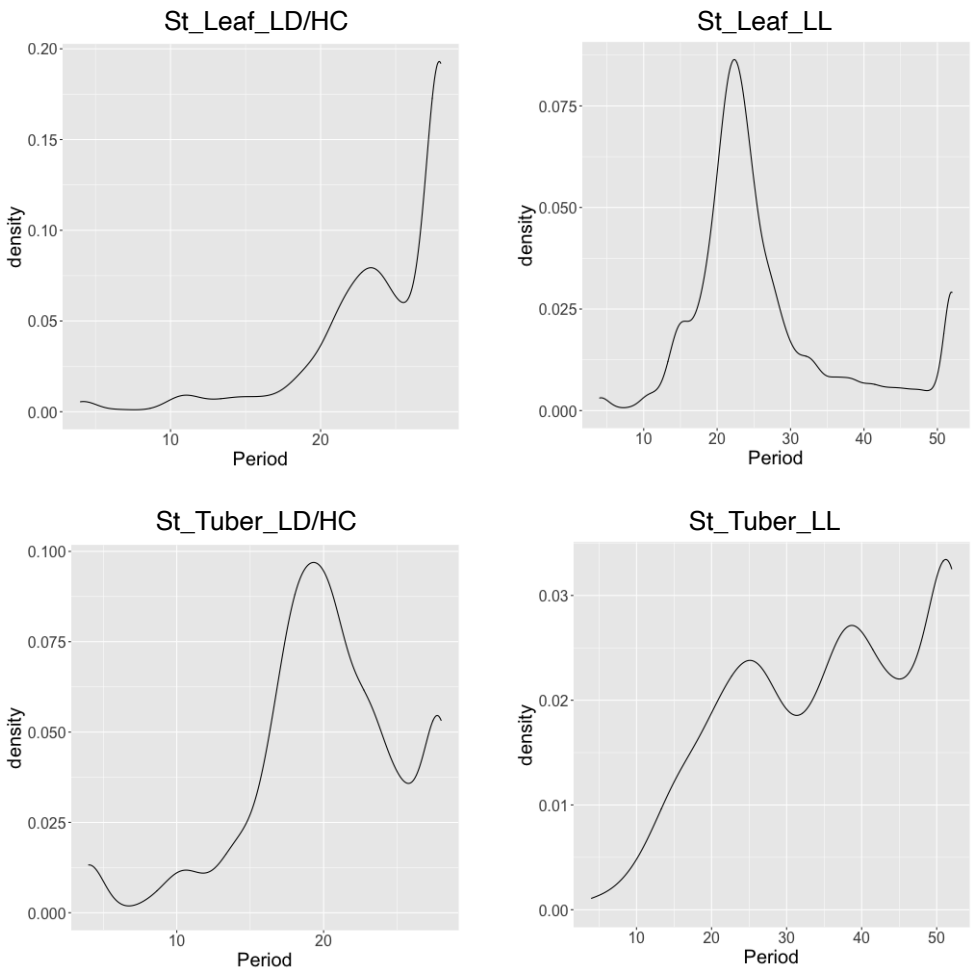

Figure S4.2: ECHO Period Distributions.
The gene-wise period length is plotted against the density of genes with significant rhythmic gene expression (BH p-value < 0.01) identified in ECHO. Separate plots are present for each species, tissue, and experiment combination. 'St' indicates potato. 'LD/HC' indicates rhythmic light and temperature conditions, while 'LL' indicates constant light and temperature conditions.

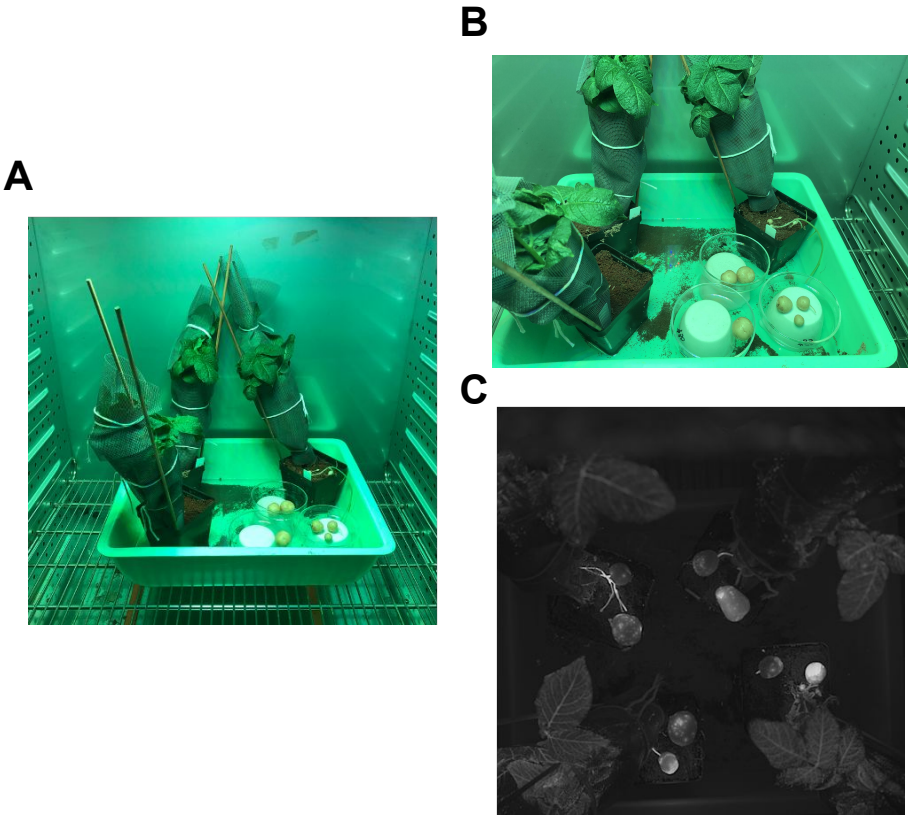

Figure S4.3: Luciferase Assays for StGH3 Lines
Pictures of the layout of the plants for luciferase assays are displayed in (**A**) and (**B**). Tubers were dug out of the soil and placed on top of the soil on the day of the assay. The shoot tissue of the plant was loosely wrapped in a screen material to prevent obstruction of imaging the tubers. An example image of the luminescence observed from the plants is present in (**C**).

