Title
Genome Scale Reverse Genetics Approaches for Novel Gene Discovery in Arabidopsis thaliana Circadian Clock Transcriptional Networks

Permalink
https://escholarship.org/uc/item/8r8546g4

Author
Nelson, Jeffrey Douglas

Publication Date
2013

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Genome Scale Reverse Genetics Approaches for Novel Gene Discovery in
*Arabidopsis thaliana* Circadian Clock Transcriptional Networks

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Biology

by

Jeffrey Douglas Nelson

Committee in charge:
Professor Martin Yanofsky, Chair
Professor Joseph Ecker, Co-Chair
Professor Steve Kay, Co-Chair
Professor Michael Gorman
Professor Donald Helinski

2013
The Dissertation of Jeffrey Douglas Nelson is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

______________________________
Co-Chair

______________________________
Co-Chair

______________________________
Chair

University of California, San Diego

2013
DEDICATION

To my love, my joy, my muse, my wife Crystal
Thank you for being my partner in life,
I couldn't have done this without you.

To my family: Doug, Karen, Ryan & Ben
Thank you for your boundless love and support,
You’ve helped shape me into the man I am today.
An artist must possess nature.
He must identify himself with her rhythm,
by efforts that will prepare the mastery which will
later enable him to express himself in his own language.

Henri Matisse
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Epigraph</td>
<td>v</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>x</td>
</tr>
<tr>
<td>Vita</td>
<td>xii</td>
</tr>
<tr>
<td>Abstract of the Dissertation</td>
<td>xiv</td>
</tr>
<tr>
<td>Chapter 1 Circadian Rhythms Throughout Biological Systems</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2 Creation of a Complete Collection of <em>Arabidopsis</em> Transcription Factors</td>
<td>10</td>
</tr>
<tr>
<td>Chapter 3 Overexpression Screen for Novel <em>Arabidopsis</em> Transcriptional Regulators</td>
<td>35</td>
</tr>
<tr>
<td>References</td>
<td>61</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 2.1 Transcription factor family abundance in *Arabidopsis* vs. human. 26
Figure 2.2 Flow chart representation of TF library and promoter collections. 27
Figure 2.3 An overview of the “Promoter Hiking” approach. 28
Figure 2.4 Automated platform for library construction and yeast one-hybrid screens. 29
Figure 2.5 Schematic for cloning transcription factors. 30
Figure 2.6 Representation of cloned and not cloned genes per TF-family. 31
Figure 2.7 Results of yeast one-hybrid screen of TF library against the promoter of CCA1. 32
Figure 2.8 FBH1 identified in yeast one-hybrid screen. 33
Figure 3.1 Initial workflow for transcription factor overexpression screen. 49
Figure 3.2 LHY overexpression screen results. 50
Figure 3.3 COL9 overexpression screen results. 51
Figure 3.4 Additional COL9 overexpression results. 52
Figure 3.5 BBX31 overexpression screen results. 53
Figure 3.6 Additional BBX31 overexpression results. 54
Figure 3.7 Flowering time in BBX31 overexpression plants. 55
Figure 3.8 Quantification of flowering time delay in BBX31 OX plants and *bbx31* mutant plants relative to controls. 56
Figure 3.9 LHY expression patterns are altered in BBX31 overexpression and *bbx31* mutant plants..................................................................................................................................... 57

Figure 3.10 BBX31 transcription factor interactors in TF library yeast two-hybrid screen........................................................................................................................................ 58

Figure 3.11 Model for regulatory role of BBX31 in clock and flowering time pathways.................................................................................................................................................. 60
LIST OF TABLES

Table 2.1 Application ready TF collections ........................................... 34

Table 3.1 List of notable hits from the BBX31 genome wide transcription factor yeast two-hybrid screen ................................................................. 59
AKNOWLEDGMENTS

I would like to acknowledge Dr. Steve Kay, his vision and leadership have been tremendous and inspirational. He has given me optimism in the potential for biology to help overcome the many difficult issues facing humanity in the 21st century.

I would also like to thank Dr. Joe Ecker. His mentorship and support were crucial, especially during difficult transitional times. In addition, his curiosity and dedication to pushing the boundaries of science and applying state of the art technologies to answer big scientific questions will guide me for the rest of my career.

More thanks go out to the other members of my dissertation committee, Dr. Marty Yanofsky, Dr. Michael Gorman, and Dr. Don Helinski for their ever-present support and guidance. I especially would like to acknowledge Don Helinski, he was a tireless mentor, advocate, and friend throughout.

I also want to thank the members of the Kay and Ecker labs, past and present, for their discussions, assistance, and friendship over the years. I’m very honored to have shared a lab with such diverse, talented, and generous scientists. In particular, Dr. Ghislain Breton and Dr. Jose Pruneda-Paz were instrumental in launching the research program undertaken herein. I am looking forward to following all of their illustrious scientific careers in the future.

This process would not have been nearly as enjoyable without the dear friends I made along the way. Dr. Ryan Tewhey, Dr. Anil Narasimha, Juan
Sebastian Gomez-Cavasos, Jonathan Goodwin, and Matt Zones were incredibly important to my life outside the lab. I’m so glad to have shared this experience with such great men and brilliant scientists.

I want to thank my family for all their love, care, and encouragement. While it has been challenging to be so geographically distant from them over these years, they’ve continued to be a strong source of support and peace that I trust and rely on immensely.

Lastly, I need to thank my wife Crystal. She has been my heart, my companion, and my dancing partner for the last 8 years. I would be a much less happy and fulfilled man were it not for her presence in my life.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Pruneda-Paz, Jose; Breton, Ghislain; Nelson, Jeffrey; Ecker, Joseph; Kay, Steve. The dissertation author is a research contributor and co-author of this work.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Nelson, Jeffrey; Breton, Ghislain; Ecker, Joseph; Kay, Steve. The dissertation author was the primary investigator and author of this material.
VITA

EDUCATION

Ph.D. Biology 2007-2013
University of California, San Diego
Principal Advisor: Steve Kay, Ph.D.
Co-Advisor: Joseph Ecker, Ph.D.
Thesis: Genome Scale Reverse Genetics Approaches for Novel Gene Discovery in *Arabidopsis thaliana* Circadian Clock Transcriptional Networks

B.A. Biochemistry and Neurobiology 2001-2005
University of Wisconsin – Madison

RESEARCH EXPERIENCE

Graduate Student, Kay lab and Ecker lab, University of California, San Diego 2007-2013
Discovery and characterization of novel transcription factors in the Arabidopsis circadian clock genetic network

Associate Research Specialist, Odorico Lab, University of Wisconsin – Madison, School of Medicine 2005-2007
Development of methods for differentiation and purification of pancreatic beta islet cells from mouse embryonic stem cells

Undergraduate Research Assistant, Kalil Lab, University of Wisconsin – Madison 2005
Capstone experience – investigate the potential of transplanted neural stem cells to repair central nervous system damage

Undergraduate Research Assistant, Lindroth Lab, University of Wisconsin – Madison 2003-2005
Undergraduate Research Scholar program – analyze the genetic variation and secondary metabolite composition of aspen tree population in response to insect predation

FELLOWSHIPS/HONORS/AWARDS

• NIGMS UCSD Genetics Training Program 2008-2012
• NSF IGERT Fellowship in Plant Systems Biology 2008-2011
MANUSCRIPTS IN PREPARATION

Nelson, J., Breton, G., Ecker, J., Kay, S. A genome scale reverse genetics overexpression screen reveals BBX31 as a novel regulator of the Arabidopsis circadian clock.


