University of Nevada, Reno

Characterizing the transcriptional regulation of crassulacean acid metabolism in *Kalanchoe*

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

by

Travis M. Garcia

Dr. John C. Cushman / Dissertation Advisor

December 2021

THE GRADUATE SCHOOL

We recommend that the dissertation prepared
under our supervision by

TRAVIS M. GARCIA

Entitled

Characterizing the transcriptional regulation of crassulacean acid metabolism in *Kalanchoe*

be accepted in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

John C. Cushman, Advisor

Jeffrey F. Harper, Committee Member

Dylan K. Kosma, Committee Member

Won C. Yim, Committee Member

Juan Solomon, Graduate School Representative

David W. Zeh, Ph. D., Dean, Graduate School
December, 2021

Abstract

Due to the agricultural challenges posed by the prospect of a hotter drier climate understanding the molecular basis of plant water-use efficiency is of increasing importance. Species performing crassulacean acid metabolism (CAM) photosynthesis have evolved to be naturally water-use efficient primarily through shifting their carbon uptake to night to minimize water-loss. Relative to C_3 and C_4 photosynthesis species, CAM plants are enriched for rhythmic circadian clock-dependent regulation of metabolic processes. However, the transcriptional regulation of CAM remains largely uncharacterized. Using *Kalanchoe fedtschenkoi*, in which CAM develops along a leaf developmental gradient, candidate transcription factors with possible CAM-related functions were identified. The mRNA abundance of these transcription factors increases upon the transition from C_3 photosynthesis to CAM and they appear to exhibit a circadian phase-dependent pattern of regulation. To better characterize the transcriptional control circuits underlying CAM, three such of these transcription factors, *KfNF-YB3*, *KfHomeodomain-like*, and *KfMYB59* were selected for chromatin immunoprecipitation-sequencing (ChIP-seq). However, these experiments failed to identify enriched target genomic loci possibly as a consequence of the unique challenges of adapting experimental protocols designed for model C_3 photosynthesis plant species to a succulent plant such as *Kalanchoe*. Additionally, this work focuses on elucidating the *cis*-regulatory elements and the *trans*-acting factors governing the transcriptional control of the phospho*enol*pyruvate carboxylase gene (*Ppc1*) in *Kalanchoe*. Despite this enzyme's importance in catalyzing the primary nocturnal fixation of CO_2 in CAM species, the complex regulatory mechanisms underlying its expression are not well-studied. We

examined the *Kalanchoe Ppc1* promoter and identified numerous *cis*-regulatory elements on the basis of their sequence conservation with known regulatory modules. These individual elements along with two-hundred base pair region segments of the *Kalanchoe Ppc1* promoter were used at bait probes in yeast one-hybrid (Y1H) assays. From this analysis, several high-confidence interacting transcriptional regulators were identified including *ERF9*, *ERF106*, *TCP4*, and *PIF1*. *In silico* examination of the *Ppc1* promoter revealed likely binding sites for these factors based on homology to validated preferred binding sequences in *Arabidopsis*. The specific transcription factors identified through this work can now serve as the basis for further experiments to confirm interaction with the *Ppc1* promoter and elucidate the nature of their regulatory effects. Overall, the work presented in this dissertation attempts to investigate the transcriptional control of crassulacean acid metabolism using the developmental CAM model *Kalanchoe*.

Acknowledgements

This dissertation would never have been possible without the expertise and support I have received from so many over these years. I would like to extend a huge thanks to Dr. John Cushman for giving me the opportunity to work in his lab on so many interesting and challenging projects. Your experience, knowledge, and insights have been such a valuable resource throughout so many of the technical hurdles encountered in the course of my experiments. I would like to thank all members of the Cushman lab, past and present, whose time and effort has helped facilitate the completion of my projects. I would especially like to thank Dr. Won C. Yim, Dr. Sung Don Lim, Dr. Bernard Wone, and Dr. Jungmin Ha whose detailed knowledge in such areas as bioinformatics and cloning has been so valuable to me over these years. My deepest gratitude to my past undergraduate research interns, especially Kristen Straight. Your time spent assisting me with laborious and time-consuming plant transformation and tissue culture work is greatly appreciated. A huge thanks to our lab manager Lisa Petrusa whose excellent work in ensuring the smooth operation of our lab has been so valuable in the planning and implementation of my experiments.

I would like to express my deepest thanks to the members of my advisory committee Dr. Won C. Yim, Dr. Jeffery Harper, Dr. Dylan Kosma, Dr. Juan Solomon, and my former committee members Dr. David K. Shintani, Dr. Hongfei Lin and Dr. Glenn Miller. All of your input and guidance has helped me grow my research and scientific writing skills. I would also like to thank many other members of the students, faculty, and staff of the UNR Biochemistry and Molecular Biology Department in particular Dr. Ian Wallace, Elizabeth Brown, Rebecca Young, Chrystle Weigand, Elijah

Holschbach, and Dr. James Davis. Your advice has been so helpful to me when troubleshooting unfamiliar techniques and protocols. Lastly, I would like to express my most sincere thanks to my parents Ken and Audrey whose constant unwavering support has allowed me to successfully bring this work to completion.

Table of Contents

List of Tables……viii

List of Figures……..ix

Chapter 1: Introduction……...1

The threat of climate change to agriculture and the relevance of CAM……..2

The biochemistry and regulation of CAM……...5

CAM *vs.* C_3 photosynthesis comparative transcriptomics…….....12

Transcription factors and plant stress tolerance…….....14

Approaches to study DNA-protein interactions…….....21

Kalanchoe species as models for studying CAM…….....26

References…….....28

Chapter 2: Genome-wide binding site profiling of *KfNF-YB3, KfHD-like,* and *KfMYB59,* three transcription factors from *Kalanchoe fedtschenkoi* with putative roles in the regulation of crassulacean acid metabolism…….39

Abstract……..40

Introduction…….....41

Materials and Methods…….....46

C_3 *vs.* CAM comparative RNA-seq and choice of transcription factors for ChIP-seq…….....46

Plant material, maintenance, and growth conditions…….....47

Design and construction of *Kalanchoe* expression constructs…….....47

Transient expression subcellular localization assays in *Kalanchoe fedtschenkoi*…….....49

Stable transformation of *Kalanchoe fedtschenkoi*……50

Confirmation and expression analysis of stably transformed *Kalanchoe* lines……52

Kalanchoe nuclear isolation…….....53

Chromatin immunoprecipitation……..56

ChIP-sequencing and bioinformatics processing of data…….....59

Results…….....59

NF-YB3, *HD-like*, and *MYB59* are differentially expressed in C_3 photosynthesis *vs*. CAM……..59

KfNF-YB3, *KfHD-like*, and *KfMYB59* are nuclear localized in *Kalanchoe fedtschenkoi*……....60

The INTACT system was successfully expressed in *Kalanchoe fedtschenkoi*……61

ChIP-seq analysis of *KfNF-YB3*, *KfHD-like*, and *KfMYB59* failed to identify enriched loci…….61

Predicted binding motifs lack significance and are composed of repetitive elements…….62

Discussion……....63

Acknowledgments……..68

References…….....69

Chapter 3. Identification of putative transcriptional regulators using the *Kalanchoe* phospho*enol*pyruvate carboxylase 1 (*Ppc1*) promoter…….95

Abstract……..96

Introduction……..97

Materials and Methods…….....102

Plant material, maintenance, and growth conditions……..102

Kalanchoe Ppc1 promoter analysis and design of yeast one-hybrid bait probes……..102

Yeast one-hybrid experiments……104

In silico analysis of the *Kalanchoe fedtschenkoi Ppc1* promoter……105

Transient expression subcellular localization assays in *Kalanchoe fedtschenkoi*…….....106

Results……...108

Y1H assays reveal several high-confidence interactions with *Ppc1* promoter elements……....108

The *Ppc1* promoter contains likely binding sites for Y1H-predicted interacting factors……109

KfERF9, *KfERF106*, and *KfTCP4* are nuclear localized in *Kalanchoe fedtschenkoi*……....110

Discussion……...111

Acknowledgments…….115

References……....116

Chapter 4. Concluding remarks……..132

References……...139

List of Tables

Chapter 2

Supplemental Table 1: Primers used in this study……83

Supplemental Table 2: Highly abundant (C_3/CAM mean > 100 FPKM) *Kalanchoe laxiflora* transcription factors displaying a $\log_2$ fold-change of at least 2 in CAM *vs.* C_3……...84

Chapter 3

Table 1: High-confidence interacting transcription factors from yeast one-hybrid screens……120

Supplemental Table 1: Bait probes for yeast one-hybrid screens…..125

Supplemental Table 2: Primers used in this study…….126

Appendix Table 1. Full list of prey clone interactions for all yeast one-hybrid bait probes……..142

List of Figures

Chapter 2

Figure 1: Fold-induction of *NF-YB3*, *homeodomain-like superfamily protein*, and *MYB59* from C_3 to CAM in *K. laxiflora* as determined by comparative transcriptomics RNA-seq analysis between leaf pair 1 (C_3) and leaf pair 6 (CAM)…….73

Figure 2: 24 h diel time course of mRNA abundance of *K. fedtschenkoi NF-YB3*, *homeodomain-like superfamily protein*, and *MYB59*……...74

Figure 3: NF-YB3, homeodomain-like superfamily protein, and MYB59 are nuclear localized in *K. fedtschenkoi*……..75

Figure 4: The NTF protein is expressed in *K. fedtschenkoi* mesophyll cells under the control of the 2.8 kb *KlPpc1* promoter……76

Figure 5: Phred quality score distribution of each base position from Illumina PE reads of ChIP-seq sample replicates and controls……77

Figure 6: Percentages of mapped *vs.* unmapped reads for each ChIP-seq sample replicate and controls……77

Figure 7: Total called peak count for each *KfNF-YB3*, *KfHD-like* and *KfMYB59* sample replicate……78

Figure 8: Genome-wide distribution of ChIP-seq sample peaks in relation to gene bodies……79

Figure 9: Top 25 most significant motifs at peak regions for *KfNF-YB3*……80

Figure 10: Top 25 most significant motifs at peak regions for *KfHD-like*……81

Figure 11: Top 25 most significant motifs at peak regions for *KfMYB59*…….82

Supplemental Figure 1: Complete coding sequence and amino acid sequence of *K. fedtschenkoi NF-YB3*……...84

Supplemental Figure 2: Complete coding sequence and amino acid sequence of *K. fedtschenkoi HD-like*……85

Supplemental Figure 3: Complete coding sequence and amino acid sequence of *K. fedtschenkoi MYB59*…….87

Supplemental Figure 4: Proximal *K. laxiflora Ppc1* promoter region used to drive expression of the NTF peptide in the *Kl*p*Ppc1*-INTACT construct……88

Supplemental Figure 5: Plasmid constructs used in this study…….90

Supplemental Figure 6: *Agrobacterium*-mediated stable transformation protocol for *K. fedtschenkoi*……92

Supplemental Figure 7: Validation of *K. fedtschenkoi* stable transformants and semi-quantitative expression analysis of TF-3xHA fusion transgenic lines……..93

Chapter 3

Figure 1: *KfPpc1* promoter locations of MEME IDs associated with putative binding sites of high-confidence interacting transcription factors from yeast one-hybrid screens……...121

Figure 2: *KfPpc1* promoter locations of putative *cis*-regulatory elements used as yeast one-hybrid screen bait probes…….122

Figure 3: *KfPpc1* promoter locations of MEME IDs associated with putative binding sites of high-confidence interacting transcription factors from yeast one-hybrid screens and locations of putative *cis*-regulatory elements used as bait in these screens…….123

Figure 4: ERF9, ERF106, and TCP4 are nuclear localized in *K. fedtschenkoi*…….124

Supplemental Figure 1: Local (Smith-Waterman) alignment of the *Kalanchoe fedtschenkoi* and *Kalanchoe laxiflora Ppc1* promoters……127

Supplemental Figure 2: Complete coding sequence and amino acid sequence of *K. fedtschenkoi ERF9*……...128

Supplemental Figure 3: Complete coding sequence and amino acid sequence of *K. fedtschenkoi ERF106*……129

Supplemental Figure 4: Complete coding sequence and amino acid sequence of *K. fedtschenkoi TCP4*……130

Supplemental Figure 5: Plasmid constructs used in this study……..131

Chapter 1. Introduction

Authors:

Travis M. Garcia[1]

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557-0330 USA

Introduction

The threat of climate change to agriculture and the relevance of CAM

An overwhelming body of scientific evidence now supports the conclusion that anthropogenic climate change is a real phenomenon and is occurring at an increasingly accelerated rate due to the unprecedented release of greenhouse gases (GHGs) into the atmosphere (Pareek, 2010). Recent analysis conducted by NASA found that the average global temperature has increased by approximately 0.8 °C since 1880 and that two-thirds of this temperature increase has occurred from 1975 to the present date (Janni et al., 2020). This warming trend poses a severe threat to crop productivity at a time when models suggest that agricultural productivity might need to be increased as much as 70 to 100% by the year 2050 to ensure a stable food supply for a projected population of 9.8 billion people (Wang et al., 2018). Climate projections indicate that the average global temperature will increase by 2.5 to 4.5 °C by the end of the 21st century, which will not only pose severe challenges to agricultural sustainability, but also submerge large areas of costal crop-growing regions further limiting land area available for agriculture (Pareek, 2010; Wang et al., 2018). Even with stringent mitigation measures, modeling indicates that the magnitude of droughts could double for 30% of global land area by the end of the 21st century. This could result in a marked increase in the frequency of drought to the extent that current 100-year drought events could occur once every 2 to 5 years (Naumann et al., 2018).