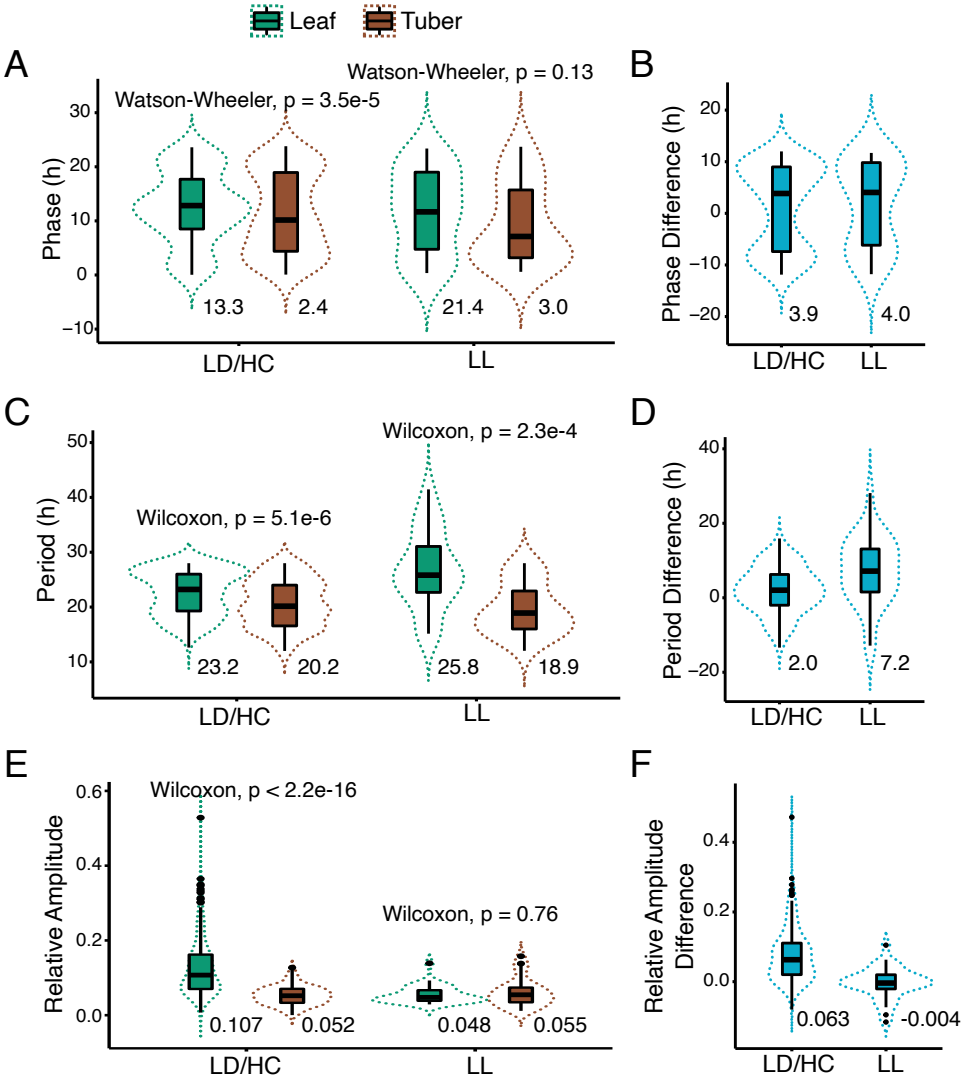

Figure S4.4: Phase, Period, and Amplitude for DiPALM Results

Violin plots overlaid with boxplots are reported for the phase (**A**), period (**C**), and relative amplitude (**E**) for the genes which displayed significant (adj. p-value < 0.01) time-dependent differences in expression patterns between leaf and tuber tissue. Data are plotted by time course experimental conditions (LD/HC for cycling light/dark and temperature and LL for constant light and temperature) and colored by tissue. Paired Wilcoxon and Watson-Wheeler tests were performed, and the p-values are reported. The gene-wise difference in phase (**B**), period (**D**), and relative amplitude (**F**) between the tissues for each experimental condition are also provided in a violin plot overlaid with a box plot. The median values are reported along the bottom of each plot.

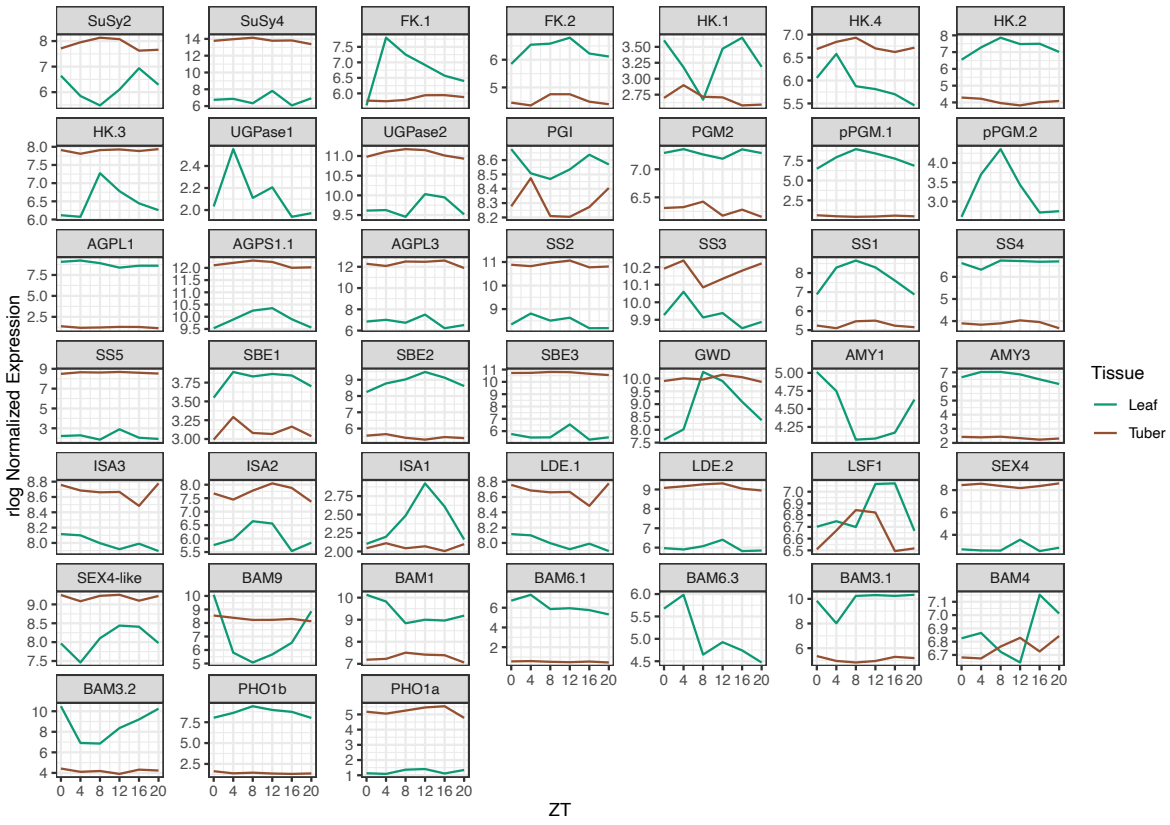

Figure S4.5: Diel Rhythms Among Carbohydrate Metabolism Enzymes

DESeq2 rlog normalized gene expression abundances are plotted against Zeitgeber time (ZT) for the LD/HC time course experiments for the carbohydrate metabolism enzymes displaying diel expression patterns. Colors correspond to tissue type. '*SuSy*': *SUCROSE SYNTHASE*; '*HK*': *HEXOKINASE*; '*FK*': *FRUCTOKINASE*; '*UGPase*': *UDP-GLUCOSE PYROPHOSPHORYLASE*; '*PGI*': *PHOSPHOGLUCOISOMERASE*; '*PGM*': *PHOSPHOGLUCOMUTASE*; '*AGPase*': *ADP-GLUCOSE PYROPHOSPHORYLASE*; '*SS*': *STARCH SYNTHASE*; '*SBE*': *STARCH BRANCHING ENZYME*; '*PWD*': *PHOSPHO-GLUCAN, WATER DIKINASE*; '*GWD*': *GLUCAN, WATER DIKINASE*; '*ISA*': *ISOAMYLASE*; '*LDE*': *LIMIT DEXTRINASE*; '*AMY*': *ALPHA-AMYLASE*; '*SEX4*': *STARCH EXCESS 4*; '*LSF*': *LIKE-STARCH EXCESS 4*; '*BAM*': *BETA-AMYLASE*; '*PHO1*': *ALPHA-GLUCAN PHOSPHORYLASE 1*; '*DPE1*': *DISPROPORTIONATING ENZYME 1*.