PUBLICATIONS


PRESENTATIONS

• Annual Plant Systems Biology IGERT Symposium, Speaker – San Diego, CA 2012
• International Congress on Plant Molecular Biology, Poster Presentation – Jeju, South Korea 2012
• International Congress on Arabidopsis Research, Poster Presentation – Madison, WI 2011
• Center for Chronobiology Symposium, Poster Presentation – San Diego, CA 2011
• Annual Plant Systems Biology IGERT Symposium, Speaker – San Diego, CA 2010
• Center for Chronobiology Symposium, Poster Presentation – San Diego, CA 2009
• American Society for Plant Biology, Poster Presentation – Honolulu, HI 2009
• Keystone Symposium – Plant Sensing, Response and Adaptation to the Environment, Conference Assistant – Big Sky, MT 2009

TEACHING EXPERIENCE

Teaching Assistant – University of California, San Diego

• BICD 100: Genetics 2013
• BIMM 116: Circadian Rhythms – Biological Clocks 2010
• BICD 120: Fundamentals of Plant Biology 2009
• BICD 123: Plant Molecular Genetics and Biotechnology Laboratory 2008
ABSTRACT OF THE DISSERTATION

Genome Scale Reverse Genetics Approaches for Novel Gene Discovery in Arabidopsis thaliana Circadian Clock Transcriptional Networks

by

Jeffrey Douglas Nelson

Doctor of Philosophy in Biology

University of California, San Diego, 2013

Professor Martin Yanofsky, Chair
Professor Joseph Ecker, Co-Chair
Professor Steve Kay, Co-Chair

The rotation of the earth on its axis results in alternating light and temperature conditions that have guided the evolution of nearly all forms of life. The resultant biological timekeepers are known as circadian clocks. This internal 24-hour oscillator orchestrates critical aspects of growth, development, and metabolism to synchronize with the appropriate time of day and impart adaptive
advantages. In the plant model organism *Arabidopsis thaliana*, a number of the molecular components of the circadian clock have been identified. However, critical nodes of this complex transcriptional network remain undiscovered. To address this deficiency, we created a collection of nearly every transcription factor encoded in the *Arabidopsis* genome. This transcription factor library platform allows us to conduct unbiased genome-wide screens for new genes functioning within the circadian clock. Using heterologous expression systems such as yeast one-hybrid and yeast two-hybrid, we have detected many protein-DNA and protein-protein interactions that indicate heretofore unrecognized regulatory connections. Initial investigations using an abridged transcription factor collection led to the discovery of *CCA1 HIKING EXPEDITION (CHE)* as a component of the *Arabidopsis* circadian clock. Further experiments with this platform are enabling the identification of a number of new circadian clock components, thereby greatly expanding our understanding of this genetic network and informing in-silico models of the clock. In addition, we have used the transcription factor library collection to conduct a systematic in-planta overexpression screen for novel regulators of the circadian clock. This approach has helped overcome many of the barriers to gene discovery associated with forward genetic screens. Proving the effectiveness of this method, we were able to identify members of the B-Box Zinc Finger protein family that display previously undiscovered robust circadian phenotypes when overexpressed. Functional genomic toolsets such as this transcription factor library are becoming
essential to gaining a systems level understanding of complex biological processes.
Chapter 1

Circadian Rhythms Throughout Biological Systems

1.1 Introduction to Circadian Rhythms

The first demonstration that organisms possess intrinsic daily rhythmic behavior occurred in the early 18th century when Jean Jacques Ortous de Mairan, intrigued by the daily opening and closing of the leaves of the heliotrope plant *Mimosa pudica*, moved an individual into constant dark conditions and observed that this rhythmic activity was maintained in the absence of cues from the sun (de Mairan, 1729). Since that initial experiment, the field of chronobiology has progressed tremendously. Researchers have shown that nearly every organism, from bacteria to humans, has evolved precise and robust internal oscillatory mechanisms that impart a predictable daily pattern of activity (Bell-Pedersen et al., 2005). This rhythmicity, governed by complex molecular signaling pathways, shapes many aspects of biology.

As our planet rotates about its axis, the surface of the earth experiences alternating conditions of light and dark. Over the course of billions of years, life has evolved to align behavior and physiology with these daily changes and gain adaptive advantages. Indeed, the very oxygen-producing organisms that generated the atmospheric conditions necessary for terrestrial life are
fundamentally linked to these patterns. Such cyanobacteria exhibit a daily repetition of photosynthesis when exposed to sunlight and carbon fixation when facing away from the sun. There are countless other elegant examples of such synchronization of the earth’s rotation and evolved rhythms of life. The fungus *Neurospora* restricts its DNA replication activities to the nighttime to prevent UV-radiation induced genomic damage (Pregueiro et al., 2006). Monarch butterflies use the daily setting sun as a compass, guiding their cross continental migratory patterns (Merlin et al., 2009). Seasonal flowering patterns of plants are governed by their precise monitoring of changes in day length throughout the year (Sawa et al., 2007). In humans, striking increases in metabolic diseases are correlated with shift work, in which activity and sleep are desynchronized with the standard day/night behavioral patterns (Karlsson et al., 2001).

The general term for the biological mechanism underlying these daily rhythms is the circadian clock, from the Latin ‘circa’ and ‘dies’ meaning ‘approximately a day’. The rigorous study of the field of 24-hour biological rhythms began in earnest in 1960 at the 25th Cold Spring Harbor Symposia on Quantitative Biology. It was here that such luminaries as Erwin Bunning, Colin Pittendrigh, Jurgen Aschoff, and Franz Halberg met, shared their work, and established conventions for the study of circadian rhythms that are still in use to this day.

The fundamental features of circadian rhythms are the following. First, they repeat once a day at the same time, i.e. they have a 24-hour period.
Second, these rhythms persist in the absence of external cues, as in rhythmicity is maintained in constant conditions. This criterion distinguishes these phenomena from 24-hour rhythmic events that occur simply in response to daily changes in light or temperature. Third, these rhythms are entrainable; meaning when local conditions change, circadian rhythms respond to the altered external cues and re-set to the local environment. Such entrainable cues are called zeitgebers (German for time giver). Zeitgebers can be light, temperature, feeding, activity, or other external signals. This allows organisms, for example, to respond to the changing timing of sunrise and sunset. Finally, in order for a biological pattern to be dubbed circadian, it must maintain 24-hour periodicity over a range of physiological temperatures. This concept is oftentimes referred to as temperature compensation. As opposed to most biochemical processes, which are exquisitely responsive to changes in temperature (speeding up as temperature increases and vice-versa), the biochemical reactions governing circadian clocks are buffered against a range of temperature conditions. Again, this allows for the maintenance of 24-hour periodicity throughout warm days and cool nights.

1.2 Circadian Clocks in Plants

Plants are no exception to these principles; they possess a complex circadian clock genetic network that governs activity such as leaf movement,
germination, growth, flowering time, stomatal opening, photosynthesis, and fragrance emission (McClung, 2006).

Research into the molecular mechanisms governing the circadian clock in plants has largely taken place with the plant reference organism, *Arabidopsis thaliana*. *Arabidopsis* is a powerful genetic model system. It has a fully sequenced and highly annotated genome (Arabidopsis Genome Initiative, 2000). The genome is compact with 125 mega base pairs, 27,000 genes, and 5 chromosomes. It is a small flowering plant with a short life cycle, taking approximately six weeks from germination to mature seed. In addition, *Arabidopsis* self-pollinates and produces several thousand seeds per plant. Furthermore, it is easy to generate transgenic organisms by employing modified strains of the soil bacteria *Agrobacterium tumefaciens* (Clough and Bent, 1998). Finally, due to its adoption and use by hundreds of research laboratories around the world, many resources exist to aid in its study. A extensive collection of knockout strains helps to identify gene function (Alonso et al., 2003). Also, available accessions from diverse native growing regions, ranging from northern Sweden to equatorial Cape Verde Island, helps to support research in plant evolution (Gan et al., 2011).

Research into the circadian clock in *Arabidopsis thaliana* was greatly accelerated by the ability to monitor a reporter gene, in vivo, whose activity reflects the daily cycles of gene expression. It was observed that the mRNA of the CHLOROPHYLL A/B BINDING protein (CAB2) exhibited a circadian rhythm
in its abundance (Millar and Kay, 1991). Researchers then took advantage of prior work demonstrating the ability to express luminescent proteins from fireflies in plants (Ow et al., 1986). A fragment of the Arabidopsis (CAB2) promoter fused to a firefly luciferase gene (CAB2::LUC) allowed scientists to visualize luciferase transcript levels via the rhythmic light emission from individual seedlings and gave them a powerful genetic system in which to screen mutagenized seeds for defects in circadian gene expression in living plants (Millar et al., 1992). Those mutants that altered the rhythmicity, period length, or amplitude served as strong candidates for mapping to a gene associated with the circadian clock. Indeed, this method led to the discovery of the first clock gene in Arabidopsis, TIMING OF CAB EXPRESSION1 (TOC1) (Millar et al., 1995). TOC1 encodes a pseudo response regulator protein with a mutant phenotype of a shortened period in CAB2::LUC expression. Further experiments led to the simultaneous reporting of CIRCADIAN CLOCK ASSOCIATED1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY) (Schaffer et al., 1998; Wang and Tobin, 1998). CCA1 was identified by its ability to bind the CAB2 promoter and LHY was found in an activation-tag screen for long hypocotyl mutants. These two genes are closely related by amino acid sequence, both encode single domain MYB transcription factors with a peak of transcription at dawn. Mutations in either CCA1 or LHY result in a shortened period phenotype. Their functional redundancy is indicated by the observation that the cca1 lhy double mutant is arrhythmic (Alabadi et al., 2002). In addition, overexpression of any of these three genes leads to a
arrhythmic phenotypes of multiple clock outputs (Wang and Tobin, 1998). A primary model for the core circadian oscillator was proposed after the observation that CCA1 and LHY bind to the evening element in the \textit{TOC1} promoter and negatively regulate its expression (Alabadi et al., 2001). Conversely, \textit{toc1} mutants display reduced \textit{CCA1} and \textit{LHY} expression. In addition, TOC1 was recently demonstrated to bind the promoters of \textit{CCA1} and \textit{LHY} via the TOC1 CCT domain (Gendron et al., 2012). Another mechanism by which TOC1 regulates \textit{CCA1} expression was recently identified to be via its interaction with the newly characterized transcription factor \textit{CCA1 HIKING EXPEDITION} (CHE). CHE is a member of the TCP (TB1, CYC, PCFs) transcription factor family and was discovered in a functional genomic screen as binding to the TCP binding element in the \textit{CCA1} promoter and negatively regulating its expression (Pruneda-Paz et al., 2009). This discovery linked together a longstanding model of the core circadian clock as a transcriptional feedback loop.