Numerous examples exist of the effect climate change is already having on production of important crop species. An analysis of global production of wheat, barley,

and maize from 1981 to 2002 found a small but measurable negative effect of climate change on the production of these crops that likely offsets some of the increases in yields of these species observed in recent decades due to advances in agricultural technology (Lobell & Field, 2007). When factors such as CO_2 levels and advances in agronomy are accounted for global simulations show a 4.1, 1.8 and 4.5% climate-dependent reduction in the yield of maize, wheat, and soybean, respectively, relative to a pre-industrial climate (Iizumi et al., 2018). An additional agricultural burden caused by climate change is likely to be interannual yield variability. A meta-analysis conducted by Challinor et al. (2014) predicts yield variability with an overall loss in aggregate production for wheat, rice, and maize even with just 2 °C in global warming.

In view of the challenges posed to sustainable agriculture as a result of climate change, crop-improvement research efforts have increasingly focused on understanding the biochemistry and molecular basis of the variety of adaptive traits plants have evolved to combat abiotic stress (Niechayev et al., 2019). As drought has and will continue to be a major component of climate-related stress much of this research has been directed toward gaining a better understanding of photosynthetic pathways that confer a high degree of plant water-use efficiency (WUE), defined as unit CO_2 fixed per unit H_2O lost (Davis et al., 2014; Chaudhry & Sidhu, 2021). In comparison with species that perform C_3 photosynthesis, plants possessing either C_4 or crassulacean acid metabolism (CAM) photosynthesis have evolved improved WUE and unique CO_2 acquisition strategies in the form of carbon-concentrating mechanisms to concentrate CO_2 around Rubisco and thus suppress this enzyme's oxygenase activity. These evolutionary adaptations, driven ultimately by the poor productivity of C_3 photosynthesis species in high-temperature and

high-light environments due to photorespiration, likely evolved as a consequence of both water-limited environments and fluctuating CO_2 levels over Earth's history (Keeley & Rundel, 2003). Today, both pathways are widespread among terrestrial plants with C_4 photosynthesis present in at least 66 lineages of angiosperms across 19 plant families, while CAM is found in over 400 genera across 36 families (Edwards, 2019; Kellog, 2013; Yang et al., 2015).

Owing to their improved WUE over C_3 photosynthesis crops and hence potential for significantly greater productivity on arid marginal lands, both C_4 photosynthesis and CAM have been considered as candidates for pathway engineering into C_3 photosynthesis species. Perhaps the most well-known effort of this kind is the C_4 Rice Project whose end goal is to engineer full C_4 photosynthesis into *O. sativa* to create a high-yielding variety of rice (Karki et al., 2013). Alternatively, engineering of CAM into C_3 photosynthesis species has also been proposed as a potentially more feasible approach to improve crop productivity in arid marginal environments (Borland et al., 2014; Yang et al., 2015; Schiller & Bräutigam, 2021).

Unlike C_4 photosynthesis, which requires the integrated functioning of specialized mesophyll and bundle sheath cells to pump CO_2 to the site of Rubisco, CAM does not require these two differentiated cell types with divergent metabolic functions and specializations. Rather, CAM is a single-cell trait that concentrates CO_2 within each individual mesophyll cell through a temporal separation between carboxylation and decarboxylation. Many facultative CAM species exist such as *M. crystallinum* (common or crystalline ice plant), which can transition between C_3 photosynthesis and CAM depending on environmental conditions (Cushman & Borland, 2002). Such plasticity

suggests that a cell performing C_3 photosynthesis might be successfully induced to perform the entire CAM pathway given the appropriate transcriptional and metabolic re-programming (Borland et al., 2014; Schlüter & Weber, 2016; Yang et al., 2015).

The biochemistry and regulation of CAM

The superior water-use efficiency of CAM, which is at least six-fold higher than that of C_3 photosynthesis plants and as much as three-fold higher than C_4 species, is due primarily to the nocturnal uptake and primary fixation of CO_2 with subsequent decarboxylation and fixation by Rubisco in the following light period behind closed stomata (Borland et al., 2009). This entire process can be divided into four basic phases across a 24 h period (Osmond, 1978). Phase I commences at the end of the photoperiod with the opening of stomata. In the cytosol, the action of carbonic anhydrase (CA) converts CO_2 into HCO_3^- that is then fixed by the action of phospho*enol*pyruvate carboxylase (PEPC) using the 3C substrate phospho*enol*pyruvate. This reaction forms oxaloacetate that is then converted to malate by the action of malate dehydrogenase. Malate represents the primary 4C storage molecule for this initial nocturnal fixation of CO_2 and is then transported into the central vacuole via a putative voltage-gated ion channel. CO_2 uptake and malate accumulation continue for much of the dark period (Borland et al., 2009; Dodd et al., 2002). Phase II occurs around dawn and is characterized by some direct fixation of CO_2 by the action of Rubisco along with some continued contribution to fixation by PEPC, which at this time becomes deactivated (Ceusters et al., 2011). Phase III of CAM involves the decarboxylation of malate through

the action of various malic enzymes. The precise malic enzyme(s) used in this phase vary by species and may include mitochondrial NAD^+-ME, cytosolic $NADP^+$-ME, chloroplastic $NADP^+$ -ME, or cytosolic PEPCK (Lim et al., 2019). This decarboxylation of malate then yields pyruvate, phospho*enol*pyruvate and CO_2. High internal partial pressures of CO_2 are observed at this stage promoting the carboxylase activity of Rubisco while suppressing its oxygenase function (Luttge, 2002). The high partial pressure of CO_2 may contribute to stomatal closure during this phase. Phase IV of CAM occurs near the end of the photoperiod and involves the reactivation of PEPC. In well-watered CAM plants this phase may also involve the re-opening of stomata and direct fixation of atmospheric CO_2 by Rubisco. Drought stress has been found to suppress this stage and instead promotes a shift to greater CO_2 uptake at night (Borland et al., 2009; Dodd et al., 2002; Males & Griffiths, 2017). These phases are generalities and exhibit much variation among species and are highly plastic based on environmental conditions. This plasticity allows for the maximizing of carbon gain while minimizing water loss (Ceusters et al., 2011). Due to the key role of phospho*enol*pyruvate carboxylase in CAM and the potential for improvement of photosynthetic and WUE, which may be possible through manipulation of this enzyme in important crop species, much research has been directed toward understanding its function and regulation (Deng et al., 2016; Liu et al., 2021). The transcriptional regulation of *Ppc1* and the *cis*-regulatory elements governing its expression characteristics will be a particular focus in this body of work. In vascular plants, *PEPC* isoforms are encoded by a multi-gene family comprising both photosynthetic and non-photosynthetic isoforms. Non-photosynthetic forms of *PEPC* play diverse roles in cellular metabolism including maintenance of cellular pH,

replenishment of citric acid cycle enzymes, transport of ions in root cells, and control of stomata. Isoforms have also been identified with roles in seed germination and fruit ripening (Christin et al., 2014; Silvera et al., 2014). The *Ppc* gene family in plants generally consists of three subfamilies, *Ppc1*, *Ppc2*, and *Ppc3*. Phylogeny analysis indicates that all of these modern forms of *Ppc* in vascular plants share a common ancestor in green algae (Chlorophyta) (Deng et al., 2016). Both C_4 photosynthesis and CAM-specific isoforms of *Ppc* appear to have arisen through the recruitment of *Ppc1* isoforms within the caryophyllales eudicot clade via multiple independent evolutionary events (Christin et al., 2014; Deng et al., 2016).

Regulation of photosynthetic CAM PEPC appears to be largely governed by circadian rhythms and occurs at both the transcriptional and post-transcriptional level (Hartwell et al., 2002; Theng et al., 2007). At the end of the photoperiod, PEPC undergoes a reversible phosphorylation at a specific serine residue near the N-terminus through the action of phospho*enol*pyruvate carboxylase kinase (PEPC kinase). This regulatory kinase, which itself is under strict circadian control, modulates the allosteric sensitivity of PEPC to both positive and negative regulation (Taybi et al., 2000). In its phosphorylated state, PEPC displays a decreased sensitivity to its negative allosteric regulator L-malate and an increased sensitivity to its substrate PEP and positive allosteric effectors such as glucose-6-phosphate (Bakrim et al., 2001; Theng et al., 2007). Dephosphorylation of PEPC occurs as a result of the action of protein phosphatase 2A, an enzyme which displays constant activity across the entire circadian cycle. This dephosphorylation then renders PEPC more sensitive to allosteric inhibition by malate as the next photoperiod begins (Nimmo, 2000; Taybi et al., 2004). Considering the

correlation in protein abundance of phosphorylated PEPC with PEPC kinase activity along with the observation that protein phosphatase 2A exhibits little change in activity across a 24 h period, a major control "switch" for PEPC activity likely lies with the action of PEPC kinase (Hartwell at al., 2002; Nimmo, 2000).

Along with its phosphorylation status, transcriptional control of the expression of CAM *Ppc* isoforms is also likely to be an important component of their spatiotemporal regulation (Silvera et al., 2014). Indeed, a major aspect of this dissertation is the investigation of the *cis*-regulatory elements and their trans-acting factors controlling the mesophyll specificity and nocturnal-timed expression of CAM photosynthetic *Ppc*. Beyond *Ppc*, the *cis*-regulatory modules and their interacting factors of other abiotic stress-responsive and photosynthetic genes are also the focus of much research due to their potential as targets for crop improvement (Franco-Zorrilla et al., 2014; Galli et al., 2020).

Cis-elements controlling temporally-phased gene expression

Results of a genome-wide study in *Arabidopsis* by Michael et al. showed that 89% of transcripts cycle under thermocycles, photocycles, the circadian clock, or a combination of two or all of these conditions. Through a scan of overrepresented *cis*-regulatory elements from 500 bp promoter regions, three distinct time-of-day transcriptional modules were identified that collectively cover the entire 24 h cycle. Furthermore, the common core sequences of these elements were conserved in other selected species (Michael et al., 2008). Analysis of *cis*-regulatory elements (CREs) in the

pineapple genome by Ming et al. established the first link between CAM photosynthesis and the circadian clock through leveraging the data obtained by Michael et al. Through examination of 1 kb promoter regions of CAM isoform carbon fixing genes the authors discovered an enrichment of circadian associated CREs, establishing a basis for the observed diel expression patterns of CAM genes in photosynthetic cells (Ming et al., 2015). Later research showed that the phasing of CAM photosynthetic genes was not only due to the acquisition of divergent circadian clock-associated CREs relative to C_3 photosynthesis species, but also was due to the evolution of novel stress-responsive CREs that can affect time-of-day gene expression (Wai et al., 2017).

Through whole-transcriptome RNA-seq analysis of *Sedum album*, a species that switches from C_3 photosynthesis to CAM under conditions of drought stress, Wai et al. (2019) found 8 *cis*-elements enriched in the promoters of genes upregulated in response to drought, which were not enriched in genes expressed under well-watered conditions. Studies such as this illustrate the need for facultative CAM models in order to fully understand the transcriptome re-wiring between these two types of photosynthesis. As both C_3 photosynthesis and CAM can be studied within a single species, insights can be gained that cannot be obtained through studying obligate CAM species only (Wai et al., 2019). Research is also underway with regards to understanding the *cis*-regulatory basis of the reverse pattern of stomatal conductance observed in CAM species relative to C_3 and C_4 photosynthesis plants. Promoter CRE scanning of genes with stomata-related functions in pineapple, orchid, *Agave*, *Arabidopsis*, rice, and *Sorghum* by Chen at al. discovered an enrichment of motifs associated with *ERF7*, *ERF73*, and *ABR1* transcription factor binding in the CAM-performing species in addition to the expected

circadian-associated motifs, which were enriched across all types of photosynthesis (Chen et al., 2020). Other studies such as Mosely et al., who performed a diel expression time-course comparing *Kalanchoe fedtschenkoi*, *Arabidopsis*, and *Solanum lycopersicum*, identified new stomata-related gene candidates with a re-scheduled expression in CAM *versus* C_3 photosynthesis, which could now be investigated for overrepresented *cis*-regulatory elements (Moseley et al., 2019). Returning to *Ppc*, various studies have shed light on the *cis*-regulatory basis of the stress-responsiveness and mesophyll-specific expression of CAM and C_4 photosynthetic *Ppc1* isoforms. Work by Cushman & Bohnert reported the discovery of two A/T rich *Ppc1* promoter sites with a loose consensus sequence in the facultative CAM model *Mesembryanthemum crystallinum*. Electrophoretic mobility shift assays (EMSAs) indicated that the binding activity of these sites increased when nuclear extracts from salt-stressed CAM-performing *M. crystallinum* plants were used relative to extracts from unstressed C_3 photosynthesis-performing tissue (Cushman & Bohnert, 1992). Further research on the *M. crystallinum Ppc* promoter by Schaeffer et al. discovered several sites with sequence similarity to MYB transcription factor binding sites. Through a *Ppc* promoter deletion series using a GUS reporter, this study also identified several enhancer regions and one silencer region that contribute to the induction of *Ppc* in response to salt stress (Schaeffer et al., 1995). The induction of *Ppc* in response to abiotic stress is not a feature unique to CAM species. A study by Wang et al. surveyed the genome of the C_3 photosynthesis crop *Glycine max* (soybean) and identified a total of 10 *Ppc* genes. All ten of these isoforms shared a total of 19 conserved CREs in the 1500 bp region upstream of their transcriptional start sites. Many of these CREs are known to be associated with transcriptional responses to both

abiotic and biotic stresses. Almost all identified soybean *Ppc* isoforms displayed transcript abundance changes in response to at least one abiotic stress with four of the ten isoforms responding to salinity and/or water-deficit stress (Wang et al., 2016). Lastly, the importance of CREs to tissue-specific expression patterns of photosynthetic genes has also been examined. Due to the cell-type partitioning of photosynthetic enzymes in C_4 photosynthesis species these plants are particularly useful models for these studies (Schlüter & Weber, 2016). Akyildiz et al. reported on the importance of a mesophyll expression CRE module, MEM1, in the C_4 species *Flaveria trinervia*. This module consists of two submodules, one of which has undergone an SNP mutation and one which has gained a tetranucleotide insertion compared to the ancestral C_3 photosynthesis MEM1 module. These sequence changes, which were found to be necessary and sufficient for mesophyll-specific expression of C_4 photosynthesis genes, also suppressed transcription of genes exhibiting bundle sheath cell expression (Akyildiz et al., 2007). Bundle sheath specific expression of the photosynthetic maize C_4 malic enzyme NADP-ME was examined in Borba et al. (2018). Their analysis showed that the promoter of this gene has acquired a pair of *cis*-elements with G-box-like architecture, which appears to have arisen from a single ancestral C_3 photosynthesis *cis*-element. Gel-shift assays using this pair of CREs from the *NADP-ME* promoter showed the binding of the maize transcription factors *ZmbHLH128* and *ZmbHLH129* in an enhanced manner relative to the single C_3 photosynthesis *cis*-element (Borba et al., 2018).