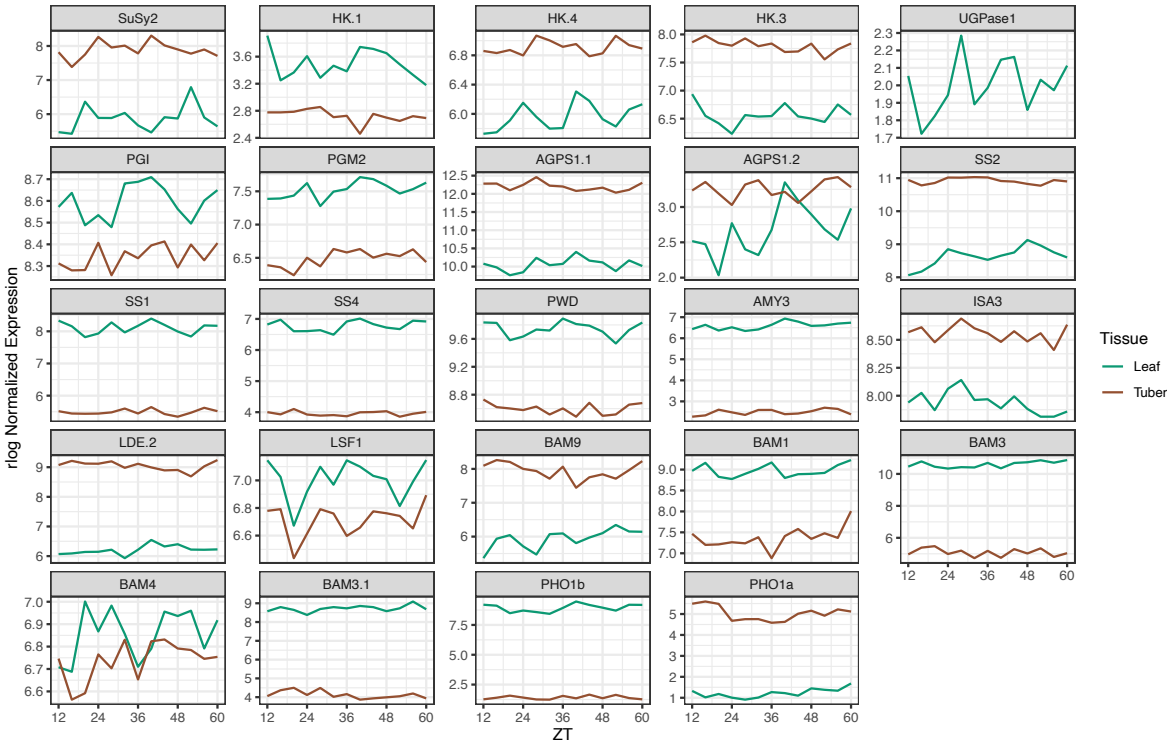

Figure S4.6: Circadian Rhythms Among Carbohydrate Metabolism Enzymes

DESeq2 rlog normalized gene expression abundances are plotted against Zeitgeber time (ZT) for the LL time course experiments for the carbohydrate metabolism enzymes displaying circadian expression patterns in at least one tissue. Colors correspond to tissue type. '*SuSy*': *SUCROSE SYNTHASE*; '*HK*': *HEXOKINASE*; '*FK*': *FRUCTOKINASE*; '*UGPase*': *UDP-GLUCOSE PYROPHOSPHORYLASE*; '*PGI*': *PHOSPHOGLUCOISOMERASE*; '*PGM*': *PHOSPHOGLUCOMUTASE*; '*AGPase*': *ADP-GLUCOSE PYROPHOSPHORYLASE*; '*SS*': *STARCH SYNTHASE*; '*SBE*': *STARCH BRANCHING ENZYME*; '*PWD*': *PHOSPHO-GLUCAN, WATER DIKINASE*; '*GWD*': *GLUCAN, WATER DIKINASE*; '*ISA*': *ISOAMYLASE*; '*LDE*': *LIMIT DEXTRINASE*; '*AMY*': *ALPHA-AMYLASE*; '*SEX4*': *STARCH EXCESS 4*; '*LSF*': *LIKE-STARCH EXCESS 4*; '*BAM*': *BETA-AMYLASE*; '*PHO1*': *ALPHA-GLUCAN PHOSPHORYLASE 1*; '*DPE1*': *DISPROPORTIONATING ENZYME 1*.