Additional genes have also been found to play a role in the regulation of the clock. For example, TOC1 is but one member of a family of pseudo response regulators (PRRs) that include PRR3, PRR5, PRR7, and PRR9, each of which plays a role in the progression of the daily clock activity (Nakamichi et al., 2010). In particular, PRR7 and PRR9 comprise another ‘morning’ regulatory loop by repressing \textit{CCA1} and \textit{LHY} (Zeilinger et al., 2006). In addition \textit{CCA1} and \textit{LHY} have been shown to activate the expression of \textit{PRR7} and \textit{PRR9} (Farré et al.,
Furthermore, a third ‘evening’ feedback loop exists between CCA1/LHY and the evening expressed genes *EARLY FLOWERING3* (*ELF3*), *EARLY FLOWERING4* (*ELF4*), and *LUX ARRHYTHMO* (*LUX*) (Locke et al., 2006). ELF3, ELF4, and LUX act as a circadian regulated complex to restrict hypocotyl elongation and growth to the end of the night (Nusinow et al., 2011). LUX encodes a putative MYB domain transcription factor essential for rhythmicity (Hazen et al., 2005). Mutations in *LUX* affect the amplitude and robustness of circadian rhythms. CCA1 and LHY bind to evening element motif in the promoter of *LUX*. As was the case for *toc1*, the expression of CCA1 and LHY is severely reduced in *lux* mutants, indicating a transcriptional feedback mechanism.

While these interlocking transcriptional feedback loops represent the foundation of the *Arabidopsis* clock, this is still a developing framework, with many nodes in the network remaining unclear (Locke et al., 2005; 2006; Pokhilko et al., 2010; 2012). Additionally, as demonstrated by the compensatory relationship between CCA1 and LHY and among the PRRs, functional redundancy seems to be an underlying feature in the *Arabidopsis* circadian clock, hinting at the existence of further, functionally overlapping, undiscovered genes. Because of its importance in plant growth and development, this redundancy likely confers an overall resistance to environmental perturbation and ensures the clock maintains its rhythmicity over a variety of changing conditions.

The importance of the circadian clock to plant growth and photosynthesis is well demonstrated; a competitive advantage is conferred upon plants whose
expression of core clock and essential metabolic genes coincide with the phase of the photoperiod and anticipate environmental changes (Dodd et al., 2005). Moreover, a recent study indicated that the increased growth and vigor displayed in hybrid and allopolyploid plants is due to a modulation of the circadian clock network (Ni et al., 2009). As these reports indicate, a more thorough understanding of this crucial system will give us insight into approaches that might be used to adjust and improve crop growth, flowering time, and overall yield. These are important goals in feeding a global population that is increasing exponentially without a comparable increase in the availability of arable land (Takeda and Matsuoka, 2008). Furthermore, biological processes uncovered in Arabidopsis may play a role in helping to understand analogous circadian clocks in humans and treat jet lag and metabolic diseases (Jones et al., 2008).

1.3 Plant Biology Research for Feeding a Growing Population

As the human population continues to grow at an unprecedented pace, we will face an ever-increasing strain on worldwide food supplies. Furthermore, with global climate becoming more volatile, extreme weather events more commonplace, and suitable fertile farmland diminishing - agricultural output will undoubtedly suffer. Unfortunately, the regions of the planet most affected by these trends are those developing nations least able to overcome such stresses due to lack of infrastructure and resources. Indeed, as was witnessed in early 2008, spikes in the price of grain in Africa and southern Asia resulted in massive
riots and civil unrest. Preventative measures to mitigate the future recurrence of such events are of the utmost scientific priority. Further research that contributes to our understanding of the biological mechanisms underlying beneficial crop traits will be essential in increasing our food output and ensuring the basic nutritional needs of humanity are met.
Chapter 2

Creation of a Complete Collection of Arabidopsis Transcription Factors

2.1 Rationale for the Functional Genomic Transcription Factor Library

Arabidopsis thaliana has gained wide acceptance as a model organism due to its power as a genetic system. Throughout the years, numerous biochemical, cytological and genomic tools have been developed, helping to considerably expand our knowledge of plant biology. More often than not, the discoveries made in Arabidopsis transcend the boundaries of plant sciences and have broader implications for the understanding of general biological processes (Jones et al., 2008). Although genetics is a powerful tool, its use for gene discovery is limited by a high degree of functional and genetic redundancies present among gene family members. Even reverse genetics approaches are often confounded by this redundancy issue, leading to the use of combinatorial knock down strategies such as artificial miRNA (Schwab et al., 2006). Gain-of-function screens are an appropriate alternative strategy to circumvent this issue and are increasingly being used in model species for which the open reading frame (ORF) reagents are available (Sopko et al., 2006; Irelan et al., 2009). Despite considerable effort by the international Arabidopsis community to create ORFeome collections that rival their human and mouse counterparts (Lamesch et
al., 2007; Rolfs et al., 2008) the lack of term funding led to the termination of several projects. Unintentionally, this situation led to the creation of unverified collections that are not being widely used, possibly because the cost to reanalyze the clones is too high to be assumed by a single laboratory. However, by combining these previous collections, and performing extensive de-novo cloning, our laboratory has created a high quality *Arabidopsis* transcription factor and regulator ORFeome collection that can be used in a wide variety of genomics applications.

Transcription factors constitute a large functional group that makes up almost 4% of all the genes in unicellular eukaryotes and 10% in multicellular eukaryotes (Kummerfeld and Teichmann, 2006). Due to their crucial role in controlling gene expression, exploring TF targets, binding partners, and modes of regulation is essential to understand any biological process. *Arabidopsis* offers some unique advantages that would make considerably important the development of large-scale genomic approaches for the study of TFs. First, the ease and low cost of gene transformation can be used to generate large transgenic collections that allow genome-wide overexpression screens. A second advantage is the propensity of most gene-regulatory regions to be isolated to a short region immediately upstream of the transcription start site, which greatly simplifies the design of large-scale yeast one-hybrid screens. Furthermore, phylogenetic analyses revealed that most TF families are larger in plants compared to humans (Shiu et al., 2005), suggesting that plants rely more
on specific DNA-binding proteins to modulate gene expression than human cells. For these reasons, the creation of functional genomics toolsets in Arabidopsis would allow the research community to perform large-scale studies and gather crucially informative data on global transcriptional networks in a multicellular organism. This would subsequently allow the construction of transcriptional maps with unprecedented definition and exponentially advance our understanding of the interrelation between modules of co-regulated genes, which will help identify and dissect combinatorial control of gene regulation. There are nearly 3,000 laboratories in the United States that are registered on the Arabidopsis Information Resource (www.arabidopsis.org) website that could potentially benefit from these resources. In addition, this collection could be used for plant research using any plant species; thus, the impact of the collections could be substantial in all fields of plant science and will greatly accelerate scientific progress in the future.

The creation of a partial TF ORFeome collection by our laboratory and its use in a yeast-one-hybrid-derived application was instrumental in cracking the transcriptional regulation of a major circadian clock gene (Pruneda-Paz et al., 2009). Comments on the discovery have acknowledged the innovative use of ORFeome collections and stated the advantage that a complete collection could have for the scientific community regarding the study of complex transcriptional circuits (Harmer, 2009; McClung, 2009; Robertson and Webb, 2009). In 2008, our group initiated the assessment of constructing a full transcription factor and
regulator collection that, according to a compilation of TF predictions from several
databases (Guo et al., 2005; Iida et al., 2005; Palaniswamy et al., 2006; Riaño-
Pachón et al., 2007), would contain 2492 TFs. In order to take advantage of
previous ORFeome efforts, we gathered all the clones that were available from
different distribution centers. In collaboration with Dr. Joseph Ecker at the Salk
Institute, we first gathered all the available TF clones in Gateway format from the
SALK gold standard collection (988 TFs) (Yamada et al., 2003). We then
acquired a quarter of the collection from the TIGR collection (83 TFs)
(Underwood et al., 2006), the REGIA collection (340 TFs) (Paz-Ares and The
Regia Consortium, 2002), and the Yale-Peking collection (252 TFs) (Gong et al.,
2004). In order to get a final and accurate count, we re-sequenced each clone
and determined that 68 were mis-annotated and needed to be re-isolated. This
meant that 540 TFs are still missing from any currently publicly available
ORFeome collection. In addition, since all these collections were created with
different cloning strategies, the frame and inserted sequence are variable, which
complicates the subsequent transfer of the full TF-cDNAs collection into a
particular destination vector.