CAM vs. C_3 photosynthesis comparative transcriptomics

In addition to studies which focus mainly on understanding the regulation of known photosynthetic genes or on identifying gene candidates involved in certain biological functions, genome-wide comparative transcriptomics between C_3 photosynthesis and CAM is also providing promising candidate genes for synthetic biology approaches to crop improvement. Through computational modeling and biological network analysis of C_3 photosynthesis *vs.* CAM differential transcript profiling data critical cellular control points and their associated regulatory factors could potentially be identified (DePaoli et al., 2014). One of the first reports on differential gene expression profiles in C_3 photosynthesis and CAM was Kore-eda et al. which performed large-scale EST profiling in well-watered and salinity-stressed *M. crystallinum*. A transition from C_3 to CAM was reflected in the decrease in transcripts encoding C_3 photosynthesis-associated enzymes, such as Rubisco, and an increase in CAM-related photosynthetic enzymes. Additionally, EST libraries from salinity-stressed samples were enriched for transcripts with assigned functions in diverse abiotic and biotic stress responses (Kore-eda et al., 2004). *M. crystallinum* was again used in Cushman et al. to investigate the circadian-dependent gene expression changes in C_3 photosynthesis *vs.* CAM. This study compared mRNA abundance across a 24 h time period for both well-watered and salinity-stressed plants with custom microarrays. The results of this study illustrate the marked circadian clock-dependent phase-shift of many CAM-related transcripts to the subjective evening and highlights the need for a better understanding of circadian clock-regulated transcriptional activators and repressors (Cushman et al., 2008). The discovery of transcription factors potentially involved in CAM-induction has been

accelerated by recent studies in which RNA-seq has been utilized for C_3 photosynthesis *vs*. CAM comparative gene expression profiling. The facultative CAM tropical dicot species *Talinum triangulare* is capable of transitioning from C_3 photosynthesis to weak CAM in response to drought stress and was used in Brilhaus et al. to generate a list of candidate transcription factors which underlie this photosynthetic plasticity. This study identified a total of 1,449 transcription factors in *T. triangulare* 40% of which exhibited differential regulation between well-watered and drought-stressed plants. A closer examination of the functional annotations of the top 25 upregulated transcription factors revealed roles in abscisic acid (ABA) pathway signaling, chlorophyll degradation, heat shock responses, and seed germination under salt-stress. Moreover, the authors found that the differentially expressed transcription factors could be broadly categorized into those exhibiting a sustained response under drought conditions and those that displayed a transient response. Enrichment of roles in ABA pathway signaling were observed for those transcription factors with a sustained response, while transcription factors transiently upregulated in response to drought stress were often components of the dehydration-responsive element binding protein / C-repeat binding factor (DREB/CBF) pathway. Overall, these drought-stress upregulated transcription factors in *T. triangulare* spanned numerous diverse structural classes with known roles in both biotic and abiotic stress signaling (Brilhaus et al., 2016). Candidate CAM-related transcription factors were also identified in a report by Zhang et al. using the model CAM species *Kalanchoe fedtschenkoi*. This study, which examined the genome-wide transcriptional changes in this species in response to light of different qualities and intensities across a photoperiod, generated co-expression networks of transcription factors with core circadian clock

genes, CAM enzymes, and stomatal biology genes. Using this approach, several specific transcription factors were identified as nodes within these networks. While overall these co-expression networks were found to be largely conserved between *K. fedtschenkoi* and *Arabidopsis,* several transcription factors with putative CAM-related functions were elucidated. These transcription factors, whose roles are currently unknown, may function as novel regulators of circadian clock-dependent CAM regulation (Zhang et al., 2020).

Transcription factors and plant stress tolerance

A detailed study of a number of specific transcription factors with putative CAM-related functions is a major overarching goal of this dissertation. The motivation for this detailed characterization of particular *trans*-acting factors is the potential utility of a transcription factor-centered approach to CAM engineering and abiotic stress tolerance improvement more generally. As transcription factors are frequently master regulators of diverse cellular responses to the environment and often regulate and interact with other transcriptional activators and repressors, bioengineering of a small number of transcription factors may enable genome-wide transcriptional repatterning (Amin et al., 2019). Numerous reports detail the successful improvement of abiotic stress tolerance and biomass productivity in model and crop plants through the overexpression of single transcription factors. For example, overexpression of the maize transcription factor *ZmWRKY58* in rice led to higher rates of survival following drought and salinity stress (Cai et al., 2014). Similarly, Chen et al. demonstrated improved tolerance to multiple abiotic stresses in rice following the overexpression of the rice transcription factor

OsNAP. Microarray analysis of overexpressing lines revealed that this transcription factor upregulated multiple downstream transcriptional activators and activated several stress responsive pathways (Chen et al., 2014). Lim et al. described the improvement of biomass productivity and reproductive yield in *Arabidopsis thaliana* and *Nicotiana sylvestris* through the overexpression of the wine grape bHLH transcription factor *VvCEB1*. RNA-seq analysis of overexpressing lines revealed a large-scale reprogramming of auxin-mediated cellular functions (Lim et al., 2018). Further phenotypic characterization of *VvCEB1*-overexpressing *Arabidopsis* lines showed an increase in tissue succulence, cell size, and a decrease in intercellular air space. WUE and salinity tolerance were also improved, presumably due to increased stomatal closure and dilution of ions as a result of enhanced succulence, respectively (Lim et al., 2020). However, relatively unexplored as an approach is the identification of high hierarchical transcriptional regulators in CAM species, halophytes, and resurrection plants followed by their heterologous expression in C_3 photosynthesis model organisms or crop species (Amin et al., 2019). In one of the few studies of this kind, the *CpMYB10* from *Craterostigma plantagineum*, a resurrection plant with the ability to survive complete desiccation of its vegetative tissue, was functionally characterized in *Arabidopsis*. In *Craterostigma MYB10* has no detectable expression under fully-hydrated conditions, but was rapidly induced under dehydration or ABA treatment. Enhanced tolerance to dessication and salt stress was observed upon overexpression in *Arabidopsis* along with an ABA hypersensitive phenotype (Villalobos et al., 2004). The identification and characterization of putative master regulators of CAM-associated transcriptional circuits is a major focus of the work presented in this dissertation and will be discussed in-detail

in the following chapters. Through the use of *Kalanchoe* species as a developmental CAM model, several transcription factors were identified that display a marked increase in transcript abundance upon the transition from C_3 photosynthesis to CAM or appear to be involved in the direct regulation of photosynthetic *Ppc* through interactions with CREs in the *Kalanchoe Ppc* promoter. These specific *trans*-acting factors, which will form the basis of the experiments in this dissertation's ensuing chapters, include transcription factors from the NF-Y, homeodomain, MYB, AP2/ERF, and TCP families.

Numerous reports have established the critical role of members from all the transcription factor classes listed above in plant responses to both biotic and abiotic stimuli. Present in all eukaryotes, transcription factors belonging to the Nuclear Factor Y (NF-Y) class, also known as CCAAT-binding factors (CBFs), are known to generally function as heterotrimeric complexes. The existence of multiple NF-YA, NF-YB, and NF-YC subunits provides for hundreds of theoretically possible heterotrimeric combinations, although only a subset of all possible combinations has been observed *in vivo*. The particular subunits that assemble to form a complete heterotrimer appear to be expressed in both a tissue-specific and species-specific manner. Furthermore, some subunits interact more promiscuously than others, are only expressed at certain developmental stages, and can even interact with other classes of transcription factors while not associated with other NF-Y subunits (Laloum et al., 2013; Petroni et al., 2013). Exhibiting both positive and negative regulatory effects on their target genes, NF-Y transcription factors have been implicated in plant responses to drought, salinity, pathogen attack, and phytohormone signals. Developmentally, genes encoding NF-Y transcription factors have known roles in root development, seed germination, and

photoperiod-dependent flowering (Petroni et al., 2013; Swain et al., 2017; Zanetti et al., 2017).

Transcription factors of the homeodomain superfamily share a strongly evolutionarily conserved domain consisting of a helix-turn-helix (HTH) structure, which comprises the DNA binding domain (Kappen, 2000). Originally discovered in *Drosophila* homeodomain proteins, also known as homeoboxes, have now been found to exist across phylogenetically divergent organisms and preferentially bind DNA sequences containing a TAAT core (Chan et al., 1998). Although the amino acids near the site of DNA binding are nearly 100% conserved, sequence divergences in adjacent regions allows for the sub-classification of homeodomain genes into approximately 20 subclasses with plants possessing members from at least seven of these subclasses. Some of these subclasses, such as the homeodomain-leucine zipper class I, II, III, and IV proteins, appear to be plant-specific (Chan et al., 1998; Roodbarkelari & Groot, 2017). Often upregulated in response to ABA or gibberellic acid (GA) signaling, plant homeodomain transcription factors play roles in mediating the signaling pathways of these phytohormones and act as positive regulators of jasmonic acid and ethylene production. Numerous homeobox transcription factors have been shown to be positive regulators of plant responses to drought, salinity, cold, high-light stress, oxidative stress, and metallotoxic stress (Gong et al., 2019). Some homeodomain genes are also important components of developmental transcriptional networks, with critical roles in embryo development and tissue differentiation. Acting frequently via their interaction with auxin biosynthetic and transport networks, homeodomain transcription factors are necessary for root development, cotyledon establishment, apical meristem maintenance,

vascularization, and abaxial/adaxial specification (Roodbarkelari & Groot, 2017).

Representing one of the largest transcription factor families in plants is the MYB family. In *Arabidopsis*, this class of transcriptional regulators encompasses approximately 9% of the entire set of known transcription factors. Present in all eukaryotes, this family is characterized by a variable number of conserved repeats near the N-terminus that function in DNA binding. These repeats, which have been found to vary from one to four in number, are each composed of a helix-turn-helix structure that extends into the DNA major groove through regularly-spaced tryptophan residues. The exact number and precise positioning of these repeats is used to classify plant MYB family members into distinct subfamilies (Li et al., 2015; Roy, 2016). Consistent with the generally expanded set of MYB transcription factors in plants compared with other eukaryotes, the different functional roles of this class of transcription factor are truly extensive with *MYB* genes having been found to play key roles in essentially all aspects of plant development and physiology (Du et al., 2012; Wang et al., 2021). Many closely-related transcription factors of the MYB class appear to have functional redundancy and recognize similar DNA motifs (Du et al., 2012; Kelemen et al., 2015). Extensive evaluation of MYB transcription factors in *Arabidopsis* and to a lesser extent soybean indicates that members of the MYB family are generally positive regulators of transcription with diverse roles in responses to biotic and abiotic stress, development, and epigenetics. Notably, regulatory networks governing the expression of anthocyanin biosynthesis genes appear to be particularly enriched in MYB family members, which often undergo extensive crosstalk with other regulators within these networks (Du et al., 2012; Li et al., 2015). Numerous examples of improvement of abiotic stress tolerance

and increased resistance to pathogens through the overexpression of MYB transcription factors are summarized in a review by Erpen et al. (2018).

Transcription factors belonging to the APETALA2/Ethylene Responsive Factor (AP2/ERF) family form a large group of plant-specific regulatory factors that is comprised of four main subfamilies (AP2, ERF, RAV, and DREB). Generally, transcription factors of the AP2/ERF class show tight regulation with expression only being induced in response to particular environmental stimuli, phytohormone signals, or during specific developmental stages (Mizoi et al., 2012; Xie et al., 2014). Structurally, all members of the AP2/ERF family share a somewhat-conserved DNA binding domain consisting of approximately 60 amino acids arranged into three antiparallel beta-sheets followed by a single alpha-helix. Differences in amino acid residues within the DNA binding domain and hence differences in preferred DNA binding motifs are present among the different subfamilies. Members of the ERF subfamily, which will be the focus of the research presented in chapter three of this dissertation, have been reported to bind *in vitro* to the sequence AGCCGCC. This element, known as a GCC-box, is often present in the promoter regions of genes that respond to pathogens, wounding, and ethylene (Licausi et al., 2013). A large number of studies conducted in C_3 photosynthesis model organisms and crop species report the improvement in tolerance to drought, salinity, cold, heat, flood, and an increase in disease resistance to fungal and bacterial pathogens upon overexpression of select ERF transcription factors (Debbarma et al., 2019; Xu et al., 2011). Broadly, the actions of ERFs appear to be frequently exerted through their roles in integrating phytohormone signaling pathways, particularly those of ethylene, jasmonic acid, and salicylic acid (Gutterson & Reuber, 2004; Xu et al., 2011).