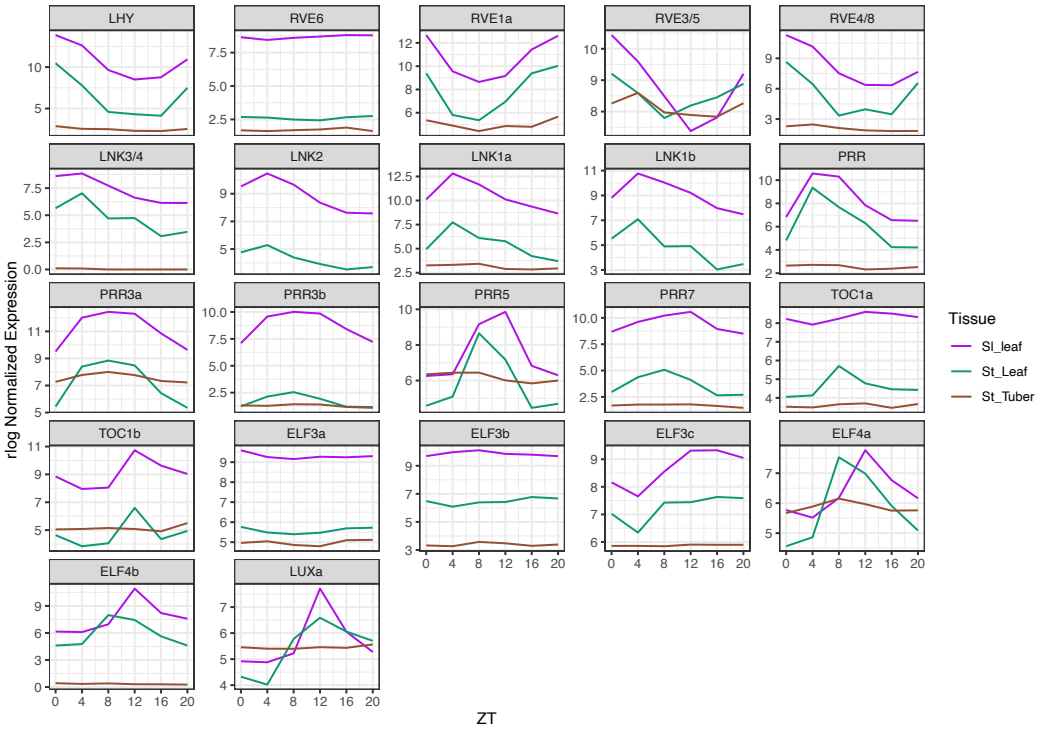

Figure S4.7: Diel Rhythms Among Core Circadian Clock Genes

DESeq2 rlog normalized gene expression abundances are plotted against Zeitgeber time (ZT) for the LD/HC time course experiments for the core circadian clock genes displaying diel expression patterns. Colors correspond to species and tissue type. 'Sl_leaf': tomato leaf; 'St_leaf: potato leaf; 'St_tuber': potato tuber. '*LHY*': *LATE ELONGATED HYPOCOTYL*; '*RVE*': *REVEILLES*; '*LNK*': *NIGHT LIGHT INDUCIBLE AND CLOCK-REGULATED*; '*PRR*': *PSEUDO-RESPONSE REGULATOR*; '*TOC1*': *TIMING OF CAB1*; '*ELF3*': *EARLY FLOWERING 3*; '*ELF4*': *EARLY FLOWERING 4*; '*LUX*': *LUX ARRHYTHMO*.

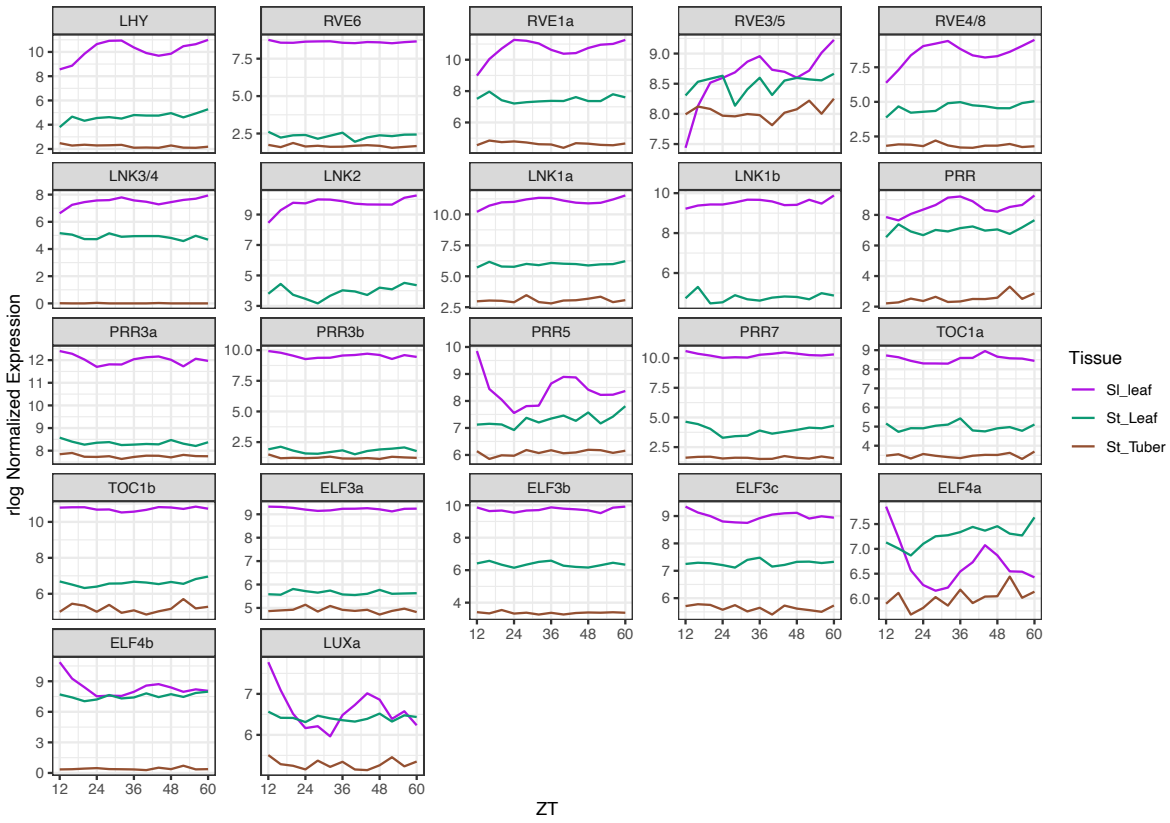

Figure S4.8: Circadian Rhythms Among Core Circadian Clock Genes

DESeq2 rlog normalized gene expression abundances are plotted against Zeitgeber time (ZT) for the LL time course experiments for the core circadian clock genes displaying diel expression patterns. Colors correspond to species and tissue type. 'Sl_leaf': tomato leaf; 'St_leaf: potato leaf; 'St_tuber': potato tuber. '*LHY*': *LATE ELONGATED HYPOCOTYL*; '*RVE*': *REVEILLES*; '*LNK*': *NIGHT LIGHT INDUCIBLE AND CLOCK-REGULATED*; '*PRR*': *PSEUDO-RESPONSE REGULATOR*; '*TOC1*': *TIMING OF CAB1*; '*ELF3*': *EARLY FLOWERING 3*; '*ELF4*': *EARLY FLOWERING 4*; '*LUX*': *LUX ARRHYTHMO*.

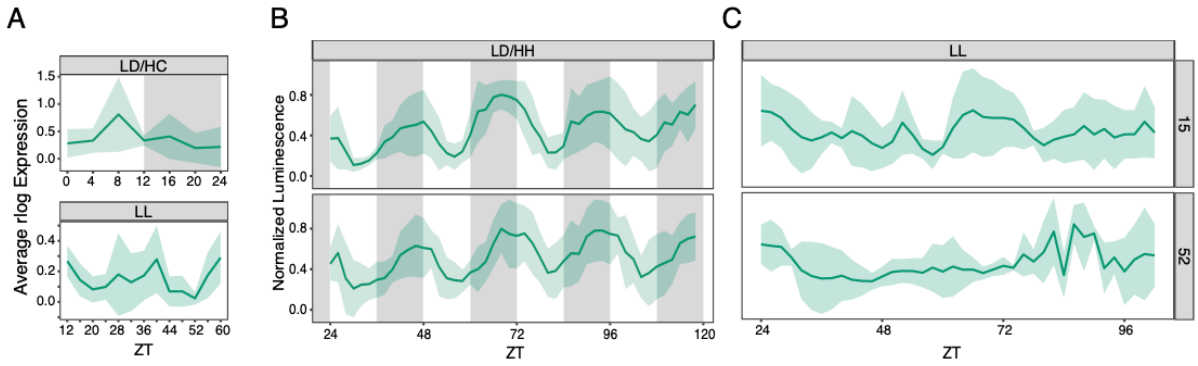

Figure S4.9: StGH3 Leaf Gene Expression and Transcriptional Patterns

(**A**). Leaf DESeq2 rlog normalized gene expression abundances are plotted against Zeitgeber time (ZT) for both the LD/HC and LL time course experiments for *StGH3* (Soltu.DM.02G028070). The ribbon corresponds to the range of one standard deviation among the replicates. (**B**). Moving mean and standard deviation normalized luminescence from two independently transformed lines (#15 and #52) is plotted against ZT for leaf tissue. Plants were imaged in cycling light/dark (LD/HH) and constant light (LL) conditions. The grey boxes indicate dark periods for the LD experiments. The ribbon corresponds to the range of one standard deviation among the replicates.