In this project we aim to complete the following goals. (1) Create a
complete gold standard Gateway-donor TF collection (2492 clones) in open
(without stop codon) and closed (with stop codon) configurations using the same
cloning strategy for each clone. (2) Transfer each clone into different destination
vectors and generate nine application-ready collections to perform functional
genomics experiments. (3) Organize and distribute each clone collection to maximize its use as a community resource. (4) Modify and optimize the standard yeast one-hybrid assay to create an affordable strategy that could be performed by any laboratory interested in using the full collection. (5) Implement an automated platform for high-throughput screens.

2.2 Methods of Transcription Factor Library Construction

The goal of this work is to generate a complete and homogeneous collection containing every *Arabidopsis* transcription factor and transcriptional regulator. Toward this end, we PCR-amplified and cloned the coding sequences into a Gateway-compatible vector in closed (with stop codon) and open (without stop codon) configurations using the same cloning strategy. Thus, this collection will serve as a universal source of TF cDNAs that could be easily transferred to any destination vector that could be used for genomics applications.

The *Arabidopsis* genome encodes 2492 transcription factors and transcriptional regulators. About 75% of them have been previously cloned and are present in TF ORFeome collections that our lab has acquired. However, these collections have been created following different cloning strategies and in different vector backbones, which greatly complicates their compilation and use in downstream applications. For this reason, we used the available cDNAs (~75% of the full collection) as templates for PCR amplification, and we re-cloned them in open (without stop codon) and closed (with stop codon) configurations.
using a uniform strategy. In order to validate the open and closed gold standard libraries and be sure of their accuracy, we fully sequenced each clone, from each end and internally when necessary. An equivalent procedure was pursued to clone the remaining 25% of the collection, but in this case different template sources were used for PCR amplification: (1) genomic DNA for TFs lacking any described intron sequence, (2) cDNA clones available on the Riken Arabidopsis cDNA Encyclopedia DNABook (Kawai and Hayashizaki, 2003), (3) diverse cDNA pools available in the Kay laboratory (unpublished), or (4) using a two-round cloning approach in which the genomic sequence for the desired TF will be first amplified, cloned into Gateway-compatible pENTR and transferred into a vector for transient expression in tobacco leaves. Assuming that the overexpressed genomic DNA was spliced in tobacco leaves, the second step included extracting RNA from each of the transiently transformed tobacco leaves, synthesizing cDNA and re-amplifying the selected TF using specific primers. The amplified cDNA was then fully sequenced to determine its accuracy and the absence of intronic sequences. If none of these approaches was successful in providing the desired clones, we amplified each exon for a particular TF from genomic DNA and used a ligation independent strategy to assemble the clones. This strategy is feasible by using oligonucleotides containing sequences from the previous or following exons and the use of the In-fusion cloning system (Clontech).

2.3 Creation of Application-Ready Transcription Factor Collections
For this aspect of the project we used the collections generated as described above and transferred each TF into different destination vectors in order to create nine application-ready collections (Table 1). These collections will allow researchers from the Arabidopsis community to perform a wide variety of functional genomics approaches without spending the time and resources necessary to generate the initial reagents.

We took advantage of the Gateway cloning technology to transfer each coding sequence to multiple destination vectors in a high-throughput fashion. We first transferred each TF into three different vectors for overexpression (35S promoter) in planta. One of them will be to overexpress the “wild-type” version of the TF, and the other two to overexpress translational fusions of the TF to the EAR repressor motif (Hiratsu et al., 2003) or the VP64 activation domain (Bensmihen et al., 2004). These last two vectors can be used to perform dominant negative screens. The applications of these three vectors are complimentary and will allow the community to perform overexpression screens in different Arabidopsis ecotypes or sensitized backgrounds. The library was also transferred into two vectors, designed to maintain the plasmid in low or high copy number, that confer different sensitivity levels when used in GAL4-based yeast one- and two-hybrid screens. In addition, we also transferred the TFs to a vector suitable for split-ubiquitin yeast two-hybrid screens (Fields and Song, 1989), which is a useful strategy to analyze interactions between TFs since the protein-protein association occurs at the cytoplasmic membrane surface and
cannot produce auto-activation of the reporter in the nucleus such as with the GAL4-based system. We also transferred the TFs into one vector for bacterial expression of glutathione-S-transferase (GST)-tagged recombinant proteins that can be utilized to purify each TF for protein-binding microarray experiments (Berger and Bulyk, 2009) and into one vector for expression of green fluorescent protein (GFP)-fusions for the analysis of sub-cellular and sub-nuclear localization in-planta. All the plasmids mentioned above have been tested in our laboratory.

The final application-ready collection will allow the determination of TF targets in vivo. Identification of TF targets is an important aspect of the characterization of transcriptional networks. Although relatively new, the combination of optimized chromatin immunoprecipitation protocols with deep sequencing technologies (ChIP-seq) is quickly becoming a standard in the field providing valuable genome-wide information. A group recently identified genome-wide targets for SEPALLATA3 (SEP3), an Arabidopsis MADS Box TF, using the ChIP-seq method (Kaufmann et al., 2010). This group was fortunate to have in their possession a native antibody that specifically recognizes SEP3. In reality, it is not feasible or cost efficient to generate antibodies against each Arabidopsis TF. For this reason, we are addressing which is the best epitope tag from a series of four that can be used to generate a TF collection that would be suitable for ChIP-seq in Arabidopsis. We are working in collaboration with the Ecker laboratory at the Salk Institute for Biological Sciences, which is equipped with multiple next generation sequencing machines and has already completed
ChIP-seq experiments with native antibodies against ETHYLENE INSENSITIVE 3 (EIN3). Thus, this group is proficient in each step of the ChIP-seq method.

The ChIP-seq method begins with isolation of DNA-protein complexes using an antibody that recognizes a protein of interest or in our case an epitope. DNA fragments are isolated away from proteins and adaptor-ligated to generate a sequencing library. The sequencing libraries are next amplified on a flow cell to generate clusters of homogenous DNA fragments using an Illumina Genetic Analyzer cluster station. Next, these libraries are sequenced using an Illumina Genetic Analyzer II. This process takes approximately 3 days for a 36 base pair run and currently generates approximately 10 million mappable reads per lane (8 lanes total per flow cell). Sequence information will be extracted from the image files with the Illumina algorithms and mapped to the *Arabidopsis* Columbia reference genome with the Illumina Eland algorithm. Finally, these data sets are analyzed for an over-enrichment of reads using Peakseq, a publicly available algorithm.

In collaboration with the Ecker Lab, we will compare the immunoprecipitation efficiency and ChIP-seq profiles for TFs selected from ten different gene families, which will be tagged with GFP, a biotinylated peptide, 3xHA, or HIS-FLAG. It should be noted that the success of any ChIP-seq experiment is dependent on the quality of the antibody. Fortunately, commercial high-quality antibodies are available for all four types of epitopes mentioned above and can be accessed by anyone in the community. From these
experiments we will select the vector to be used to create the ninth and final application-ready collection.

2.4 Organization and Distribution of the Transcription Factor Collections

We aim to assure the broad distribution and long term maintenance of the resources described above. The wide and easy access to the TF collections by any scientist will maximize their usage and thus benefit the scientific community.

The collections were symmetrically organized in 22 x 96-well plates, giving preference to TFs belonging to the same family to be in the same plate. In this way, it will be easier to distribute the library to laboratories interested in a specific TF family. However, to maintain the integrity of the collections and reduce handling costs, we would recommend that the distribution should be performed in fractions not smaller than one full plate and, ideally, as a whole. A second important aspect have learned from previous TF ORFeome efforts is the tendency for some clones to be unstable when stored as glycerol stocks. For this reason, we have added a second level of security by using the QIAsafe DNA stabilizer from Biometrica to deliver and store (at room temperature) additional copies of the collections as DNA stocks.