Genome-wide expression profiling of ERF subfamily transcription factors in maize by Hao et al. showed high levels of expression of multiple ERFs in root tissue, pollen, and endosperm indicating that this class of transcription factor might play critical roles in plant development and reproduction in addition to functioning in stress signaling (Hao et al., 2020).

The Teosinte branched1/Cycloidea/Proliferating cell factor (TCP) class of transcription factors is a family of plant-specific regulatory proteins that shares a conserved N-terminal 60-amino acid DNA binding domain (the TCP domain), which also serves in protein-protein interactions and contains a nuclear localization sequence (Feng et al., 2018; Leng et al., 2019). The TCP domain exhibits a non-canonical basic helix–loop–helix motif and transcription factors of this family can be divided into two broad classes, I and II, depending on specific variations in amino acid residues within or near the helix-loop-helix motif. The TCP domain preferentially binds the core motif sequence TGGGC(C/T) and transcription factors of this family typically function as either homo- or heterodimers (Danisman et al., 2012). Early studies on the function of *TCP* genes elucidated their roles as critical components of the transcriptional regulation of the cell cycle and cellular proliferation. More recently, increasing evidence points to their key roles in integrating environmental signals and stress stimuli with phytohormone signaling networks. In many cases TCP transcription factors act downstream of a particular phytohormone signal to mediate the cell cycle, yet also act upstream of the same signaling pathway to activate or repress gene expression directly involved in the biosynthesis of that phytohormone (Nicolas & Cubas, 2016). Diverse environmental signals are now known to affect TCP functions in plants including drought, salt, osmotic,

high light, cold, and pathogenic stress. Consistent with their roles in both development and stress responses, multiple functional characterization studies in *Arabidopsis* have observed marked changes in development and stress tolerance in mutant and overexpressing lines (Feng et al., 2018). Regulation of TCP transcription factor expression appears to be complex. However, microRNAs appear to be a well-conserved mechanism to target TCP transcripts for repression. Evolutionary and comparative expression analysis of TCP transcription factors in angiosperms identified 60 microRNAs predicted to interact with TCP transcripts. In particular, miR319 was found to target 90 TCP transcripts across 18 species (Liu et al., 2019).

Approaches to study DNA-protein interactions

The common theme of the research presented in the following chapters of this dissertation is the investigation of protein-DNA interactions. As such, extensive use is made of the techniques chromatin immunoprecipitation sequencing (ChIP-seq) and yeast one-hybrid screening (Y1H). In recent years ChIP-seq has emerged as the gold standard for the genome-wide mapping of protein-DNA interactions *in vivo*. Through ChIP-sequencing genomic sites bound by transcription factors, transcriptional machinery, and epigenetic regulatory proteins can be identified at base-pair resolution. The basic procedure for ChIP-seq involves the treatment of cells or tissue samples with formaldehyde to crosslink DNA to its associated DNA-binding proteins followed by cell and nuclear lysis. The resulting freed chromatin is then sheared into smaller fragments of approximately a few hundred base pairs through sonication or digestion with nucleases.

Next, an antibody against the specific protein of interest or against an epitope tag fused to the relevant protein is used to capture these crosslinked protein-DNA complexes. This resulting immune complex is then immobilized often through the use of a solid resin that contains IgG-derived binding domains for the antibody used. Lastly, these captured immune complexes are eluted and treated with heat, high-salt concentration solutions, and proteinases to reverse the formaldehyde crosslinks and digest the antibody and its target protein. The resulting DNA is then used to create libraries for short-read sequencing platforms, typically Illumina (Furey, 2012; Park, 2009). As the DNA recovered from a successful ChIP experiment is expected to be enriched for binding sites of the protein of interest, the resulting sequence reads when mapped back to the reference genome of the organism should produce localized "peaks". These peak regions corresponding to overrepresented genomic loci can then be used to draw conclusions about the functional role, target genes, and preferred DNA-binding motifs of the protein of interest through various available bioinformatics tools (Ouma et al., 2015). While a powerful technique in its own right, a number of improvements have been made in recent years to the basic ChIP-seq experimental design. The technique of ChIP-exo improves upon the binding site resolution of traditional ChIP-seq through the incorporation of a step in which Lambda exonuclease is used to digest immunoprecipitated DNA in a strand specific 5' to 3' direction. This exonuclease digestion continues until it is blocked by a DNA-bound protein. Through the inclusion of this step, contaminating background DNA fragments are eliminated and much sharper peaks are obtained post-sequencing due to the digestion of DNA not directly bound by the immunoprecipitated protein (Mahony & Pugh, 2015). The relatively large number of cells and quantity of immunoprecipitated

DNA required to obtain useable sequencing data in standard ChIP-seq served as the motivation behind the development of nano-ChIP-seq. This approach utilizes a partial DNA amplification step in which scarce amounts of immunoprecipitated DNA are PCR amplified with primers, which yield products with ends compatible with Illumina adapters after a specific restriction enzyme digestion step and requires two to three orders of magnitude less cells than conventional ChIP-seq (Adli & Bernstein, 2011).

Due to the unique transcriptional regulation and chromatin architecture, which is often present at the individual-cell level, ChIP-seq data from only a particular cell type within an organ or tissue is often desirable (Taylor-Teeples et al., 2011). Several techniques are available for this and include laser-capture microdissection (LCM) and fluorescent-activated cell sorting (FACS). LCM involves the fixing of tissues followed by laser-assisted dissection under microscopic observation. Although no transformation or modification of the selected cell type is required, this technique is labor-intensive and can be difficult to perform when the cell type of interest is present as only a thin layer. FACS is achieved through the use of cell type-specific promoters or enhancers, which are used to drive the expression of a fluorescent protein in isolated protoplasts. These fluorescent cells can then be separated from non-fluorescing ones via fluorescence-activated cell sorting, a type of specialized flow cytometry. Although a rapid and robust technique, which has provided much useful cell-type specific transcriptional data in *Arabidopsis,* FACS is not compatible with species that lack an efficient protoplast generation protocol (Long, 2011).

More recently, the isolation of nuclei tagged in specific cell types (INTACT) approach has been used for cell-type specific gene expression and chromatin profiling

studies. INTACT is a method that biotinylates the nuclear envelope only in the specific cell type of interest then utilizes streptavidin affinity purification to selectively isolate these nuclei from the entire pool of nuclei present in a nuclear extract. INTACT is often a superior method for the study of individual cell types within a tissue as compared with FACS or LCM as it does not require the generation of protoplasts or involve the laborious and technically complex technique of laser microdissection. However, INTACT does require a stable transformation protocol for the organism of study and also requires that one has identified a promoter, which is highly cell-type specific, to drive the expression of the biotin acceptor protein (Deal & Henikoff, 2010). Although INTACT was implemented unsuccessfully in the *Kalanchoe* ChIP-seq project described in chapter two of this dissertation, there are multiple reports of it being used successfully in model plant species and for ChIP-seq studies (Wang et al., 2015).

As ChIP-seq and its variations involve determining the DNA sequences and genomic sites bound by a particular defined protein of interest, these techniques can be thought of as "protein-centered" approaches. Yeast one-hybrid (Y1H) screens, on the other hand, are "DNA-centered" approaches because they utilize a defined DNA sequence, such as a motif or promoter, which is then used to screen large collections of proteins for specific interactions with the defined DNA sequence (Reece-Hoyes & Walhout, 2012). The basic strategy of a Y1H screen involves transformation of a yeast strain with two constructs: a reporter construct consisting of a "bait" DNA sequence cloned upstream of a gene encoding an easily detectable reporter and an expression construct composed of a gene encoding a "prey" protein to be screened against the bait sequence that is fused to a yeast transcriptional activation domain. If a protein-DNA

interaction occurs between the prey/activation domain fusion and the bait DNA sequence the activation domain will be brought into close enough contact to induce expression of the chosen reporter. Often, yeast possessing auxotrophic phenotypes are chosen for Y1H screens with a rescue of cell growth on media lacking the specified nutrient serving as the reporter system (Reece-Hoyes et al., 2011; Reece-Hoyes & Walhout, 2012). Y1H screens have proven to be very versatile as the scale of a particular experiment can be adjusted to investigate the interaction of a bait DNA sequence with a small number of specific proteins or might include hundreds or even thousands of prey proteins each expressed as a member of a cDNA library. Additionally, the Y1H technique allows the investigation of protein-DNA interactions that might not be feasible through other methods, such as ChIP-seq, due to reasons such as antibody availability or low-abundance of a transcription factor or cell/tissue type (Reece-Hoyes & Walhout, 2012; Yang et al., 2019). Although a robust and well-established technique, Y1H screens do have limitations. False positives are possible, which can sometimes be the result of artificial binding sites generated by the junction between a bait DNA sequence and the yeast transformation vector. Weak interactions might be missed due to insufficient activation of the reporter. Lastly, Y1H assays in which large cDNA libraries of prey constructs are being screened require a substantial amount of time to perform. Higher throughput is possible, but this typically requires the implementation of expensive robotic systems (Reece-Hoyes & Walhout, 2012).

Kalanchoe species as models for studying CAM

Lastly, an introduction is owed to the model CAM species used for the research presented in this dissertation as the experiments in the following chapters make extensive use of diploid *Kalanchoe fedtschenkoi* and the very closely-related tetraploid *Kalanchoe laxiflora*. The genus *Kalanchoe* encompasses an estimated 125 succulent eudicot species native to Africa, Madagascar, and Asia. Most *Kalanchoe* species grow as small shrubs or small trees and, as members of the Crassulaceae, appear to all be capable of performing CAM. A clear gradient exists in the degree of CAM, which various species within the *Kalanchoe* genus perform, at least in their native environment (Hartwell et al., 2016; Kluge et al., 1991). The various species of *Kalanchoe* have been the subject of molecular phylogeny studies the results of which have established three major phylogenetic groups, which correlate with the degree to or the flexibility with which they perform CAM. Group I (Kitchingia) *Kalanchoe* mainly inhabit humid regions with high annual rainfall and primarily perform C_3 photosynthesis although some species have been observed to transition to CAM under drought stress. Group II (Bryophyllum) *Kalanchoe* species perform CAM in a flexible facultative manner and often display CAM along a developmental gradient with younger leaves performing C_3 photosynthesis and mature leaves performing CAM. Included in the Bryophyllum class are both *Kalanchoe fedtschenkoi* and *Kalanchoe laxiflora*. Species of this class primarily inhabit seasonally dry regions, with a significant amount of rainfall for part of the year. Group III *Kalanchoe* (Eukalanchoe) exhibit strong CAM, possess the greatest degree of tissue succulence within the *Kalanchoe* genus, and inhabit arid regions (Hartwell et al., 2016).

Kalanchoe species have been instrumental to our current understanding of CAM.

Data obtained in *Kalanchoe daigremontiana* provided the basis for a CAM system dynamics model, which describes the four phases of CAM detailed earlier (Borland & Yang, 2013; Owen & Griffiths, 2013). This species has also served as a model system for measuring the degree of CAM within individual leaves in response to light availability (Winter, 1987). *Kalanchoe fedtschenkoi*, in particular, has emerged as an ideal model for studying the functional genomics of CAM. This species is very easy to grow, grows quickly, can be clonally propagated from plantlets which develop from intact leaves or leaf explants (Borland et al., 2009; Hartwell at al., 2016), is amenable to stable transformation (Wang et al., 2019), and possesses a relatively small diploid genome which has been sequenced (Yang et al., 2017). *K. fedtschenkoi* has served as the model organism in multiple studies to elucidate the function of core CAM genes and their relationship with the circadian clock (Boxall et al., 2017; Dever et al., 2015; Hartwell et al., 2002), to identify genes involved in CAM stomatal dynamics (Moseley et al., 2019), and to characterize CAM mesophyll *vs*. epidermal proteomes (Abraham et al., 2020). Lastly, *K. fedtschenkoi* and *K. laxiflora* are useful CAM models in that C_3 photosynthesis and CAM are both exhibited in a single plant along a leaf developmental gradient. This feature allows for both of these types of photosynthesis to be studied without the need to apply an external complicating factor such as drought or salinity stress (Borland et al., 2009). Unlike *Arabidopsis,* which has now been an established model system for several decades, *Kalanchoe* has only recently become a molecular CAM model of serious interest (Hartwell et al., 2016). It has been both satisfying and encouraging to see the substantial growth in the amount of functional genomic resources available for these species over the time spent on this dissertation.