Species and Tissue	Experiment	Gene Set	COV Filter Threshold	Soft Power	Tree Cut Height	Tree Merge Height	Number of Genes in Gene Set
St_Leaf	LD/HC	COV	0.05	12	0.975	0.4	15,219
	LL	COV	0.05	9	0.975	0.4	13,355
St_tuber	LD/HC	COV	0.05	12	0.99	0.4	9,020
	LL	COV	0.05	9	0.975	0.4	8,986

Table S4.1: WGCNA Parameters for Identifying Rhythmic Gene Expression

The coefficient of variance filter threshold, soft power, tree cut height, tree merge height, and number of genes present in the gene set are included in the table for each of the gene sets used in WGCNA. 'St_leaf' refers to potato leaf tissue, 'St_tuber' refers to potato tuber tissue. 'LD/HC' experiment refers to cycling light and temperature conditions, and 'LL' experiment refers to constant light and temperature conditions. 'COV' gene set refers to genes passing the coefficient of variance filter.

Species and Tissue	Experiment	Gene Set	COV Filter Threshold	Soft Power	Tree Cut Height	Tree Merge Height	Number of Genes in Gene Set
St_Leaf	LD/HC	COV	0.05	12	0.975	0.4	7,487
	LL	COV	0.05	9	0.975	0.4	7,025
St_tuber	LD/HC	COV	0.05	15	0.975	0.4	7,487
	LL	COV	0.05	8	0.975	0.4	7,025

Table S4.2: WGCNA Parameters for DiPALM Analyses

The coefficient of variance filter threshold, soft power, tree cut height, tree merge height, and number of genes present in the gene set are included in the table for each of the gene sets used in WGCNA. 'St_leaf' refers to potato leaf tissue, 'St_tuber' refers to potato tuber tissue. 'LD/HC' experiment refers to cycling light and temperature conditions, and 'LL' experiment refers to constant light and temperature conditions. 'COV' gene set refers to genes passing the coefficient of variance filter.

Gene Set	GO Term	GO Name	Number of Annotations in All Genes	Number of Annotations in the Gene Set	Adjusted P-value	Average Log-2 Fold Change
Shared LD/HC	GO:0009768	Light harvesting in photosystem I	36	31	7.41e-30	6.88
	GO:0009765	Light harvesting	6	5	2.77e-24	6.86
	GO:0019684	Light reaction	20	2	8.21e-19	6.26
	GO:0015979	Photosynthesis	86	10	7.64e-15	6.52
Shared LL	GO:0009768	Light harvesting in photosystem I	36	19	2.20e-23	6.95
	GO:0009765	Light harvesting	6	2	8.37e-20	6.83
	GO:0019684	Light reaction	20	6	5.19e-21	5.79
	GO:0015979	Photosynthesis	86	2	1.35e-17	7.08
Tuber Unique LD/HC	GO:0015979	Photosynthesis	86	20	1.40e+31	6.58
	GO:0019684	Light Reaction	20	8	6.45e-25	6.00
	GO:0009767	Photosynthetic Electron Transport Chain	22	8	2.64e-16	6.80
Tuber Unique LL	GO:0015979	Photosynthesis	86	28	6.75e-67	6.80
	GO:0019684	Light reaction	20	4	9.65e-53	6.98

Table S4.3: Photosynthetic Related GO term enrichments

The significant gene ontology (GO) terms related to photosynthetic pathways for shared diurnal rhythmic genes (LD/HC), shared circadian rhythmic genes (LL), tuber-specific diurnal rhythmic genes (LD/HC), and tuber-specific circadian rhythmic genes (LL). The GO term, GO name, total number of genes annotated with the GO term among all genes, number of genes annotated with the GO term among the gene set, the False Discovery Rate adjusted p-value, and the average gene expression log-2 fold change for leaf tissue compared to tuber tissue among the genes annotated with the GO term in the gene set.

Gene ID	Name	Functional Annotation	Log-2 Fold Change
Soltu.DM.10G000090	LHY	Homeodomain-like superfamily protein	5.60
Soltu.DM.10G000080	RVE2	Homeodomain-like superfamily protein	5.67
Soltu.DM.06G013470	RVE6	Homeodomain-like superfamily protein	-
Soltu.DM.01G034130	RVE3/5	Homeodomain-like superfamily protein	-
Soltu.DM.02G004510	RVE1a	Homeodomain-like superfamily protein	3.32
Soltu.DM.04G007910	RVE1b	hypothetical protein	-
Soltu.DM.05G015800	RVE1c	hypothetical protein	-
Soltu.DM.11G017930	RVE1d	hypothetical protein	-
Soltu.DM.10G023770	RVE4/8	Homeodomain-like superfamily protein	4.63
Soltu.DM.01G035820	LNK3/4	conserved hypothetical protein	5.20
Soltu.DM.04G005260	LNK1b	dentin sialophosphoprotein-related	5.62
Soltu.DM.01G044350	LNK1a	dentin sialophosphoprotein-related	2.79
Soltu.DM.01G025720	LNK2	conserved hypothetical protein	5.32
Soltu.DM.03G021500	PRR5	pseudo-response regulator	-
Soltu.DM.10G021220	PRR7	pseudo-response regulator	2.29
Soltu.DM.10G000040	PRR	pseudo-response regulator	4.58
Soltu.DM.04G018960	PRR3a	pseudo-response regulator	-
Soltu.DM.04G018970	PRR3b	pseudo-response regulator	-
Soltu.DM.06G025760	TOC1a	CCT motif -containing response regulator protein	-
Soltu.DM.03G029900	TOC1b	CCT motif -containing response regulator protein	-
Soltu.DM.08G014030	ELF3c	hydroxyproline-rich glycoprotein family protein	-
Soltu.DM.11G023720	ELF3a	hydroxyproline-rich glycoprotein family protein	-
Soltu.DM.12G004460	ELF3b	hydroxyproline-rich glycoprotein family protein	2.81
Soltu.DM.06G009180	ELF4b	Protein of unknown function (DUF1313)	6.87
Soltu.DM.06G009240	ELF4a	Protein of unknown function (DUF1313)	-
Soltu.DM.06G009220	ELF4c	Protein of unknown function (DUF1313)	2.15
Soltu.DM.06G003930	LUXb	Homeodomain-like superfamily protein	2.16
Soltu.DM.06G031660	LUXa	Homeodomain-like superfamily protein	-

Table S4.4: Core Circadian Clock Orthologs in Potato

The gene ID, clock gene name, functional annotation, and log-2 fold change for the potato orthologs of *Arabidopsis thaliana* core circadian clock genes are provided. The log-2 fold change is for the comparison of leaf v. tuber tissue with positive numbers indicating higher expression in the leaf. '-' indicates no significant difference in expression (adjusted p-value < 0.01) and/or the log-2 fold change was not greater than 2 or less than -2.