The maintenance and distribution of the collections will be mainly performed by the *Arabidopsis* Biological Resource Center (ABRC) that is located at Ohio State University. The ABRC was established to acquire, preserve and distribute seed and DNA resources that are useful to the *Arabidopsis* research
community. Since they handle and ship more than 100,000 stocks annually to scientists from 60 countries, we believe ABRC constitutes a reliable option for the long-term preservation and sharing of the collections. We plan to provide the ABRC with glycerol and DNA stocks as well as 50 copies of each library containing 50-100ng of each QIAsafe-stabilized plasmid for their initial distribution. Finally, as a backup option to further reinforce the long term and universal usage of the collections, we plan to deliver them to different laboratories or institutions around the world.

The distribution of the collections will be also accompanied by an electronic data sheet containing the composition of each library including basic information for each TF that will be posted in the ABRC and/or our laboratory website. Furthermore, this information is already required by the ABRC as part of the donation process and therefore will be available on their website. During the construction of the collections we have generated data regarding each clone such as sequencing information for the TF-donor collection and plasmid maps of each TF and destination vector. We consider that this information could be useful for any user who wants visual confirmation of the validity of a specific clone. In order to organize and distribute this information, the data will be entered in a database containing each family and each TF and saved using the Geneious software platform. The Basic version of this software is freely available and contains all the features necessary to open a file containing the nucleotide
sequence and sequencing reads. The final file will be posted on our server and accessible through the web.

2.5 Optimization of a Yeast One-Hybrid Protocol Adapted for the TF Library

We hope to provide the community with an affordable and simple procedure to perform screens using the yeast one-hybrid compatible TF collections generated as described above. We recently established a functional genomic approach coined “promoter hiking” in which a library of 200 TFs in GAL4-AD fusions was tested using the yeast one-hybrid system on five contiguous fragments of the clock gene CCA1 promoter. Our approach greatly accelerated the screening procedure and was instrumental in the discovery of CHE, a new clock gene (Pruneda-Paz et al., 2009). One advantage of this approach is that since each TF is assayed separately in a 96-well plate format, there is no absolute requirement, such as when performing a screen using a random cDNA library, to sequence the constructs associated with the positive hits at the end of the procedure. Briefly, using the LacZ gene as reporter, we performed the β-galactosidase assay and determined a fold activation level for each TF and compared it to the average fold activation. This procedure is done independently for each promoter fragment, and gives us a level of sensitivity that would be costly to obtain by sequencing each hit in a random screen. While this protocol was manageable in our initial low throughput screen, it would not be cost effective or time efficient to screen multiple promoter regions using a full 1956-
clone library. Therefore, we have optimized the initial protocol in order to create an approach where the library could be screened in an ordered fashion.

Our optimization procedure has focused on: (1) yeast transformation and mating procedures, (2) alternatives to measure β-galactosidase activity such as the utilization of different substrates or cell lysis protocols, and (3) the use of fluorescent or bioluminescent markers as reporters. In addition, after a thorough assessment of the reporter activity sensitivity, we have tested the feasibility of applying a smart pooling strategy in order to reduce the size of the library to be screened.

Finally, in order to fully take advantage of this resource and more importantly, to use its potential to considerably accelerate research in the field of transcriptional regulation, we have implemented an automated platform. The selected platform is flexible enough to be used for the construction of the libraries and to perform high throughput screens after the completion of the project. The layout allows a fully automated yeast one-hybrid screen, excluding the incubation time at 30°C, contains a robotic arm that transfers plates between the following stations: 2 centrifuges, 1 plate hotel, 1 delidder, 1 plate sealer, 10 plate stacks, 1 pipetting station containing a tip washing station and 1 spectrophotometer. This platform and layout were chosen for their extreme flexibility and could be modified at any time to be adapted to any modification of the protocol performed during the optimization time. In addition, this platform can be used for the
construction of the libraries to perform tasks such as transferring clones for clonase reaction and preparing plasmids for transformation.

2.6 Initial Results from the Transcription Factor Library

Employing a variety of methods: aggregation of previous collections, re-sequencing, and de-novo cloning, we have generated the most comprehensive collection of *Arabidopsis* transcription factors and transcriptional regulators yet created, containing 1956 clones. This library is already fulfilling the great potential inherent in such functional genomics approaches. By employing the automated system described above, we were able to collect a rich dataset of potential transcriptional regulators of key *Arabidopsis* gene promoters. In particular, fragmenting the upstream promoter and 5’ UTR fragments of the *CCA1* promoter, as described in (Pruneda-Paz et al., 2009), and querying against the transcription factor library has yielded some intriguing novel clock regulators. When a genome wide yeast one-hybrid TF screen was conducted against the *CCA1* promoter, a bHLH transcription factor (At1g35460) was found to display a highly significant affinity for the fragment -213/-42 bp upstream of the transcriptional start site. This gene was previously described to activate the expression of the key flowering gene *CONSTANS (CO)* and was given the name *FLOWERING BHLH 1 (FBH1)* (Ito et al., 2012). FBH1 binds specifically to the E-Box motif (CATATG) in the *CCA1* promoter; the β-galactosidase reporter activity is significantly reduced in constructs containing mutated E-Box and E-Box like
motifs. Furthermore, GFP affinity tagged versions of FBH1 under the control of the constitutive 35S promoter display strong affinity for the genomic region of the CCA1 promoter containing the E-Box motif via chromatin immunoprecipitation. In addition, the activity of FBH1 was investigated using in-planta clock gene promoter::luciferase reporter experiments, which are commonly used in our laboratory to display the rhythmic gene expression patterns of the circadian clockwork. In these experiments, transgenic plants overexpressing FBH1 were generated in a CCA1::LUC+ genetic background. Owing to the transcriptional regulation of CCA1 displayed in previous experiments, FBH1 overexpression resulted in a repression of the amplitude of CCA1::LUC+ oscillation. Furthermore, in one of the FBH1 overexpressing lines examined, a shortened period phenotype was displayed, indicating a sped up circadian clock. Finally, the peak of the CCA1::LUC+ reporter expression, its phase, was significantly shifted earlier in the FBH1 overexpression plants. In addition to the prior data, these in-vivo experiments provide further evidence for the importance of FBH1 in the transcriptional regulation of CCA1 and by extension critical importance for the maintenance of proper timekeeping in the Arabidopsis circadian clock. The success in identifying this novel circadian transcriptional regulator will likely been duplicated many times over as the use of this vital functional genomic resource continues to expand. We envision a point in the near future when the data integration of many yeast one-hybrid and yeast two-hybrid experiments results in the ability to generate robust genetic networks of transcriptional regulation that
can be biologically validated and incorporated into ever more accurate predictive models of plant circadian clock biology.

Importantly this transcription factor library system is not limited to the investigation of circadian clock genetic networks. The transcriptional regulation of any Arabidopsis biological process, from drought stress to light perception to growth, might be explored with a previously impossible breadth with this powerful platform.

2.7 Acknowledgements

Chapter 2, in part, is currently being prepared for submission for publication of the material. Pruneda-Paz, Jose; Breton, Ghislain; Nelson, Jeffrey; Ecker, Joseph; Kay, Steve. The dissertation author is a research contributor and co-author of this work.
Figure 2.1 Transcription factor family abundance in Arabidopsis vs. human. A Comparison of TF number per family for Arabidopsis (green) and human (orange) (Adapted from (Shiu et al., 2005)).
**Figure 2.2** Flow chart representation of TF library and promoter collections.
Figure 2.3 An overview of the “Promoter Hiking” approach. A library of transcription factors is assembled and transferred to a destination vector in order to get translational fusions to the GAL4 activation domain (gal4AD). Yeast bait strains are constructed by genomic insertion of transcriptional fusions of tiled-promoter fragments to the β-galactosidase gene (lacZ) and each one is individually challenged against the transcription factors present in the library. The DNA-binding is followed by the quantification of β-galactosidase activity.
Figure 2.4 Automated platform for library construction and yeast one-hybrid screens. This platform contains all the components to pipette, wash, transfer, shake, incubate, centrifuge and measure absorbance or luminescent signals in a 96 or 384 well format.
Figure 2.5 Schematic for cloning transcription factors A. Cloning workflow B. Percentage of total number of transcription factors encoded in the *Arabidopsis* genome successfully cloned and present in library C. Transcription factors binned by presence in pre-existing transcription factor databases D. Of the 1956 successfully cloned transcription factors, number containing mutations or alternative isoforms E. Of the 536 transcription factors not successfully cloned, reasons for failure of cloning
Figure 2.6 Representation of cloned and not cloned genes per TF-family. (small_fam groups represents all families with 1-4 members).
Figure 2.7 Results of yeast one-hybrid screen of TF library against the promoter of CCA1