References

Abraham, P. E., Hurtado Castano, N., Cowan-Turner, D., Barnes, J., Poudel, S., Hettich, R., Flütsch, S., Santelia, D., & Borland, A. M. (2020). Peeling back the layers of crassulacean acid metabolism: Functional differentiation between *Kalanchoë fedtschenkoi* epidermis and mesophyll proteomes. *The Plant Journal*, *103*(2), 869–888. https://doi.org/10.1111/tpj.14757

Adli, M., & Bernstein, B. E. (2011). Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature Protocols*, *6*(10), 1656–1668. https://doi.org/10.1038/nprot.2011.402

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., & Westhoff, P. (2007). Evolution and Function of a *cis* -Regulatory Module for Mesophyll-Specific Gene Expression in the C4 Dicot *Flaveria trinervia*. *The Plant Cell*, *19*(11), 3391–3402. https://doi.org/10.1105/tpc.107.053322

Amin, A. B., Rathnayake, K. N., Yim, W. C., Garcia, T. M., Wone, B., Cushman, J. C., & Wone, B. W. M. (2019). Crassulacean Acid Metabolism Abiotic Stress-Responsive Transcription Factors: A Potential Genetic Engineering Approach for Improving Crop Tolerance to Abiotic Stress. *Frontiers in Plant Science*, *10*, 129. https://doi.org/10.3389/fpls.2019.00129

Bakrim, N., Brulfert, J., Vidal, J., & Chollet, R. (2001). Phosphoenolpyruvate Carboxylase Kinase Is Controlled by a Similar Signaling Cascade in CAM and C4 Plants. *Biochemical and Biophysical Research Communications*, *286*(5), 1158–1162. https://doi.org/10.1006/bbrc.2001.5527

Bohnert, H. J., & Cushman, J. C. (2000). The Ice Plant Cometh: Lessons in Abiotic Stress Tolerance. *Journal of Plant Growth Regulation*, *19*(3), 334–346. https://doi.org/10.1007/s003440000033

Borland, A. M., Griffiths, H., Hartwell, J., & Smith, J. A. C. (2009). Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *Journal of Experimental Botany*, *60*(10), 2879–2896. https://doi.org/10.1093/jxb/erp118

Borland, A. M., Hartwell, J., Weston, D. J., Schlauch, K. A., Tschaplinski, T. J., Tuskan, G. A., Yang, X., & Cushman, J. C. (2014). Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends in Plant Science*, *19*(5), 327–338. https://doi.org/10.1016/j.tplants.2014.01.006

Borland, A. M., & Yang, X. (2013). Informing the improvement and biodesign of crassulacean acid metabolism via system dynamics modelling. *New Phytologist*, *200*(4), 946–949. https://doi.org/10.1111/nph.12529

Boxall, S. F., Dever, L. V., Kneřová, J., Gould, P. D., & Hartwell, J. (2017). Phosphorylation of Phospho *enol* pyruvate Carboxylase Is Essential for Maximal and Sustained Dark CO_2 Fixation and Core Circadian Clock Operation in the Obligate Crassulacean Acid Metabolism Species *Kalanchoë fedtschenkoi*. *The Plant Cell*, *29*(10), 2519–2536. https://doi.org/10.1105/tpc.17.00301

Brilhaus, D., Bräutigam, A., Mettler-Altmann, T., Winter, K., & Weber, A. P. M. (2016). Reversible Burst of Transcriptional Changes during Induction of Crassulacean Acid Metabolism in *Talinum triangulare*. *Plant Physiology*, *170*(1), 102–122. https://doi.org/10.1104/pp.15.01076

Cai, Y.-H., & Huang, H. (2012). Advances in the study of protein–DNA interaction. *Amino Acids*, *43*(3), 1141–1146. https://doi.org/10.1007/s00726-012-1377-9

Ceusters, J., Borland, A. M., Godts, C., Londers, E., Croonenborghs, S., Van Goethem, D., & De Proft, M. P. (2011). Crassulacean acid metabolism under severe light limitation: A matter of plasticity in the shadows? *Journal of Experimental Botany*, *62*(1), 283–291. https://doi.org/10.1093/jxb/erq264

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, *4*(4), 287–291. https://doi.org/10.1038/nclimate2153

Chan, R. L., Gago, G. M., Palena, C. M., & Gonzalez, D. H. (1998). Homeoboxes in plant development. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1442*(1), 1–19. https://doi.org/10.1016/S0167-4781(98)00119-5

Chaudhry, S., & Sidhu, G. P. S. (2021). Climate change regulated abiotic stress mechanisms in plants: A comprehensive review. *Plant Cell Reports*. https://doi.org/10.1007/s00299-021-02759-5

Chen, C.-Y., Lin, P.-H., Chen, K.-H., & Cheng, Y.-S. (2020). Structural insights into Arabidopsis ethylene response factor 96 with an extended N-terminal binding to GCC box. *Plant Molecular Biology*, *104*(4–5), 483–498. https://doi.org/10.1007/s11103-020-01052-5

Chen, X., Wang, Y., Lv, B., Li, J., Luo, L., Lu, S., Zhang, X., Ma, H., & Ming, F. (2014). The NAC Family Transcription Factor OsNAP Confers Abiotic Stress Response Through the ABA Pathway. *Plant and Cell Physiology*, *55*(3), 604–619. https://doi.org/10.1093/pcp/pct204

Christin, P.-A., Arakaki, M., Osborne, C. P., Bräutigam, A., Sage, R. F., Hibberd, J. M., Kelly, S., Covshoff, S., Wong, G. K.-S., Hancock, L., & Edwards, E. J. (2014). Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *Journal of Experimental Botany*, *65*(13), 3609–3621. https://doi.org/10.1093/jxb/eru087

Cushman, J. C., & Bohnert, H. J. (1992). Salt stress alters A/T-rich DNA-binding factor interactions within the phosphoenolpyruvate carboxylase promoter from *Mesembryanthemum crystallinum*. *Plant Molecular Biology*, *20*(3), 411–424. https://doi.org/10.1007/BF00040601

Cushman, J. C., & Borland, A. M. (2002). Induction of Crassulacean acid metabolism by water limitation: CAM induction by water limitation. *Plant, Cell & Environment*, *25*(2), 295–310. https://doi.org/10.1046/j.0016-8025.2001.00760.x

Cushman, J. C., Tillett, R. L., Wood, J. A., Branco, J. M., & Schlauch, K. A. (2008). Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C3 photosynthesis and Crassulacean acid metabolism (CAM). *Journal of Experimental Botany*, *59*(7), 1875–1894. https://doi.org/10.1093/jxb/ern008

Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A. D., Muino, J. M., Cutri, L., Dornelas, M. C., Angenent, G. C., & Immink, R. G. H. (2012). Arabidopsis Class I and Class II TCP Transcription Factors Regulate Jasmonic Acid Metabolism and Leaf Development Antagonistically. *Plant Physiology*, *159*(4), 1511–1523. https://doi.org/10.1104/pp.112.200303

Davis, S. C., LeBauer, D. S., & Long, S. P. (2014). Light to liquid fuel: Theoretical and realized energy conversion efficiency of plants using Crassulacean Acid Metabolism (CAM) in arid conditions. *Journal of Experimental Botany*, *65*(13), 3471–3478. https://doi.org/10.1093/jxb/eru163

Deal, R. B., & Henikoff, S. (2010). A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. *Developmental Cell*, *18*(6), 1030–1040. https://doi.org/10.1016/j.devcel.2010.05.013

Debbarma, J., Sarki, Y. N., Saikia, B., Boruah, H. P. D., Singha, D. L., & Chikkaputtaiah, C. (2019). Ethylene Response Factor (ERF) Family Proteins in Abiotic Stresses and CRISPR–Cas9 Genome Editing of ERFs for Multiple Abiotic Stress Tolerance in Crop Plants: A Review. *Molecular Biotechnology*, *61*(2), 153–172. https://doi.org/10.1007/s12033-018-0144-x

Deng, H., Zhang, L.-S., Zhang, G.-Q., Zheng, B.-Q., Liu, Z.-J., & Wang, Y. (2016). Evolutionary history of PEPC genes in green plants: Implications for the evolution of CAM in orchids. *Molecular Phylogenetics and Evolution*, *94*, 559–564. https://doi.org/10.1016/j.ympev.2015.10.007

DePaoli, H. C., Borland, A. M., Tuskan, G. A., Cushman, J. C., & Yang, X. (2014). Synthetic biology as it relates to CAM photosynthesis: Challenges and opportunities. *Journal of Experimental Botany*, *65*(13), 3381–3393. https://doi.org/10.1093/jxb/eru038

Dever, L. V., Boxall, S. F., Kneřová, J., & Hartwell, J. (2015). Transgenic Perturbation of the Decarboxylation Phase of Crassulacean Acid Metabolism Alters Physiology and Metabolism But Has Only a Small Effect on Growth. *Plant Physiology*, *167*(1), 44–59. https://doi.org/10.1104/pp.114.251827

Dodd, A. N., Borland, A. M., Haslam, R. P., Griffiths, H., & Maxwell, K. (2002). Crassulacean acid metabolism: Plastic, fantastic. *Journal of Experimental Botany*, *53*(369), 569–580. https://doi.org/10.1093/jexbot/53.369.569

Du, H., Yang, S.-S., Liang, Z., Feng, B.-R., Liu, L., Huang, Y.-B., & Tang, Y.-X. (2012). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biology*, *12*(1), 106. https://doi.org/10.1186/1471-2229-12-106

Edwards, E. J. (2019). Evolutionary trajectories, accessibility and other metaphors: The case of C $_4$ and CAM photosynthesis. *New Phytologist*, *223*(4), 1742–1755. https://doi.org/10.1111/nph.15851

Erpen, L., Devi, H. S., Grosser, J. W., & Dutt, M. (2018). Potential use of the DREB/ERF, MYB, NAC and WRKY transcription factors to improve abiotic and biotic stress in transgenic plants. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *132*(1), 1–25. https://doi.org/10.1007/s11240-017-1320-6

Feng, Z.-J., Xu, S.-C., Liu, N., Zhang, G.-W., Hu, Q.-Z., & Gong, Y.-M. (2018). Soybean TCP transcription factors: Evolution, classification, protein interaction and stress and hormone responsiveness. *Plant Physiology and Biochemistry*, *127*, 129–142. https://doi.org/10.1016/j.plaphy.2018.03.020

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., & Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *PLANT BIOLOGY*, 6. https://doi.org/10.1016/j.bbagrm.2016.05.001

Furey, T. S. (2012). ChIP–seq and beyond: New and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics*, *13*(12), 840–852. https://doi.org/10.1038/nrg3306

Galli, M., Feng, F., & Gallavotti, A. (2020). Mapping Regulatory Determinants in Plants. *Frontiers in Genetics*, *11*, 591194. https://doi.org/10.3389/fgene.2020.591194

Gong, S., Ding, Y., Hu, S., Ding, L., Chen, Z., & Zhu, C. (2019). The role of HD-Zip class I transcription factors in plant response to abiotic stresses. *Physiologia Plantarum*, *167*(4), 516–525. https://doi.org/10.1111/ppl.12965

Gutterson, N., & Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Current Opinion in Plant Biology*, *7*(4), 465–471. https://doi.org/10.1016/j.pbi.2004.04.007

Hao, L., Shi, S., Guo, H., Li, M., Hu, P., Wei, Y., & Feng, Y. (2020). Genome-wide identification and expression profiles of ERF subfamily transcription factors in *Zea mays*. *PeerJ*, *8*, e9551. https://doi.org/10.7717/peerj.9551

Hartwell, J., Dever, L. V., & Boxall, S. F. (2016). Emerging model systems for functional genomics analysis of Crassulacean acid metabolism. *Current Opinion in Plant Biology*, *31*, 100–108. https://doi.org/10.1016/j.pbi.2016.03.019

Hartwell, J., Nimmo, G. A., Wilkins, M. B., Jenkins, G. I., & Nimmo, H. G. (2002). Probing the circadian control of phosphoenolpyruvate carboxylase kinase expression in Kalanchoë fedtschenkoi. *Functional Plant Biology*, *29*(6), 663. https://doi.org/10.1071/PP01208

Iizumi, T., Shiogama, H., Imada, Y., Hanasaki, N., Takikawa, H., & Nishimori, M. (2018). Crop production losses associated with anthropogenic climate change for 1981-2010 compared with preindustrial levels. *International Journal of Climatology*, *38*(14), 5405–5417. https://doi.org/10.1002/joc.5818

Janni, M., Gullì, M., Maestri, E., Marmiroli, M., Valliyodan, B., Nguyen, H. T., & Marmiroli, N. (2020). Molecular and genetic bases of heat stress responses in crop plants and breeding for increased resilience and productivity. *Journal of Experimental Botany*, *71*(13), 3780–3802. https://doi.org/10.1093/jxb/eraa034

Karki, S., Rizal, G., & Quick, W. P. (2013). Improvement of photosynthesis in rice (*Oryza sativa L.*) by inserting the C4 pathway. *Rice*, *6*(1), 28. https://doi.org/10.1186/1939-8433-6-28

Keeley, J. E., & Rundel, P. W. (2003). Evolution of CAM and C_4 carbon-concentrating mechanisms. International Journal of Plant Sciences, 164(3), 555–577. USGS Publications Warehouse. https://doi.org/10.1086/374192

Kelemen, Z., Sebastian, A., Xu, W., Grain, D., Salsac, F., Avon, A., Berger, N., Tran, J., Dubreucq, B., Lurin, C., Lepiniec, L., Contreras-Moreira, B., & Dubos, C. (2015). Analysis of the DNA-Binding Activities of the Arabidopsis R2R3-MYB Transcription Factor Family by One-Hybrid Experiments in Yeast. *PLOS ONE*, *10*(10), e0141044. https://doi.org/10.1371/journal.pone.0141044

Kluge, M., Brulfert, J., Ravelomanana, D., Lipp, J., & Ziegler, H. (1991). Crassulacean acid metabolism in Kalanchoe species collected in various climatic zones of Madagascar: A survey by 13C analysis. *Oecologia*, *88*(3), 407–414. https://doi.org/10.1007/BF00317586