REFERENCES

REFERENCES

Baker, D.A. and Moorby, J. (1969) The Transport of Sugar , Water , and Ions into Developing Potato Tubers. *Ann. Bot.*, **33**, 729–741.

Bläsing, O.E., Gibon, Y., Günther, M., et al. (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. *Plant Cell*, **17**, 3257–3281.

Bolger, A., Scossa, F., Bolger, M.E., et al. (2014) The genome of the stress-tolerant wild tomato species Solanum pennellii. *Nat. Genet.*, **46**, 1034–1038.

Bombarely, A., Moser, M., Amrad, A., et al. (2016) Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nat. Plants*, **2**, 1–9.

Bordage, S., Sullivan, S., Laird, J., Millar, A.J. and Nimmo, H.G. (2016) Organ specificity in the plant circadian system is explained by different light inputs to the shoot and root clocks. *New Phytol.*, **212**, 136–149.

Butt, G.R., Qayyum, Z.A. and Jones, M.A. (2020) Plant defence mechanisms are modulated by the circadian system. *Biology (Basel).*, **9**, 454.

Chen, W.W., Takahashi, N., Hirata, Y., Ronald, J., Porco, S., Davis, S.J., Nusinow, D.A., Kay, S.A. and Mas, P. (2020) A mobile ELF4 delivers circadian temperature information from shoots to roots. *Nat. Plants*, **6**, 416–426.

Covington, M.F. and Harmer, S.L. (2007) The circadian clock regulates auxin signaling and responses in Arabidopsis. *PLoS Biol.*, **5**, e222.

Dodd, A.N., Kusakina, J., Hall, A., Gould, P.D. and Hanaoka, M. (2014) The circadian regulation of photosynthesis. *Photosynth. Res.*, **119**, 181–190.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.*, **45**, 616–629.

Edwards, K.D., Anderson, P.E., Hall, A., Salathia, N.S., Locke, J.C.W., Lynn, J.R., Straume, M., Smith, J.Q. and Millar, A.J. (2006) FLOWERING LOCUS C mediates natural variation in the high-temperature response of the Arabidopsis circadian clock. *Plant Cell*, **18**, 639–650.

Emms, D.M. and Kelly, S. (2019) OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.*, **20**, 238.

Endo, M., Shimizu, H., Nohales, M.A., Araki, T. and Kay, S.A. (2014) Tissue-specific clocks in Arabidopsis show asymmetric coupling. *Nature*, **515**, 419–422.

Filichkin, S.A., Breton, G., Priest, H.D., et al. (2011) Global Profiling of Rice and Poplar Transcriptomes Highlights Key Conserved Circadian-Controlled Pathways and cis-Regulatory Modules. *PLoS One*, **6**, e16907.

Geigenberger, P. and Stitt, M. (2000) Diurnal changes in sucrose, nucleotides, starch synthesis and AGPS transcript in growing potato tubers that are suppressed by decreased expression of sucrose phosphate synthase. *Plant J.*, **23**, 795–806.

Graf, A., Schlereth, A., Stitt, M. and Smith, A.M. (2010) Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. *Proc. Natl. Acad. Sci.*, **107**, 9458–9463.

Greenham, K. and McClung, C.R. (2015) Integrating circadian dynamics with physiological processes in plants. *Nat. Rev. Genet.*, **16**, 598–610.

Greenham, K., Sartor, R.C., Zorich, S., Lou, P., Mockler, T.C. and McClung, C.R. (2020) Expansion of the circadian transcriptome in Brassica rapa and genome-wide diversification of paralog expression patterns. *Elife*, **9**, e58993.

Greenwood, M., Domijan, M., Gould, P.D., Hall, A.J.W. and Locke, J.C.W. (2019) Coordinated circadian timing through the integration of local inputs in Arabidopsis thaliana. *PLoS Biol.*, **17**, e3000407.

Hagen, G. and Guilfoyle, T. (2002) Auxin-responsive gene expression: Genes, promoters and regulatory factors. *Plant Mol. Biol.*, **49**, 373–385.

Hardigan, M.A., Laimbeer, F.P.E., Newton, L., et al. (2017) Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci.*, **114**, E9999–E10008.

Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A. and Kay, S.A. (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science (80-.).*, **290**, 2110–2113.

Harsselaar, J.K. Van, Lorenz, J., Senning, M., Sonnewald, U. and Sonnewald, S. (2017) Genome-wide analysis of starch metabolism genes in potato (Solanum tuberosum L.). *BMC Genomics*, **18**, 37.

Hayes, K.R., Beatty, M., Meng, X., Simmons, C.R., Habben, J.E. and Danilevskaya, O.N. (2010) Maize global transcriptomics reveals pervasive leaf diurnal rhythms but rhythms in developing ears are largely limited to the core oscillator. *PLoS One*, **5**, e12887.

Higashi, T., Aoki, K., Nagano, A.J., Honjo, M.N. and Fukuda, H. (2016) Circadian oscillation of the lettuce transcriptome under constant light and light-dark conditions. *Front. Plant Sci.*, **7**, 1114.

Higuchi, T., Suga, S., Tsuchiya, T., Hisada, H., Morishima, S., Okada, Y. and Maeshima, M. (1998) Molecular cloning, water channel activity and tissue specific expression of two isoforms of radish vacuolar aquaporin. *Plant Cell Physiol.*, **39**, 905–913.

Hoover R. (2001) Composition, molecular structure, and physicochemical properties of tuber and root starches: a review. *Carbohydr. Polym.*, **45**, 253–267.

Hosmani, P.S., Flores-Gonzalez, M., Geest, H. van de, et al. (2019) An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *bioRxiv*, doi:10.1101/767764.

Janse van Rensburg, H.C. and Ende, W. Van den (2018) UDP-glucose: A potential signaling molecule in plants? *Front. Plant Sci.*, **8**, 2230.

Jones, M.A., Hu, W., Litthauer, S., Lagarias, J.C. and Harmer, S.L. (2015) A constitutively active allele of phytochrome B maintains circadian robustness in the absence of light. *Plant Physiol.*, **169**, 814–825.

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, **37**, 907–915.

Kötting, O., Kossmann, J., Zeeman, S.C. and Lloyd, J.R. (2010) Regulation of starch metabolism: The age of enlightenment? *Curr. Opin. Plant Biol.*, **13**, 321–329.

Lamesch, P., Berardini, T.Z., Li, D., et al. (2012) The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, 1202–1210.

Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Li, W., Zarka, K.A., Douches, D.S., Coombs, J.J., Pett, W.L. and Grafius, E.J. (1999) Coexpression of potato PVY(o) coat protein and cry V-Bt genes in potato. *J. Am. Soc. Hortic. Sci.*, **124**, 218–223.

Liao, Y., Smyth, G.K. and Shi, W. (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

Lin, T., Zhu, G., Zhang, J., et al. (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.*, **46**, 1220–1226.