A. Workflow of cloning process

B. Model of tiled CCA1 promoter and 5’ UTR

C. Scatter plot of β-gal activity of full length promoter vs. promoter fragments

D. Number of TF interactions by promoter fragment

E. Number of promoter fragments bound by TFs
**Figure 2.8** FBH1 identified in yeast one-hybrid screen. A. β-gal activity of FBH1 binding to -213/-42 CCA1 promoter fragment. B. WT and mutated forms of E-Box FBH1 binding sites present in CCA1 promoter constructs. C. Altered affinity of FBH1 for WT and mutated promoters. D. Specific immunoprecipitation of GFP tagged FBH1 on E-Box region of CCA1 promoter. E. Suppressed amplitude of luciferase output in FBH1 overexpressing plants. F. Shortened period of FBH1 overexpressing plants. G. Phase shifts of FBH1 overexpressing plants.
Table 2.1 Application ready TF collections

<table>
<thead>
<tr>
<th>Application</th>
<th>Gene fusion</th>
<th>Destination vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overexpression <em>in-planta</em></td>
<td>none</td>
<td>pB7GW2 (Karimi, 2002)</td>
</tr>
<tr>
<td>Overexpression <em>in-planta</em></td>
<td>EAR repressor motif</td>
<td>pB7GW2-EAR</td>
</tr>
<tr>
<td>Overexpression <em>in-planta</em></td>
<td>VP64 activation domain</td>
<td>pB7GW2-VP64</td>
</tr>
<tr>
<td>Yeast one- and two-hybrid</td>
<td>GAL4-activation domain</td>
<td>pDEST22 (Invitrogen)</td>
</tr>
<tr>
<td>Yeast one- and two-hybrid</td>
<td>GAL4-activation domain</td>
<td>pACT-GW (Shimoda, 2008)</td>
</tr>
<tr>
<td>Split-ubiquitin yeast two-hybrid</td>
<td>Ubiquitin ligase N-terminus</td>
<td>pMet YC-DEST (Grefen, 2007)</td>
</tr>
<tr>
<td>Bacterial expression</td>
<td>Glutathione-S-transferase (GST)</td>
<td>pDEST15 (Invitrogen)</td>
</tr>
<tr>
<td>Subcellular localization <em>in-planta</em></td>
<td>Green fluorescent protein (GFP)</td>
<td>pMDC83(Curtis, 2003)</td>
</tr>
<tr>
<td>Chromatin immunoprecipitation</td>
<td>To be determined</td>
<td>To be determined</td>
</tr>
</tbody>
</table>
Chapter 3

Overexpression Screen for Novel Arabidopsis Transcriptional Regulators

3.1 Introduction to the Circadian Clock in Arabidopsis

Organisms that live on the surface of the planet experience alternating cycles of light and dark as the earth rotates on its axis. As a result, they have evolved precise molecular mechanisms that align their behavior and physiology with specific times of day. Everything ranging from fungi to fruit flies to humans all display such circadian rhythms, which are governed by a network of genes collectively known as the circadian clock. These networks possess a strikingly similar general architecture across taxa that include input sensors, a central oscillator, and output pathways. The fundamental features of circadian clocks are that they possess an approximately 24 hour period; their activity is generated endogenously, persisting even in the absence of external signals; and they can maintain their biochemical rhythmicity over a range of temperatures. Plants are no exception to these principles; they possess a complex circadian clock genetic network that governs a diverse set of activities such as leaf movement, germination, growth, flowering time, stomatal opening, pathogen resistance, photosynthesis, and fragrance emission (McClung, 2006).
Research into the molecular mechanisms governing the circadian clock in the plant reference organism, *Arabidopsis thaliana* was greatly accelerated by the ability to monitor a reporter gene, in vivo, whose levels reflect the daily cycles in gene expression. The first gene reported to display such cyclic expression was *CHLOROPHYLL A/B BINDING PROTEIN 2 (CAB2)*, which exhibited a robust 24-hour rhythm in its abundance (Millar and Kay, 1991). This observation led to the insight that a fragment of the *Arabidopsis CAB2* promoter fused to a firefly luciferase gene (*CAB2::LUC*) would allow researchers the ability to visualize clock activity via the rhythmic light emission from individual seedlings. Furthermore, it gave them a powerful genetic system in which to screen mutagenized seedlings (Millar et al., 1992). Mutants that showed alterations in the rhythmicity, period length, or amplitude of the reporter were candidates for genes associated with the circadian clock. Indeed, this method led to the discovery of the first clock gene in *Arabidopsis*, *TIMING OF CAB EXPRESSION 1 (TOC1)* (Millar et al., 1995). *TOC1* encodes a pseudo response regulator protein with a mutant phenotype of a shortened period in *CAB2::LUC* expression. Further experiments led to the near simultaneous reporting of *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)* and *LATE ELONGATED HYPOCOTYL (LHY)* (Schaffer et al., 1998; Wang and Tobin, 1998). *CCA1* was identified by its ability to bind the *CAB2* promoter. *LHY* was discovered in a forward genetic screen for long hypocotyls. These two genes are closely related by sequence, both encode MYB domain transcription factors and mutations in either gene results in a
shortened period phenotype. Their functional redundancy is indicated by the observation that the \emph{cca1/lhy} double mutant enhances the clock phenotype, yielding arrhythmic plants (Alabadi et al., 2002). In addition, overexpression of any of these three genes also leads to an arrhythmic clock phenotype. The initial model for the core circadian oscillator arose with the observation that CCA1 and LHY bind to the ‘evening element’ in the \emph{TOC1} promoter and negatively regulate its expression (Alabadi et al., 2001). Ten years later, this model of a negative transcriptional feedback loop was resolved with the discovery that TOC1 directly represses \emph{CCA1} and \emph{LHY} expression throughout the night (Gendron et al., 2012).

Additional genes have also been found to play a role in the regulation of the clock. For example, TOC1 is but one member of a family of pseudo response regulators (PRRs) that include PRR3, PRR5, PRR7, and PRR9, each of which likely contributes to the progression of daily clock activity. In particular, PRR5, PRR7, and PRR9 comprise another ‘morning’ regulatory loop by repressing \emph{CCA1} and \emph{LHY} throughout the daytime (Nakamichi et al., 2010). Furthermore, a third ‘evening’ feedback loop has been reported between CCA1/LHY and the evening expressed genes \emph{EARLY FLOWERING4} (\emph{ELF4}), \emph{LUX ARRHYTHMO} (\emph{LUX}) and \emph{EARLY FLOWERING 3} (\emph{ELF3}) (Nusinow et al., 2011).

These multiple interlocking transcriptional feedback loops represent the foundation of the \emph{Arabidopsis} circadian clock timekeeping mechanisms (Doherty and Kay, 2010). However, this model remains a relatively open framework, with
nodes in the network remaining uncharacterized (Locke et al., 2005). For example, computational models reconstructing *Arabidopsis* circadian clocks indicate the necessity for additional ‘night inhibitor’ and ‘TOC1 activator’ nodes (Pokhilko et al., 2010). Additionally, as demonstrated by the compensatory relationship between CCA1/LHY and among the PRRs, functional redundancy seems to be an underlying feature in the *Arabidopsis* circadian clock, hinting at the possibility of further undiscovered, functionally overlapping, genes. Because of the clock’s fundamental significance in a multitude of plant processes, this redundancy likely confers an overall robustness in the face of environmental perturbation and ensures the clock maintains its appropriate rhythmicity over a variety of changing conditions.

Indeed, the importance of the circadian clock to plants is well demonstrated. First, it is widely influential, it has been estimated that up to 89% of the total transcriptional output of the *Arabidopsis* genome cycles under circadian or diurnal conditions (Michael et al., 2008). Second, plants with clock gene mutations resulting in long or short circadian rhythms are out of sync with their environment and as a result, contain less chlorophyll and fix less carbon than wild-type plants (Dodd et al., 2005). Third, increased growth and vigor displayed in hybrid and allopolyploid plants is due, at least in part, to a modulation of the circadian clock network (Ni et al., 2009). As these reports show, a more thorough understanding of this crucial system will give us insight into approaches that might be used to adjust and improve crop growth, and
overall yield. These are important goals in feeding a global population that is increasing exponentially without a comparable increase in the availability of arable land (Takeda and Matsuoka, 2008). Furthermore, biological processes uncovered in Arabidopsis may assist in our understanding of analogous circadian clocks in humans and help to treat jet lag and metabolic diseases (Jones et al., 2008).