Kore-eda, S., Cushman, M. A., Akselrod, I., Bufford, D., Fredrickson, M., Clark, E., & Cushman, J. C. (2004). Transcript profiling of salinity stress responses by large-scale expressed sequence tag analysis in *Mesembryanthemum crystallinum*. *Gene*, *341*, 83–92. https://doi.org/10.1016/j.gene.2004.06.037

Laloum, T., De Mita, S., Gamas, P., Baudin, M., & Niebel, A. (2013). CCAAT-box binding transcription factors in plants: Y so many? *Trends in Plant Science*, *18*(3), 157–166. https://doi.org/10.1016/j.tplants.2012.07.004

Leng, X., Wei, H., Xu, X., Ghuge, S. A., Jia, D., Liu, G., Wang, Y., & Yuan, Y. (2019). Genome-wide identification and transcript analysis of TCP transcription factors in grapevine. *BMC Genomics*, *20*(1), 786. https://doi.org/10.1186/s12864-019-6159-2

Li, C., Ng, C. K.-Y., & Fan, L.-M. (2015). MYB transcription factors, active players in abiotic stress signaling. *Environmental and Experimental Botany*, *114*, 80–91. https://doi.org/10.1016/j.envexpbot.2014.06.014

Licausi, F., Ohme-Takagi, M., & Perata, P. (2013). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. *New Phytologist*, *199*(3), 639–649. https://doi.org/10.1111/nph.12291

Lim, S. D., Lee, S., Choi, W.-G., Yim, W. C., & Cushman, J. C. (2019). Laying the Foundation for Crassulacean Acid Metabolism (CAM) Biodesign: Expression of the C4 Metabolism Cycle Genes of CAM in Arabidopsis. *Frontiers in Plant Science*, *10*, 101. https://doi.org/10.3389/fpls.2019.00101

Lim, S. D., Mayer, J. A., Yim, W. C., & Cushman, J. C. (2020). Plant tissue succulence engineering improves water-use efficiency, water-deficit stress attenuation and salinity tolerance in Arabidopsis. *The Plant Journal*, *103*(3), 1049–1072. https://doi.org/10.1111/tpj.14783

Liu, D., Hu, R., Zhang, J., Guo, H.-B., Cheng, H., Li, L., Borland, A. M., Qin, H., Chen, J.-G., Muchero, W., Tuskan, G. A., & Yang, X. (2021). Overexpression of an *Agave* Phosphoenolpyruvate Carboxylase Improves Plant Growth and Stress Tolerance. *Cells*, *10*(3), 582. https://doi.org/10.3390/cells10030582

Liu, M.-M., Wang, M.-M., Yang, J., Wen, J., Guo, P.-C., Wu, Y.-W., Ke, Y.-Z., Li, P.-F., Li, J.-N., & Du, H. (2019). Evolutionary and Comparative Expression Analyses of TCP Transcription Factor Gene Family in Land Plants. *International Journal of Molecular Sciences*, *20*(14), 3591. https://doi.org/10.3390/ijms20143591

Lobell, D. B., & Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, *2*(1), 014002. https://doi.org/10.1088/1748-9326/2/1/014002

Long, T. A. (2011). Many needles in a haystack: Cell-type specific abiotic stress responses. *Current Opinion in Plant Biology*, *14*(3), 325–331. https://doi.org/10.1016/j.pbi.2011.04.005

Luttge, U. (2002). CO2-concentrating: Consequences in crassulacean acid metabolism. *Journal of Experimental Botany*, *53*(378), 2131–2142. https://doi.org/10.1093/jxb/erf081

Mahony, S., & Pugh, B. F. (2015). Protein–DNA binding in high-resolution. *Critical Reviews in Biochemistry and Molecular Biology*, *50*(4), 269–283. https://doi.org/10.3109/10409238.2015.1051505

Males, J., & Griffiths, H. (2017). Stomatal Biology of CAM Plants. *Plant Physiology*, *174*(2), 550–560. https://doi.org/10.1104/pp.17.00114

Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., Hazen, S. P., Shen, R., Priest, H. D., Sullivan, C. M., Givan, S. A., Yanovsky, M., Hong, F., Kay, S. A., & Chory, J. (2008). Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific cis-Regulatory Modules. *PLoS Genetics*, *4*(2), e14. https://doi.org/10.1371/journal.pgen.0040014

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., Lyons, E., Wang, M.-L., Chen, J., Biggers, E., Zhang, J., Huang, L., Zhang, L., Miao, W., Zhang, J., Ye, Z., Miao, C., Lin, Z., Wang, H., … Yu, Q. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics*, *47*(12), 1435–1442. https://doi.org/10.1038/ng.3435

Mizoi, J., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1819*(2), 86–96. https://doi.org/10.1016/j.bbagrm.2011.08.004

Moseley, R. C., Tuskan, G. A., & Yang, X. (2019). Comparative Genomics Analysis Provides New Insight Into Molecular Basis of Stomatal Movement in *Kalanchoë fedtschenkoi*. *Frontiers in Plant Science*, *10*, 292. https://doi.org/10.3389/fpls.2019.00292

Naumann, G., Alfieri, L., Wyser, K., Mentaschi, L., Betts, R. A., Carrao, H., Spinoni, J., Vogt, J., & Feyen, L. (2018). Global Changes in Drought Conditions Under Different Levels of Warming. *Geophysical Research Letters*, *45*(7), 3285–3296. https://doi.org/10.1002/2017GL076521

Nicolas, M., & Cubas, P. (2016). TCP factors: New kids on the signaling block. *Current Opinion in Plant Biology*, *33*, 33–41. https://doi.org/10.1016/j.pbi.2016.05.006

Niechayev, N. A., Pereira, P. N., & Cushman, J. C. (2019). Understanding trait diversity associated with crassulacean acid metabolism (CAM). *Current Opinion in Plant Biology*, *49*, 74–85. https://doi.org/10.1016/j.pbi.2019.06.004

Nimmo, H. G. (2000). The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends in Plant Science*, *5*(2), 75–80. https://doi.org/10.1016/S1360-1385(99)01543-5

Osmond, C. B. (1978). Crassulacean acid metabolism: a curiosity in context. *Annual Review of Plant Physiology*, *29*(1), 379-414.

Ouma, W. Z., Mejia-Guerra, M. K., Yilmaz, A., Pareja-Tobes, P., Li, W., Doseff, A. I., & Grotewold, E. (2015). Important biological information uncovered in previously unaligned reads from chromatin immunoprecipitation experiments (ChIP-Seq). *Scientific Reports*, *5*(1), 8635. https://doi.org/10.1038/srep08635

Owen, N. A., & Griffiths, H. (2013). A system dynamics model integrating physiology and biochemical regulation predicts extent of crassulacean acid metabolism (CAM) phases. *New Phytologist*, *200*(4), 1116–1131. https://doi.org/10.1111/nph.12461

Pareek, A. (Ed.). (2010). *Abiotic stress adaptation in plants: Physiological, molecular, and genomic foundation*. Springer.

Park, P. J. (2009). ChIP–seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*, *10*(10), 669–680. https://doi.org/10.1038/nrg2641

Petroni, K., Kumimoto, R. W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., Holt, B. F., & Mantovani, R. (2013). The Promiscuous Life of Plant NUCLEAR FACTOR Y Transcription Factors. *The Plant Cell*, *24*(12), 4777–4792. https://doi.org/10.1105/tpc.112.105734

Reece-Hoyes, J. S., Diallo, A., Lajoie, B., Kent, A., Shrestha, S., Kadreppa, S., Pesyna, C., Dekker, J., Myers, C. L., & Walhout, A. J. M. (2011). Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nature Methods*, *8*(12), 1059–1064. https://doi.org/10.1038/nmeth.1748

Reece-Hoyes, J. S., & Marian Walhout, A. J. (2012). Yeast one-hybrid assays: A historical and technical perspective. *Methods*, *57*(4), 441–447. https://doi.org/10.1016/j.ymeth.2012.07.027

Roodbarkelari, F., & Groot, E. P. (2017). Regulatory function of homeodomain-leucine zipper (HD - ZIP) family proteins during embryogenesis. *New Phytologist*, *213*(1), 95–104. https://doi.org/10.1111/nph.14132

Roy, S. (2016). Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant Signaling & Behavior*, *11*(1), e1117723. https://doi.org/10.1080/15592324.2015.1117723

Schaeffer, H. J., Forsthoefel, N. R., & Cushman, J. C. (1995). Identification of enhancer and silencer regions involved in salt-responsive expression of Crassulacean acid metabolism (CAM) genes in the facultative halophyte *Mesembryanthemum crystallinum*. *Plant Molecular Biology*, *28*(2), 205–218. https://doi.org/10.1007/BF00020241

Schiller, K., & Bräutigam, A. (2021). Engineering of Crassulacean Acid Metabolism. *Annual Review of Plant Biology*, *72*(1), 77–103. https://doi.org/10.1146/annurev-arplant-071720-104814

Schlüter, U., & Weber, A. P. M. (2016). The Road to C_4 Photosynthesis: Evolution of a Complex Trait via Intermediary States. *Plant and Cell Physiology*, *57*(5), 881–889. https://doi.org/10.1093/pcp/pcw009

Silvera, K., Winter, K., Rodriguez, B. L., Albion, R. L., & Cushman, J. C. (2014). Multiple isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): Implications for the evolution of crassulacean acid metabolism. *Journal of Experimental Botany*, *65*(13), 3623–3636. https://doi.org/10.1093/jxb/eru234

Swain, S., Myers, Z. A., Siriwardana, C. L., & Holt, B. F. (2017). The multifaceted roles of NUCLEAR FACTOR-Y in *Arabidopsis thaliana* development and stress responses. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1860*(5), 636–644. https://doi.org/10.1016/j.bbagrm.2016.10.012

Taybi, T., Nimmo, H. G., & Borland, A. M. (2004). Expression of Phospho*enol*pyruvate Carboxylase and Phospho *enol* pyruvate Carboxylase Kinase Genes. Implications for Genotypic Capacity and Phenotypic Plasticity in the Expression of Crassulacean Acid Metabolism. *Plant Physiology*, *135*(1), 587–598. https://doi.org/10.1104/pp.103.036962

Taybi, T., Patil, S., Chollet, R., & Cushman, J. C. (2000). A Minimal Serine/Threonine Protein Kinase Circadianly Regulates Phospho*enol*pyruvate Carboxylase Activity in Crassulacean Acid Metabolism-Induced Leaves of the Common Ice Plant. *Plant Physiology*, *123*(4), 1471–1482. https://doi.org/10.1104/pp.123.4.1471

Taylor-Teeples, M., Ron, M., & Brady, S. M. (2011). Novel biological insights revealed from cell type-specific expression profiling. *Current Opinion in Plant Biology*, *14*(5), 601–607. https://doi.org/10.1016/j.pbi.2011.05.007

Theng, V., Agarie, S., & Nose, A. (2007). Regulatory Properties of Phospho*enol*pyruvate Carboxylase in Crassulacean Acid Metabolism Plants: Diurnal Changes in Phosphorylation State and Regulation of Gene Expression. *Plant Production Science*, *10*(2), 171–181. https://doi.org/10.1626/pps.10.171

Wai, C. M., VanBuren, R., Zhang, J., Huang, L., Miao, W., Edger, P. P., Yim, W. C., Priest, H. D., Meyers, B. C., Mockler, T., Smith, J. A. C., Cushman, J. C., & Ming, R. (2017). Temporal and spatial transcriptomic and micro RNA dynamics of CAM photosynthesis in pineapple. *The Plant Journal*, *92*(1), 19–30. https://doi.org/10.1111/tpj.13630

Wai, C. M., Weise, S. E., Ozersky, P., Mockler, T. C., Michael, T. P., & VanBuren, R. (2019). Time of day and network reprogramming during drought induced CAM photosynthesis in Sedum album. *PLOS Genetics*, *15*(6), e1008209. https://doi.org/10.1371/journal.pgen.1008209

Wang D., Deal R.B. (2015) Epigenome Profiling of Specific Plant Cell Types Using a Streamlined INTACT Protocol and ChIP-seq. In: Alonso J., Stepanova A. (eds) Plant Functional Genomics. Methods in Molecular Biology, vol 1284. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-2444-8_1

Wang, J., Vanga, S., Saxena, R., Orsat, V., & Raghavan, V. (2018). Effect of Climate Change on the Yield of Cereal Crops: A Review. *Climate*, *6*(2), 41. https://doi.org/10.3390/cli6020041

Wang, N., Zhong, X., Cong, Y., Wang, T., Yang, S., Li, Y., & Gai, J. (2016). Genome-wide Analysis of Phosphoenolpyruvate Carboxylase Gene Family and Their Response to Abiotic Stresses in Soybean. *Scientific Reports*, *6*(1), 38448. https://doi.org/10.1038/srep38448

Wang, X., Chen, X., Cheng, Q., Zhu, K., Yang, X., & Cheng, Z. (2019). Agrobacterium–mediated Transformation of *Kalanchoe laxiflora*. *Horticultural Plant Journal*, *5*(5), 221–228. https://doi.org/10.1016/j.hpj.2019.07.001

Wang, X., Niu, Y., & Zheng, Y. (2021). Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. *International Journal of Molecular Sciences*, *22*(11), 6125. https://doi.org/10.3390/ijms22116125

Winter, K. (1987). *Gradient in the degree of Crassulacean acid metabolism within leaves of Kalanchoe daigremontiana*. *Planta, 172*.