Los Santos, H. De, Collins, E.J., Mann, C., Sagan, A.W., Jankowski, M.S., Bennett, K.P. and Hurley, J.M. (2020) ECHO: An application for detection and analysis of oscillators identifies metabolic regulation on genome-wide circadian output. *Bioinformatics*, **36**, 773–781.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

Lu, Y., Gehan, J.P. and Sharkey, T.D. (2005) Daylength and circadian effects on starch degradation and maltose metabolism. *Plant Physiol.*, **138**, 2280–2291.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.

Matsuda, H., Nakayasu, M., Aoki, Y., Yamazaki, S., Nagano, A., Yazaki, K. and Sugiyama, A. (2020) Diurnal metabolic regulation of isoflavones and soyasaponins in soybean roots. *Plant Direct*, **4**, e00286.

Michael, T.P., Mockler, T.C., Breton, G., et al. (2008) Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet.*, **4**, e14.

Mockler, T.C., Michael, T., Priest, H., Shen, R., Sullivan, C., Givan, S., McEntee, C., Kay, S. and Chory, J. (2007) The Diurnal Project : Diurnal and Circadian Expression Profiling , Model-based Pattern Matching , and Promoter Analysis. *Cold Spring Harb Symp Quant Biol*, **72**, 353–363.

Morris, W.L., Hancock, R.D., Ducreux, L.J.M., et al. (2014) Day length dependent restructuring of the leaf transcriptome and metabolome in potato genotypes with contrasting tuberization phenotypes. *Plant, Cell Environ.*, **37**, 1351–1363.

Müller, L.M., Mombaerts, L., Pankin, A., Davis, S.J., Webb, A.A.R., Goncalves, J. and Korff, M. Von (2020) Differential effects of day/night cues and the circadian clock on the barley transcriptome. *Plant Physiol.*, **183**, 765–779.

Müller, N.A., Wijnen, C.L., Srinivasan, A., et al. (2016) Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Genet.*, **48**, 89–93.

Nakamichi, N. (2020) The transcriptional network in the arabidopsis circadian clock system. *Genes (Basel).*, **11**, 1284.

Paulina Aguilera-Alvarado, G. and Sanchez-Nieto, S. (2017) Plant Hexokinases are Multifaceted Proteins. *Plant Cell Physiol.*, **58**, 1151–1160.

Pérez-Torres, E., Kirchgessner, N., Pfeifer, J. and Walter, A. (2015) Assessing potato tuber diel growth by means of X-ray computed tomography. *Plant Cell Environ.*, **38**, 2318–2326.

Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J. and Buell, C.R. (2020) Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience*, **9**, giaa100.

Pilling, E. and Smith, a M. (2003) Growth ring formation in the starch granules of potato tubers. *Plant Physiol*, **132**, 365–371.

Poiré, R., Wiese-Klinkenberg, A., Parent, B., Mielewczik, M., Schurr, U., Tardieu, F. and 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.

Rienth, M., Torregrosa, L., Kelly, M.T., Luchaire, N., Pellegrino, A., Grimplet, J. and Romieu, C. (2014) Is transcriptomic regulation of berry development more important at night than during the day? *PLoS One*, **9**, e88844.

Schnieders, B.J., Kerckhoffs, L.H.J. and Struik, P.C. (1988) Diel changes in tuber volume. *Potato Res.*, **31**, 129–135.

Sharma, S.K., Bolser, D., Boer, J. de, et al. (2013) Construction of reference chromosome-scale pseudomolecules for potato: Integrating the potato genome with genetic and physical maps. *G3 Genes, Genomes, Genet.*, **3**, 2031–2047.

Shim, J.S., Kubota, A. and Imaizumi, T. (2017) Circadian clock and photoperiodic flowering in arabidopsis: CONSTANS is a Hub for Signal integration. *Plant Physiol.*, **173**, 5–15.

Skryhan, K., Gurrieri, L., Sparla, F., Trost, P. and Blennow, A. (2018) Redox regulation of starch metabolism. *Front. Plant Sci.*, **9**, 1344.

Smith, S.M., Fulton, D.C., Chia, T., Thorneycroft, D., Chapple, A., Dunstan, H., Hylton, C., Zeeman, S.C. and Smith, A.M. (2004) Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in arabidopsis leaves. *Plant Physiol.*, **136**, 2687–2699.

Takahashi, N., Hirata, Y., Aihara, K. and Mas, P. (2015) A Hierarchical Multi-oscillator Network Orchestrates the Arabidopsis Circadian System. *Cell*, **163**, 148–159.

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189–195.

Vandenbrink, J.P., Brown, E.A., Harmer, S.L. and Blackman, B.K. (2014) Turning heads: The biology of solar tracking in sunflower. *Plant Sci.*, **224**, 20–26.

Vavrdová, T., Samaj, J. and Komis, G. (2019) Phosphorylation of plant microtubule-associated proteins during cell division. *Front. Plant Sci.*, **10**, 238.

Voß, U., Wilson, M.H., Kenobi, K., et al. (2015) The circadian clock rephases during lateral root organ initiation in Arabidopsis thaliana. *Nat. Commun.*, **6**, 7641.

Wang, Y., Diehl, A., Wu, F., Vrebalov, J., Giovannoni, J., Siepel, A. and Tanksley, S.D. (2008) Sequencing and Comparative Analysis of a Conserved Syntenic Segment in the Solanaceae. *Genetics*, **180**, 391–408.

Welsh, D.K., Imaizumi, T. and Kay, S.A. (2005) Real-Time Reporting of Circadian-Regulated Gene Expression by Luciferase Imaging in Plants and Mammalian Cells. *Methods Enzymol.*, **393**, 269–288.

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

Wu, G., Anafi, R.C., Hughes, M.E., Kornacker, K. and Hogenesch, J.B. (2016) MetaCycle: An integrated R package to evaluate periodicity in large scale data. *Bioinformatics*, **32**, 3351–3353.

Yanovsky, M.J., Izaguirre, M., Wagmaister, J.A., Gatz, C., Jackson, S.D., Thomas, B. and Casal, J.J. (2000) Phytochrome A resets the circadian clock and delays tuber formation under long days in potato. *Plant J.*, **23**, 223–232.

Zielinski, T., Moore, A.M., Troup, E., Halliday, K.J. and Millar, A.J. (2014) Strengths and limitations of period estimation methods for circadian data. *PLoS One*, **9**.

CHAPTER 5

CONCLUDING REMARKS

CARDENOLIDE BIOSYNTHESIS IN CALOTROPIS GIGANTEA

The medicinal properties of cardenolides and other cardiac glycosides produced by *Calotropis gigantea* (L.) W.T.Aiton have propelled research in *C. gigantea* and other species in the genus (Pandey *et al.*, 2016; Koch *et al.*, 2020). The genome assembly and annotation for *C. gigantea* presented here provide an important resource for the investigation of the enzymes involved in the biosynthetic pathways (Hoopes *et al.*, 2018), and recent advances in sequencing technology and approaches could further facilitate identification of enzymes involved in cardenolide biosynthesis. For example, a chromosome-scale assembly generated with third generation sequencing technologies, such as Oxford Nanopore Technologies, would enable a more accurate investigation of metabolic gene clusters for cardenolide biosynthesis. Furthermore, establishing a comprehensive developmental gene atlas supported by metabolomics would facilitate enzyme discovery through co-expression and regulatory network analyses. Tissue- and cell-specific production of specialized secondary metabolites has commonly been observed in the Apocynaceae family and in *C. gigantea*, cardenolides have been found in roots (Kiuchi *et al.*, 1998), leaves (Seeka and Sutthivaiyakit, 2010), and latex (Ishnava *et al.*, 2012), suggesting the involvement of specialized cell types in cardenolides. Profiling gene expression and metabolites in tissues at multiple developmental stages and at the single cell level would facilitate characterization of the localization and of the enzymes in the biosynthetic pathway, thereby enabling reconstitution of the pathway in heterologous organisms for cardenolide production.