3.2 Creating a Comprehensive Arabidopsis Transcription Factor Library

With the completion of the Arabidopsis thaliana genome (Arabidopsis Genome Initiative, 2000), the opportunities for understanding the complex genetics underlying plant biology have expanded tremendously. Indeed, research in Arabidopsis has led to breakthroughs in our understanding the mechanisms of plant growth, hormone signaling, immunity, and numerous other processes. Furthermore, it is now possible to undertake functional genomic approaches to interrogate the entire repertoire of genes, transcripts, and proteins in an organism. One functional class of proteins whose role is essential for gene regulation is the transcription factor (TF). By binding to specific elements within the promoters of target genes and recruiting or inhibiting the core transcriptional machinery, these proteins control the expression profile of an organism’s genome at a given time or state. For circadian clock networks in particular, transcription factor based feedback loops have been well established as the foundation for the precise orchestration and maintenance of rhythmic outputs (McClung, 2006). So,
by creating a large collection of open reading frames (ORFs) encoding such proteins, one would be able to conduct screens for alterations in expression activity or transcription factor binding in a given condition. In *Arabidopsis*, research groups have initiated the compilation of collections of cDNAs of every transcription factor (Paz-Ares and The Regia Consortium, 2002; Yamada et al., 2003; Gong et al., 2004). Ongoing work in the Kay lab is aimed at completing this endeavor and creating a complete transcription factor ORFeome (2492 clones) by combining existing TF collections (SALK, TIGR, REGIA, YALE) and re-validating sequence identity. To date, we have successfully compiled 1956 of these TFs. For versatility and ease of use, the collection has been cloned into the widely used GATEWAY system. This will allow the facile transfer into numerous expression vectors for a variety of downstream applications such as yeast one- and two-hybrid, overexpression *in planta*, bacterial recombinant protein expression, or chromatin immunoprecipitation.

One such functional genomic approach recently undertaken by researchers in the Kay lab led to the identification of a novel transcription factor that acts within the central oscillator of the circadian clock in *Arabidopsis*. CCA1 HIKING EXPEDITION (CHE) is a member of the TCP (TB1, CYC, PCFs) transcription factor family and was discovered in a functional screen as binding to the TCP binding element in the *CCA1* promoter and negatively regulating its expression (Pruneda-Paz et al., 2009). The experiment involved generating a collection of 200 TFs whose transcripts cycle robustly under circadian and diurnal
conditions, and screening them in a yeast one-hybrid system against various overlapping fragments of the *CCA1* promoter (Pruneda-Paz et al., 2009). This so-called ‘promoter hiking’ assay represented a crucial step forward in the development of genome-enabled screens for transcription factors affecting the circadian clock.

### 3.3 A Transcription Factor Overexpression Method for Clock Gene Discovery

As previously indicated through computational models and demonstrated by the example of CHE, there are likely additional circadian clock regulators in *Arabidopsis* that remain undiscovered. Traditional methods such as forward genetic screens, while previously fruitful, seem to be reaching a saturation point as new mutants affecting clock phenotypes simply yield new alleles of already characterized clock genes. As an alternative, I have undertaken a large-scale reverse genetics screen that leverages the collection of transcription factors compiled in our laboratory. Because of the facile nature of *Arabidopsis* transformation, I undertook an approach based on systematically generating transgenic TF overexpression lines and interrogating their effect on clock function. To bring this project into a manageable scale, I selected a subset of transcription factors whose transcripts themselves exhibit the most robust cycling over circadian and diurnal conditions (Michael et al., 2008). This biased approach is reasonable because a large majority of genes known to be involved
in the core circadian oscillator display a circadian rhythm in their own gene expression.

To initiate this genome enabled transcription factor screen, I selected a subset of 250 transcription factors whose mRNAs exhibit the most robust cycling under circadian and diurnal conditions. Next, I transferred these 250 clones via LR clonase reaction into overexpression (OX) vectors with a 35S cauliflower mosaic virus (CaMV) promoter driving the expression of the cDNA. I then transformed these vectors into agrobacterium, utilizing a 96 well electroporator, which greatly expedited the transformation process. These agrobacterium strains were transformed into Columbia ecotype Arabidopsis containing the FLAVIN-BINDING, KELCH REPEAT, F-BOX1 (FKF1) promoter fused to luciferase reporter (FKF1::LUC+) via floral dipping (Logemann et al., 2006). FKF1 is a blue light photoreceptor that regulates day length measurement and flowering time in Arabidopsis (Nelson et al., 2000). The FKF1::LUC+ reporter was selected because its expression robustly cycles and FKF1 is a direct output of the core circadian oscillator. Plants overexpressing components of the circadian clock will likely display a period effect, amplitude modulation, or arrhythmicity of this reporter. Alternatively, it is possible that OX lines affecting this reporter could be involved in FKF1 mediated flowering regulation, independent of the clock.

After the plants self-pollinated, seeds were collected and grown on MS plates in the presence of BASTA to select for T1 seeds harboring the
overexpression transgene. After growing in a 12hrs light, 12hrs dark regime for the first 10 days, transgenic seedlings were harvested and arrayed for luciferase screening (Welsh et al., 2005). The seedlings were released into constant light conditions for five days and imaged every 2.5 hours. The collected images were then stacked and quantified for luminescence using the Metamorph software. Finally, the raw luminescence data was analyzed via the BRASS Excel package that performs fast-Fourier transform-nonlinear least square (FFT-NLLS) analysis (Straume et al., 1991) to assign a period and phase score for each set of transgenic seedlings, which are then compared to control plants.

3.4 Results of Transcription Factor Overexpression Screen

Through this method, TFs that altered the phase, period, amplitude, or rhythmicity of the circadian clock were identified. Although this screen was blind to gene identity, I was able to detect period alterations produced by the overexpression of the known clock genes *CCA1*, *LHY*, and *LUX*. This indicates that the approach is sensitive enough to capture altered circadian period phenotypes from TFs identified using other screening methods. Furthermore, it demonstrates that even TFs with overlapping clock functions (*CCA1* and *LHY*) can be identified. This provided confidence in the ability of this screening procedure to enable the discovery of novel TFs regulating the circadian transcriptional network.
This screen has indeed yielded multiple promising hits, the most interesting of which are members of the B-Box (BBX) family of TFs. Specifically, when overexpressed, CONSTANS-LIKE 9 (COL9) shortens the period of luciferase expression by 1.5 hours, while B-BOX 31 (BBX31) lengthens it by 2 hours. Broadly speaking, the 32 genes that make up this family are characterized as having a combination of one or two N-terminal B-Box zinc finger domains and, in some cases, a C-terminal CCT (CONSTANS, CONSTANS-like, TIMING OF CAB 1 (TOC1)) domain (Khanna et al., 2009). The B-Box zinc finger is thought to mediate protein-protein interactions. The CCT domain contains a nuclear localization sequence and is found in both this family as well as in the PRR genes. Intriguingly, four members of the PRR family (TOC1, PRR5, PRR7, & PRR9) play fundamental roles in the core oscillatory mechanisms of the circadian clock. In addition, for certain members of the B-Box family, the CCT domain imparts DNA binding activity (Tiwari et al., 2010). COL9 possesses two zinc finger domains and a CCT domain while BBX31 contains only one B-Box zinc finger domain and no CCT. As a result of their difference in structure, investigation of these two genes will provide the opportunity to interrogate the functional significance of these domains within the clock.

The founding, and best-characterized, member of this family is CONSTANS (CO). CO is a key node in the clock-regulated photoperiod dependent flowering pathway in Arabidopsis and many other plants (Putterill et al., 1995). CO protein activates the expression of FLOWERING LOCUS T (FT)
in the leaf, which in turn travels via the phloem to the floral meristem and initiates flowering (Suárez-López et al., 2001). This process, however, occurs only in long days, when the repression of CO by CYCLING DOF FACTORS 1 & 2 (CDF1 & CDF2) is relieved. This de-repression is mediated primarily by the interaction of the clock-regulated proteins GIGANTEA (GI) and FKF1; GI and FKF1 form a light dependent complex near the end of long days and target the CDFs to the proteasome via polyubiquitination (Sawa et al., 2007). Additional regulation of CO protein arises from its interaction with CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) and SUPPRESSOR OF PHYA 1 (SPA1). COP1 and SPA1 form an ubiquitin-ligase complex that targets CO for degradation in the dark under short day conditions (Liu et al., 2008). Under long days, however, CO protein is stabilized due to the photoreceptor-mediated repression of COP1. The regulation CONSTANS transcript and protein is indeed very precise and complex, owing to its pivotal role in mediating photoperiod dependent flowering (Turck et al., 2008).

For the candidate genes identified by my screen, (COL9 and BBX31) there is very limited prior research. One publication looks into the activity of COL9 and indicates that its overexpression leads to late flowering (Cheng and Wang, 2005) (a phenotype that I have also observed in both the COL9 and BBX31 OX lines) and loss of function T-DNA lines display very slight early flowering. This finding is interesting as it represents the opposite phenotypes as are seen in CONSTANS OX and loss of function mutant lines, which flower early and late.
respectively. The authors of this paper also describe a repression of CO and FT transcripts in COL9 OX lines, which likely account for its flowering phenotype (Cheng and Wang, 2005). The role of these genes in circadian clock regulation, however, has been largely unexplored.