Xie, X., Shen, S., Yin, X., Xu, Q., Sun, C., Grierson, D., Ferguson, I., & Chen, K. (2014). Isolation, classification and transcription profiles of the AP2/ERF transcription factor superfamily in citrus. *Molecular Biology Reports*, *41*(7), 4261–4271. https://doi.org/10.1007/s11033-014-3297-0

Xu, Z.-S., Chen, M., Li, L.-C., & Ma, Y.-Z. (2011). Functions and Application of the AP2/ERF Transcription Factor Family in Crop ImprovementF: Functions of the AP2/EREBP Family. *Journal of Integrative Plant Biology*, *53*(7), 570–585. https://doi.org/10.1111/j.1744-7909.2011.01062.x

Yang, G., Chao, D., Ming, Z., & Xia, J. (2019). A Simple Method to Detect the Inhibition of Transcription Factor-DNA Binding Due to Protein–Protein Interactions In Vivo. *Genes*, *10*(9), 684. https://doi.org/10.3390/genes10090684

Yang, X., Cushman, J. C., Borland, A. M., Edwards, E. J., Wullschleger, S. D., Tuskan, G. A., Owen, N. A., Griffiths, H., Smith, J. A. C., De Paoli, H. C., Weston, D. J., Cottingham, R., Hartwell, J., Davis, S. C., Silvera, K., Ming, R., Schlauch, K., Abraham, P., Stewart, J. R., … Holtum, J. A. M. (2015). A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist*, *207*(3), 491–504. https://doi.org/10.1111/nph.13393

Yang, X., Hu, R., Yin, H., Jenkins, J., Shu, S., Tang, H., Liu, D., Weighill, D. A., Cheol Yim, W., Ha, J., Heyduk, K., Goodstein, D. M., Guo, H.-B., Moseley, R. C., Fitzek, E., Jawdy, S., Zhang, Z., Xie, M., Hartwell, J., … Tuskan, G. A. (2017). The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications*, *8*(1), 1899. https://doi.org/10.1038/s41467-017-01491-7

Zanetti, M. E., Rípodas, C., & Niebel, A. (2017). Plant NF-Y transcription factors: Key players in plant-microbe interactions, root development and adaptation to stress. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1860*(5), 645–654. https://doi.org/10.1016/j.bbagrm.2016.11.007

Zhang, J., Hu, R., Sreedasyam, A., Garcia, T. M., Lipzen, A., Wang, M., Yerramsetty, P., Liu, D., Ng, V., Schmutz, J., Cushman, J. C., Borland, A. M., Pasha, A., Provart, N. J., Chen, J.-G., Muchero, W., Tuskan, G. A., & Yang, X. (2020). Light-responsive expression atlas reveals the effects of light quality and intensity in *Kalanchoë fedtschenkoi*, a plant with crassulacean acid metabolism. *GigaScience*, *9*(3), giaa018. https://doi.org/10.1093/gigascience/giaa018

Chapter 2. Genome-wide binding site profiling of *KfNF-YB3, KfHD-like,* and *KfMYB59,* three transcription factors from *Kalanchoe fedtschenkoi* with putative roles in the regulation of crassulacean acid metabolism

Authors:

Travis M. Garcia[1], Bernard W.M. Wone[2], Jungmin Ha[3], Sung Don Lim[4], Won C. Yim[1], John C. Cushman[1]

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557-0330 USA

[2]University of South Dakota, Department of Biology, Vermillion, SD 57069 USA

[3]Department of Plant Science, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826 South Korea

[4]Department of Plant Life & Resource Science, Sangji University, 83 Sangjidae-gil, Wonju-si, Gangwon-do, 26339, South Korea

Abstract

Increasing crop production is vital to meet future food and renewable energy demands. However, much of the world is too arid for sustainable agricultural productivity. Introducing crassulacean acid metabolism (CAM) into these crops might confer the improved water-use efficiency required for growth in arid lands. CAM relies, in part, upon rhythmic circadian clock-dependent mRNA regulation. However, the transcriptional regulation of CAM remains largely uncharacterized. Using *Kalanchoe fedtschenkoi*, in which CAM develops along a leaf developmental gradient, candidate transcription factors with possible CAM-related functions were identified. The mRNA transcript abundance levels of these transcription factors increased during the transition from C_3 photosynthesis to CAM. To attempt to further elucidate the transcriptional regulation of CAM and identify important *cis*-regulatory genomic targets, ChIP-seq analysis was performed on three of these transcription factors. After confirming the nuclear localization of the expressed proteins, nuclear isolation and chromatin immunoprecipitation was performed using mature leaf tissue from transgenic lines of *K. fedtschenkoi* expressing epitope tagged fusions of *K. fedtschenkoi* NF-YB3, Homeodomain-like Superfamily Protein, and MYB59. Ultimately, although high-quality DNA was recovered from immunoprecipitated samples and subsequently sequenced at sufficient depth, no substantial enrichment of target loci could be identified over background signal. In contrast to well-studied model plant species such as *Arabidopsis* for which nuclear isolation, chromatin immunoprecipitation and other upstream protocols relevant to ChIP-seq experiments are well-established, *Kalanchoe* presents distinct challenges. These challenges of working with this species are discussed here in addition

to presenting the resulting ChIP-seq data from which, unfortunately, no meaningful biological insights can be made. However, candidate target binding sites for these transcription factors were identified for future investigations.

Introduction

A variety of emerging technologies exist to convert plant tissue to liquid fuels. Biomass from bioenergy crops such as maize, sugarcane, and poplar can be converted to such fuels as ethanol, alkanes, and terpenes. Demand will continually increase in the coming years for these forms of biofuels (Somerville et al. 2010). However, to meet future energy needs, biomass productivity must be expanded as current energy crop production is expected to fall short of demand. Unfortunately, a significant obstacle exists in expanding the production of these crops because large regions of the world are too arid to support their growth. Approximately 18% of the world's surface is drought prone, with 25% of the United States suffering from this problem. Almost all of the western United States is too arid to produce any significant biomass (Somerville et al. 2010; Borland et al. 2009; Boyer 1982).

Recent climate data predicting the incidence of drought conditions will only rise in the future points to the importance of bioenergy crops which can tolerate water-limited environments, heat, and even prolonged periods of drought. A feasible way to engineer important bioenergy crops for growth in arid or semi-arid lands might be through the introduction of the metabolic pathway crassulacean acid metabolism (CAM) (Borland et al. 2014). CAM is a common plant adaptation that has arisen through convergent evolution in response to low water availability and fluctuating CO_2 over the Earth's

history and is now present in 36 plant families and 6% of all higher plants. CAM plants are highly efficient with respect to their carbon gain and water-use efficiency (WUE). In a plant undergoing crassulacean acid metabolism, CO_2 is taken up at night through open stomata. This CO_2 is converted to HCO_3- and then added to phospho*enol*pyruvate (PEP) in a reaction which forms 4C organic acids, primarily malate. Malate is then stored in the plant's vacuoles. As carbon uptake occurs mainly at night, while stomata typically remain closed during the day, CAM species are 3 to 6 fold more water-use efficient than plants which perform C_4 and C_3 photosynthesis, respectively. During the day, the nocturnal malate stores are released from the vacuoles and decarboxylated. This carbon may then be utilized for such pathways as gluconeogenesis and carbohydrate formation and storage (Cushman et al. 2008; Kellog 2013).

Patterns of gene expression have been observed to diverge significantly in plant tissue performing CAM as compared to C_3 photosynthesis. Previous research compared gene expression changes between CAM and C_3 metabolism in the facultative CAM species *Mesembryanthemum crystallinum* (common ice plant). Under unstressed conditions this species carries out C_3 photosynthesis. However, upon drought or salinity stress CAM is induced. From large-scale mRNA abundance comparisons in ice plant numerous genes are known to either increase or decrease in their expression upon CAM induction relative to the C_3 photosynthesis state. Many genes displayed obvious expression changes or circadian-phase expression shifts between these two types of metabolism. One such subset of mRNAs from ice plant whose temporal abundance in CAM shows departure from their C_3 patterns are mRNAs for transcription factors whose induction appears to be circadian controlled (Cushman et al. 2008). Similarly, a CAM *vs*.

C_3 comparative transcriptomics study was undertaken in the facultative CAM tropical dicot *Talinum triangulare,* a species which transitions from C_3 photosynthesis to weak CAM in response to drought stress. Strikingly, 40% of the entire 1,449 annotated transcription factors in this species exhibited clear differential expression between a well-watered and drought-induced CAM state (Brilhaus et al., 2016).

The results of comparative transcript profiling studies such as those described above strongly suggest a role for circadian clock-controlled transcription factors in the regulation of CAM (Zhang et al., 2020). However, the specific *trans*-acting factors underlying the transcriptional networks of CAM and their target genomic loci remain largely unknown. As an understanding of CAM temporal regulatory networks would be vital in the ultimate engineering of CAM into important crop species, this research attempts to advance understanding of the transcriptional basis of crassulacean acid metabolism through genome-wide binding site analysis of the *Kalanchoe fedtschenkoi* transcription factors KfNF-YB3, KfHD-like, and KfMYB59. Previously, we performed C_3 photosynthesis *vs*. CAM differential RNA-seq analysis in the developmental CAM models *Kalanchoe laxiflora* and *Kalanchoe fedtschenkoi,* species in which C_3 photosynthesis is performed in immature leaves while CAM is performed in mature leaves (Yang et al., 2017). Among the differentially expressed transcription factors between C_3 and CAM, KfNF-YB3, KfHD-like, and KfMYB59 were selected for chromatin immunoprecipitation sequencing (ChIP-seq) analysis on the basis of marked transcript induction in CAM and clear circadian oscillations in mRNA abundance.

Previous literature findings support the potential roles for these transcription factors in abiotic stress-responsive pathways. Overexpression of *NF-YB3* from the

conifer species *Picea wilsonii* in *Arabidopsis* resulted in enhanced tolerance to drought, salinity, and osmotic stress along with an improvement in seed germination under these same conditions. qRT-PCR analysis of transgenic lines indicated upregulation of the salinity responsive gene *SOS3* and the drought-responsive gene *CDPK1*. Additionally, overexpressing lines displayed increased induction of multiple marker genes of the CBF pathway in response to drought (Zhang et al., 2015). A more comprehensive whole-genome RNA-seq analysis of *Arabidopsis* transcriptome changes in response to *NF-YB3* overexpression by Sato et al. discovered widespread upregulation of heat stress responsive genes. ChIP-RT-qPCR of several of these upregulated genes revealed interactions of NF-YB3 with promoter regions of heat-responsive genes containing the canonical NF-Y trimer CCAAT DNA binding motif. Phenotypic characterization of *NF-YB3* knockout mutants in this study revealed a stress-hypersensitive phenotype (Sato et al., 2019).

A sizeable body of literature now supports the role of homeodomain transcription factors in plant adaptations to abiotic stress. Several studies in *Arabidopsis* have linked multiple members of this transcription factor class to ABA-mediated physiological responses, including stomatal closure, in response to water-deficit and salinity stress. Research conducted in sunflower established a role for homeodomain transcription factors in the positive regulation of jasmonic acid and ethylene production in response to abiotic stress and wounding (Gong et al., 2019). At least some of the stress-adaptive responses mediated through homeodomain transcription factors have their basis in osmotic adjustment and enhanced ROS detoxification (Chew et al., 2013).

MYB59 has previously been reported to be an important component of the jasmonic acid gene regulatory network in which it is known to act as a positive regulator of plant defense responses to necrotrophic pathogens (Hickman et al., 2017). Additional studies on this transcription factor has revealed its role in regulation of the progression of the cell cycle, particularly in root cells, in response to Ca^{2+}-mediated stress signals and heavy metal toxicity (Fasani et al., 2019; Mu et al., 2009). Lastly, a report by Quan et al. established a link between *MYB59* and drought responses in rice. This study identified *OsMYB59* as one of the downstream effectors upregulated in response to the heterologous overexpression of tomato *TSRF1*, a stress-induced ERF transcription factor. Analysis of the *OsMYB59* promoter revealed the presence of ABA and drought-responsive elements (Quan et al., 2010).

In the course of our ChIP-seq-based analysis of *KfNF-YB3*, *KfHD-like*, and *KfMYB59*, we attempted to integrate a mesophyll cell-specific nuclei enrichment step upstream of chromatin immunoprecipitation. As CAM is a photosynthetic pathway restricted to mesophyll cells, we reasoned that a homogenous sample of nuclei from only this particular cell type would more precisely reflect the *in vivo* transcription factor-DNA interactions relevant to the regulation of CAM. To this end we employed isolation of nuclei tagged in specific cell types (INTACT) method. This approach utilizes a cell type-specific promoter to drive the expression of a nuclear envelope-localized GFP-tagged biotin acceptor peptide (NTF), which is introduced through stable transformation. The co-transformation of an additional transgene (BirA), which acts as an NTF-specific biotin ligase, results in biotinylated nuclei in the cell type of interest. Affinity purification with streptavidin can then be used to isolate these nuclei preferentially from those of other cell

types (Deal & Henikoff, 2010). Although *K. fedtschenkoi* mesophyll nuclei were apparently successfully tagged using this approach in which the *Kalanchoe Ppc* promoter was used to drive the expression of NTF, efforts to enrich these nuclei from whole-leaf nuclear extracts failed. Moreover, the isolation of nuclei generally from this species required substantial modification of protocols, which are suitable for non-succulent species. In addition to these challenges, the final ChIP-sequencing data appeared to be largely background noise. Bioinformatics analysis failed to confidently identify target loci or predict unique high-confidence DNA binding motifs for any of the three transcription factors analyzed. Possible reasons for these failures will be explored in the discussion section of this chapter along with suggestions for alternative approaches to investigate the transcriptional roles of *KfNF-YB3*, *KfHD-like*, and *KfMYB59*.