MAIZE PRESENCE-ABSENCE VARIANT EXPRESSION PATTERNS

Maize (*Zea mays* L.) is globally the second most produced food crop after sugar cane (http://www.fao.org/faostat/en/#home). Through the development of an expanded gene expression atlas across developmental stages and multiple stress conditions for the reference B73 genotype, I have generated an important resource for the functional annotation of genes (Hoopes *et al.*, 2019). These analyses have enabled a robust characterization of presence-absence variants (PAVs) and confirmed previous observations that PAVs are lowly expressed and likely functioning in environmental adaptation (Hoopes *et al.*, 2019). Recent advancements in the generation of chromosome-scale genome assemblies for multiple maize inbreds (Springer *et al.*, 2018; Sun *et al.*, 2018; Yang *et al.*, 2019; Hufford *et al.*, 2021) have shifted the field to focus on accession-specific genes and enabled the identification of PAVs without the reliance on a single reference genome assembly. Resources beyond a genome assembly are still needed to study maize in an accession-specific manner, including the development of gene expression atlases profiling a multitude of developmental stages and environmental conditions. Generation of co-expression networks coupled with phenotypic measurements for each accession would facilitate the characterization of genotype-specific gene correlations and pathways thereby providing evidence of how PAVs interact with the core genes which are present in all accessions.

TISSUE-SPECIFIC CIRCADIAN CLOCKS IN POTATO

The circadian clock regulates numerous physiological processes in plants and the identification of robust independent circadian rhythms in the below-ground specialized heterotrophic tuber tissue is a critical new finding in the field. Future studies are needed to investigate the lack of circadian expression patterns among core circadian clock genes. Allelic

variation was not accounted for as the haploid reference genome assembly for potato was used in our study and it is possible the four alleles at each locus have differential expression patterns, masking rhythms when combining expression data to a single representative allele such as that in the reference genome for potato. Indeed, extensive preferential allele expression has been identified in tetraploid potato (Pham *et al.*, 2017) and a phased, haplotype-resolved genome assembly and annotation for tetraploid potato are currently being generated that would enable quantification of expression patterns for each allele. Generation of core clock mutants and subsequent characterization of tissue-specific circadian rhythms would provide further evidence of the necessity of clock genes in each tissue. Mobile signals are known to contribute to tuberization in potato (Hannapel *et al.*, 2017) and to coordinate tissue-specific clocks in *Arabidopsis thaliana* (L.) Heynh. (Chen *et al.*, 2020). Grafting experiments between wild type plants and arrhythmic mutants would facilitate identification of potential mobile elements coordinating the clocks between the leaf and tuber.

CONCLUSION

As a rapidly evolving field, genomics has revolutionized how plant biology is investigated and enabled countless discoveries in plants. Genome assemblies have provided contextual information to genomic DNA sequence, enabling identification of physically clustered genes in metabolic gene clusters, and mRNA quantification in multiple tissues under different environmental contexts which have facilitated characterization of molecular pathways and regulatory networks, such as the circadian clock. Accession-specific genome assemblies have provided rich data resources to characterize shared and unique sequences, and reference genome assemblies are expected to become obsolete as accession-specific assemblies become more commonplace due to lower sequencing costs. Similarly, localizing mRNA abundance at the single cell level is expected to become the predominant method to quantify expression levels.

REFERENCES

REFERENCES

Chen, W.W., Takahashi, N., Hirata, Y., Ronald, J., Porco, S., Davis, S.J., Nusinow, D.A., Kay, S.A. and Mas, P. (2020) A mobile ELF4 delivers circadian temperature information from shoots to roots. *Nat. Plants*, **6**, 416–426.

Hannapel, D.J., Sharma, P., Lin, T. and Banerjee, A.K. (2017) The multiple signals that control tuber formation. *Plant Physiol.*, **174**, 845–856.

Hoopes, G.M., Hamilton, J.P., Kim, J., Zhao, D., Wiegert-Rininger, K., Crisovan, E. and Buell, C.R. (2018) Genome assembly and annotation of the medicinal plant Calotropis gigantea, a producer of anticancer and antimalarial cardenolides. *G3 Genes, Genomes, Genet.*, **8**, 385–391.

Hoopes, G.M., Hamilton, J.P., Wood, J.C., Esteban, E., Pasha, A., Vaillancourt, B., Provart, N.J. and Buell, C.R. (2019) An updated gene atlas for maize reveals organ-specific and stress-induced genes. *Plant J.*, **97**, 1154–1167.

Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., et al. (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *bioRxiv*, 2021.01.14.426684.

Ishnava, K.B., Chauhan, J.B., Garg, A.A. and Thakkar, A.M. (2012) Antibacterial and phytochemical studies on Calotropis gigantia (L.) R. Br. latex against selected cariogenic bacteria. *Saudi J. Biol. Sci.*, **19**, 87–91.

Kiuchi, F., Fukao, Y., Maruyama, T., Obata, T., Tanaka, M., Sasaki, T., Mikage, M., Haque, M.E. and Tsuda, Y. (1998) Cytotoxic Principles of a Bangladeshi Crude Drug, Akond Mul (Roots of Calotropis gigantea L.). *Chem. Pharm. Bull. (Tokyo).*, **46**, 528–530.

Koch, V., Nieger, M. and Bräse, S. (2020) Towards the synthesis of calotropin and related cardenolides from 3-epiandrosterone: A-ring related modifications. *Org. Chem. Front.*, **7**, 2670–2681.

Pandey, A., Swarnkar, V., Pandey, T., Srivastava, P., Kanojiya, S., Mishra, D.K. and Tripathi, V. (2016) Transcriptome and Metabolite analysis reveal candidate genes of the cardiac glycoside biosynthetic pathway from Calotropis procera. *Sci. Rep.*, **6**, 34464.

Pham, G.M., Hamilton, J.P., Wood, J.C., Burke, J.T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J. and Buell, C.R. (2020) Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience*, **9**, giaa100.

Pham, G.M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D.S. and Buell, C.R. (2017) Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J.*, **92**, 624–637.

Seeka, C. and Sutthivaiyakit, S. (2010) Cytotoxic cardenolides from the leaves of Calotropis gigantea. *Chem. Pharm. Bull.*, **58**, 725–728.

Springer, N.M., Anderson, S.N., Andorf, C.M., et al. (2018) The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.*, **50**, 1282–1288.

Sun, S., Zhou, Y., Chen, J., et al. (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.*, **50**, 1289–1295.

Yang, N., Liu, J., Gao, Q., et al. (2019) Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.*, **51**, 1052–1059.