In order to gain more insight into the function of these genes, I have acquired knockout (KO) T-DNA insertion lines for both COL9 and BBX31. However, these KO display no flowering time or elongated hypocotyl phenotypes that are commonly associated with clock gene mutations. I hypothesize this is due to the functional redundancy in this expanded gene family. Further analysis of published datasets has revealed additional information about these candidate genes. COL9 has a phase of expression at zeitgeber time (ZT) 12, regardless of the day length or light conditions. Additionally, its promoter contains three evening elements (EE), which are key cis regulatory motifs bound by CCA1 and LHY to repress target gene transcription during the day and impart evening phased expression. Also, COL9 displays a 4 hour delayed phase shift when LHY is overexpressed. These data point toward a direct regulation of COL9 by the clock. BBX31 mRNA peaks at ZT 0 and is unaffected by LHY overexpression. Despite its dawn-phased expression, its promoter contains four EE. However, it also possesses two G-box motifs and a morning element, indicating a potentially more complex mode of transcriptional regulation. In order to interrogate the transcriptional control of these two candidate genes, future experiments could apply the transcription factor library resource to test the affinity of TFs such as
for their promoter regions via yeast one-hybrid. In the case of COL9, given its phase of expression and cis element profile, I suspect that its transcript is under the direct regulatory control of CCA1 and/or LHY. This result could present a mechanistic link between COL9 and the core clock, providing the foundation for its integration into the circadian clock model.

In addition to determining the clock’s role in the transcriptional regulation of COL9 and BBX31, I explored the role of these novel TFs on the expression patterns of known clock genes. This approach has given insight into the mechanism of the period length phenotypes observed in the overexpression luciferase screen. That is, what clock gene transcripts are COL9 and BBX31 regulating? In order to accomplish this, I have collected a circadian time series of total mRNA from COL9 & BBX31 overexpression transgenic lines as well as their respective T-DNA insertion lines. Using a 384 well quantitative RT-PCR (qPCR) setup, I have queried the expression of potential target genes in a high-throughput manner. The results of these experiments have indicated that BBX31 OX represses the expression of LHY, the morning expressed core clock TF. In addition, the bbx31 mutant displays an elevated level of LHY expression relative to control plants. This result is consistent with the luciferase reporter phenotype; holding down the expression of LHY would lead to a drastic alteration of the rhythmicity of the clock.

3.5 Future Experimental Approaches
Future experiments will continue to provide additional insight into transcriptional regulatory targets of COL9 and BBX31 and establish their position within the circadian clock network. In order to test these potential targets in vivo, I have generated epitope tagged versions of COL9 and BBX31 using well-established C- and N-terminal His-Flag constructs generated in the Kay lab. These reagents will allow future investigators to ascertain, via chromatin immunoprecipitation (ChIP), the in-vivo binding activity of these TFs with target clock gene promoters (such as LHY) identified in qPCR experiments. I am confident that this Protein-DNA interaction data will provide additional mechanistic explanation for the clock period phenotypes and, with the data generated in the aforementioned experiments, establish a role for these BBX proteins in new mode of transcriptional regulation of the Arabidopsis circadian clock.

3.6 Acknowledgments

Chapter 3, in part, is currently being prepared for submission for publication of the material. Nelson, Jeffrey; Breton, Ghislain; Ecker, Joseph; Kay, Steve. The dissertation author was the primary investigator and author of this material.
Figure 3.1 Initial workflow for transcription factor overexpression screen. Also included are follow up characterization experiments.
Figure 3.2  LHY overexpression screen results. FKF1::LUC+ bioluminescent plants overexpressing the circadian clock gene *LHY* versus control plants. LHY overexpression plants display arrhythmicity of the reporter gene.
The image shows a graph titled "35S::COL9". The graph plots the bioluminescence over time in LL (light conditions) for plants overexpressing the COL9 gene versus control plants. The x-axis represents time in LL (hours) ranging from 0 to 120, and the y-axis represents bioluminescence (counts/seedling/25min) ranging from 0 to 30,000.

**Figure 3.3** COL9 overexpression screen results. FKF1::LUC+ bioluminescent plants overexpressing the candidate clock gene COL9 versus control plants. COL9 overexpression plants display shortened period of the reporter gene.
**Figure 3.4** Additional COL9 overexpression results. Plants overexpressing COL9 display shortened period relative to control plants.
Figure 3.5 BBX31 overexpression screen results. FKF1::LUC+ bioluminescent plants overexpressing the candidate novel clock gene BBX31 display lengthened period compared to control plants.
Figure 3.6 Additional BBX31 overexpression results. Plants overexpressing BBX31 display lengthened period relative to control plants.
Figure 3.7 Flowering time in BBX31 overexpression plants. Overexpression of BBX31 leads to delayed flowering time relative to controls.
Figure 3.8 Quantification of flowering time delay in BBX31 OX plants and bbx31 mutant plants relative to controls.
Figure 3.9 LHY expression patterns are altered in BBX31 overexpression and bbx31 mutant plants. This functionally links the phenotypes observed in the overexpression screen to the regulation of a key circadian clock gene.
Figure 3.10 BBX31 transcription factor interactors in TF library yeast two-hybrid screen. BBX family of proteins is highly overrepresented relative to their overall abundance in the library.
**Table 3.1** List of notable hits from the BBX31 genome wide transcription factor yeast two-hybrid screen.

**BBX31 Yeast 2-Hybrid ‘Hits’**

<table>
<thead>
<tr>
<th>AGI</th>
<th>Gene Name</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G24790</td>
<td>CONSTANS-LIKE 3 (COL3)</td>
<td>3.79</td>
</tr>
<tr>
<td>AT5G24930</td>
<td>CONSTANS-LIKE 4 (COL4)</td>
<td>4.40</td>
</tr>
<tr>
<td>AT5G57660</td>
<td>CONSTANS-LIKE 5 (COL5)</td>
<td>5.29</td>
</tr>
<tr>
<td>AT3G07650</td>
<td>CONSTANS-LIKE 9 (COL9)</td>
<td>3.78</td>
</tr>
<tr>
<td>AT5G48250</td>
<td>CONSTANS-LIKE 10 (COL10)</td>
<td>3.74</td>
</tr>
<tr>
<td>AT2G47890</td>
<td>CONSTANS-LIKE 13 (COL13)</td>
<td>6.02</td>
</tr>
<tr>
<td>AT1G28050</td>
<td>CONSTANS-LIKE 15 (COL15)</td>
<td>3.68</td>
</tr>
<tr>
<td>AT4G39070</td>
<td>BBX20</td>
<td>3.03</td>
</tr>
<tr>
<td>AT5G47640</td>
<td>NUCLEAR FACTOR Y, SUBUNIT B2 (NF-YB2)</td>
<td>4.60</td>
</tr>
<tr>
<td>AT3G47620</td>
<td>TCP14</td>
<td>6.32</td>
</tr>
<tr>
<td>AT1G73730</td>
<td>ETHYLENE-INSENSITIVE3-LIKE 3 (EIL3)</td>
<td>6.44</td>
</tr>
<tr>
<td>AT1G18330</td>
<td>EARLY-PHYTOCHROME-RESPONSIVE1 (EPR1) (RVE7)</td>
<td>4.00</td>
</tr>
<tr>
<td>AT3g46640</td>
<td>LUX</td>
<td>2.33</td>
</tr>
<tr>
<td>AT1G04550</td>
<td>BODENLOS</td>
<td>2.43</td>
</tr>
</tbody>
</table>
Figure 3.11 Model for regulatory role of BBX31 in clock and flowering time pathways (adapted from (de Montaigu et al., 2010)). BBX31 regulates the circadian clock via its affect on LHY expression. Additionally, BBX31 may play a role in downstream flowering time pathways.
References


Irelan JT, Gutierrez Del Arroyo A, Gutierrez A, Peters G, Quon KC, Miraglia L, et


Kaufmann K, Muiño JM, Østerås M, Farinelli L, Krajewski P, Angenent GC. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nature Protocols. 2010;5(3):457–72.

Kawai J, Hayashizaki Y. DNA book. Genome Res. 2003 Jun;13(6B):1488–95. PMCID: PMC403695


Locke JCW, Southern MM, Kozma-Bognár L, Hibberd V, Brown PE, Turner MS,


Shiu S-H, Shih M-C, Li W-H. Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiol. 2005 Sep;139(1):18–26. PMCID: PMC1203354