Materials and Methods

C_3 vs. CAM comparative RNA-seq and choice of transcription factors for ChIP-seq

Comparative transcriptome analysis between leaf pair 1 (C_3) and leaf pair 6 (CAM) in *Kalanchoe fedtschenkoi* was undertaken as described by Yang et al. (2017). Corresponding data for *Kalanchoe laxiflora* was obtained through Phytozome (*Kalanchoe laxiflora* v1.1, DOE-JGI, http://phytozome.jgi.doe.gov/). On the basis of a $\log_2$ fold-induction of greater than 2 between C_3 photosynthesis and CAM, a mean expression level of over 100 FPKM when transcript abundance in C_3 photosynthesis and CAM were averaged, and apparent circadian rhythmicity *KfNF-YB3* (Kaladp0059s0310.1) (Kalax.0351s0020.1 in *K. laxiflora*), *KfHomeodomain-like* (Kaladp0496s0018.1) (Kalax.0093s0011.1 in *K. laxiflora*), and *KfMYB59*

(Kaladp0033s0028.1) (Kalax.0291s0010.1 in *K.laxiflora*) were selected for ChIP-sequencing studies (Figures 1 and 2; Supplemental Table 2).

Plant material, maintenance, and growth conditions

Kalanchoe fedtschenkoi diploid M2 accession plants provided by Oak Ridge National Laboratory, Oak Ridge, TN were sterilized by washing for 30 s in 70% v/v ethanol followed by a 10-minute wash in 10% v/v NaOCl containing 1% v/v Tween 20. Plants were then washed five times in sterile water and transferred to 1X Murashige and Skoog media with Gamborg vitamins, 30 g/L sucrose, 0.7% Plant TC agar, pH 5.8 (MS30 media). All plants used for the experiments described in this chapter unless otherwise stated were maintained on this media in a growth chamber under a 16 h light / 8 h dark schedule at a constant 20 °C with a photon flux density of approximately 280 μmol m^{-2} s^{-1}.

Design and construction of Kalanchoe expression constructs

Total RNA was isolated from 100 mg frozen ground leaf tissue from tissue culture-grown *Kalanchoe fedtschenkoi* plants using the Qiagen RNeasy Plant Mini Kit following the manufacturer's instructions with the addition of 9 mg polyethylene glycol 20,000 (PEG-20,000) to the 450 μL of RLT buffer used for extraction. On-column DNA digestion was performed according to RNeasy kit instructions with the Qiagen RNase-Free DNase Set. Final RNA elution was performed with RNase-free water. RNA purity and approximate quantity was assessed with a Thermo Fisher Scientific NanoDrop 2000c spectrophotometer and precise quantity assessed using Thermo Fisher Scientific Quant-iT

RiboGreen fluorescence. RNA integrity was evaluated by electrophoresis on a 1% w/v agarose gel using 250 ng RNA.

cDNA synthesis was carried out on 500 ng isolated RNA using the Thermo Fisher Scientific SuperScript VILO Master Mix according to manufacturer's instructions. The full-length coding sequences (CDS) of *KfNF-YB3*, *KfHD-like*, and *KfMYB59* were amplified from the resulting cDNA using New England Biolabs High Fidelity Phusion DNA polymerase both with and without stop codons. Primer sequences are provided in Supplemental Table 1. CDS and amino acid sequences for *KfNF-YB3, KfHD-like*, and *KfMYB59* are provided in Supplemental Figures 1, 2, and 3. The resulting amplified PCR products were resolved by electrophoresis on a 1% w/v agarose gel and purified with the Qiagen QIAquick Gel Extraction Kit following the manufacturer's recommendations although nuclease-free water was used for final elution of the PCR products.

Purified transcription factor cDNAs were then used to set up cloning reactions with the Thermo Fisher Scientific pENTR D-TOPO directional cloning kit according to manufacturer's instructions with the exception that the incubation time of the ligation reaction was increased to 4 h to improve efficiency. Successful entry clones were confirmed by Sanger sequencing and used to generate plant expression constructs using the Thermo Fisher Scientific Gateway LR Clonase II Enzyme Mix according to manufacturer's instructions. CaMV 35S promoter-driven expression constructs were made using the improved Gateway binary vector plasmid series (ImpGWBs) (Nakagawa et al., 2007). N-terminus fusions to sGFP were created by cloning *KfNF-YB3, KfHD-like*, and *KfMYB59* into ImpGWB406. C-terminus fusions to sGFP were constructed with

ImpGWB405. N-terminus fusions to a 3xHA tag were generated by cloning *KfNF-YB3*, *KfHD-like*, and *KfMYB59* into ImpGWB515 while C-terminus 3xHA fusions were made with ImpGWB514. Transcription factor coding sequences amplified without stop codons were utilized for construction of the C-terminus sGFP and 3xHA constructs.

For construction of *Kl*p*Ppc*-INTACT, total genomic DNA from *Kalanchoe laxiflora* accession 1982-6028 grown under the above described conditions was used as a template to amplify a 2.8 kb section of the proximal *Ppc1* promoter (Supplemental Figure 4). The Thermo Fisher Scientific pENTR D-TOPO directional cloning kit was then used to generate an entry clone according to product instructions. Cloning of the *KlPpc1* promoter into pK7WG-INTACT upstream of *NTF* was then achieved with the Thermo Fisher Scientific Gateway LR Clonase II Enzyme Mix according to product instructions. All final plant expression clones were sequence-verified with Sanger sequencing. Diagrams of all plant expression constructs are provided in Supplemental Figure 5.

Transient expression subcellular localization assays in Kalanchoe fedtschenkoi

Agrobacterium tumefaciens GV3101 cells containing *KfNF-YB3, KfHD-like*, and *KfMYB59* sGFP fusion constructs and those carrying empty ImpGWB405/406 vectors were grown for 2 d at 28 °C in YEP containing 25 mg/L gentamycin, 50 mg/L rifampicin, and 100 mg/L spectinomycin. Cells were pelleted by centrifugation at 3,200 *x g* for 10 minutes at 4 °C and re-suspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES, 150 µM acetosyringone, pH 5.6 in sterile water). Re-suspended cells were adjusted to an OD_{600} of 0.5 and incubated at 25 °C for 3 h. Mature *Kalanchoe fedtschenkoi* leaves from tissue culture-grown plants were detached and gently punctured

several times on their abaxial leaf surfaces with an 18 gauge needle. Care was taken not to puncture the leaves all the way through. *Agrobacterium* suspensions described above were then used to syringe-infiltrate the punctured locations on the detached leaves. Several sites were infiltrated on each leaf. Infiltrated leaves were then placed on wet filter paper in petri plates and returned to the growth chamber for four days. After this time leaves were thinly sectioned by hand with a razor blade and examined for sGFP fluorescence using a Keyence BZ-X710 microscope. Fluoroshield with DAPI (Sigma Aldrich product no. F6057) was used for microscope slide mounting and visualization of nuclei. For visualization of GFP, a Keyence model OP-87763 filter was used with an excitation wavelength of 470/40 nm and an emission wavelength of 525/50 nm. For DAPI imaging, a Keyence model OP-87762 filter was used with an excitation wavelength of 360/40 nm and an emission wavelength of 460/50 nm.

Stable transformation of Kalanchoe fedtschenkoi

For stable transformation of *Kalanchoe fedtschenkoi* with *Kl*p*Ppc*-INTACT *Agrobacterium* harboring this construct was grown for 2 d at 28 °C in YEP containing 25 mg/L gentamycin, 50 mg/L rifampicin, and 100 mg/L spectinomycin. Acetosyringone was then added to a final concentration of 100 µM and cells were incubated for an additional 2 h at 28 °C. The cells were then pelleted by centrifugation at 3,200 *x g* for 10 minutes at 4 °C and re-suspended in liquid MS30 re-suspension media (MS30 with 100 µM acetosyringone, 1 mg/L thidazuron, 0.2 mg/L indole acetic acid, pH 5.8) at an OD_{600} of 0.1. Small, approximately 5 mm leaves taken from young *K. fedtschenkoi* plantlets

originally started from leaf explants were sliced in half and incubated in the re-suspended *Agrobacterium* culture for 1 h with gentle agitation. Leaves were then blotted dry on filter paper to remove excess *Agrobacterium* and placed on co-cultivation media (MS30 with 100 µM acetosyringone, 1 mg/L thidazuron, 0.2 mg/L indole acetic acid, pH 5.8) for 4 d. After co-cultivation, explants were then transferred to regeneration media (MS30 with 1 mg/L thidazuron, 0.2 mg/L indole acetic acid, 250 mg/L carbenicillin, 150 mg/L kanamycin, pH 5.8) until shoot initials of approximately 1 to 2 mm in length became apparent. At this stage explants were transferred to shoot induction media (MS30 with 1 mg/L benzylaminopurine, 0.2 mg/L indole acetic acid, 250 mg/L carbenicillin, 150 mg/L Kanamycin, pH 5.8). Once shoots reached approximately 8 to 10 mm in height and exhibited some signs of forming leaf structures, they were excised from the underlying mother explant and transferred to rooting media (MS30 with 250 mg/L carbenicillin, 150 mg/L kanamycin, pH 5.8). The above described protocol was repeated for transformation of *Kalanchoe fedtschenkoi* with the several 3xHA transcription factor fusion constructs and an empty vector ImpGWB502 control harboring only a 3xHA tag except that hygromycin at 1 mg/L was used for selection instead of kanamycin. A single transgenic line of *Kalanchoe fedtschenkoi* previously transformed with the *Klp**Ppc*-INTACT construct was used to generate all TF-3xHA fusion and control lines for ChIP-seq analysis. A graphical depiction of the stable transformation process in provided in Supplemental Figure 6.

Confirmation and expression analysis of stably transformed Kalanchoe lines

100 mg fresh leaf tissue was obtained from putative stable transformants which survived and rooted on rooting media and used for total genomic DNA isolation with the Qiagen DNeasy Plant Mini Kit according to product instructions although nuclease-free water was used for final elution of genomic DNA. To confirm the presence of *Kl*p*Ppc*-INTACT PCR was conducted using primers designed to amplify the *NTF* and *BirA* transgenes. To confirm the presence of the various TF-3xHA fusion constructs PCR was conducted using primers targeting the hpt hygromycin resistance gene. Expression of the NTF peptide in *Kl*p*Ppc*-INTACT transformants was confirmed by the visualization of a nuclear-localized GFP signal in mature leaf tissue using the microscopy protocol described above. To confirm expression of TF-3xHA fusions total RNA was isolated and cDNA synthesis was carried out as described above. Semi-quantitative RT-PCR was then conducted using primers targeting the 3xHA tag and the constitutively expressed *KfUBQ10* transcript. PCR was performed using 30 cycles as the expression of *KfUBQ10* was consistently saturated at this cycle number, but clear differences in the expression levels of the 3xHA tag among the different independent transgenic lines were apparent. Transgenic lines (3 for each construct) exhibiting an intermediate expression of each TF-3xHA fusion by visual inspection of the amplified products were selected for further experiments (pGWB515-3xHA-*KfNF-YB3* lines 2, 4, 6; pGWB515-3xHA-*KfHD-like* lines 2, 3, 4; pGWB514-*KfHD-like*-3xHA lines 2, 3, 5; pGWB515-3xHA-*KfMYB59* lines E, I, L; pGWB514-*KfMYB59*-3xHA lines 1, 7, 11) except in those cases where only 3 independent transformed lines in total were recovered (pGWB514-*KfNF-YB3*-3xHA lines 1, 2, 3; and pGWB502-3xHA lines 2, 3, 4). The sequences of primers described in this

section are provided in Supplemental Table 1. PCR and RT-PCR results described in this section are provided in Supplemental Figure 7.

Kalanchoe nuclear isolation

Three independent transformed lines for each transcription factor, *KfNF-YB3*, *KfHD-like*, and *KfMYB59* expressed separately as an N and C-terminus fusion to 3xHA along with three independent lines transformed with an empty ImpGWB502-3xHA vector were used for mesophyll nuclei enrichment attempts. Initially, a cell type-specific nuclear enrichment protocol for ChIP-seq samples developed for *Arabidopsis* (Deal & Henikoff, 2010) was followed with little modifications. 5 g of mature *Kalanchoe* leaves were collected from each transgenic line from tissue culture-grown plants and cut in half with a scalpel. Protein-DNA crosslinking was achieved by adding 35 mL of nuclear purification buffer with formaldehyde (NPBf) (20 mM MOPS, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 1% v/v formaldehyde, pH 7.0) to the cut leaves followed by vacuum infiltration at RT for 15 minutes at -0.76 kPa. 2M glycine was then added to a final concentration of 125 mM to stop the crosslinking reactions and vacuum infiltration was continued for an additional 5 minutes. Infiltrated leaves were washed twice with water and blotted dry. Leaves were then ground to a fine powder using liquid N_2 and a mortar and pestle. Frozen ground tissue was added to 10 mL of ice-cold nuclear purification buffer (NPB) (20 mM MOPS, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.2 mM spermine, 0.01X Roche Complete Mini EDTA-free Protease Inhibitor Tablets, 0.5 mg/L E64, pH 7.0). The resulting suspensions were filtered through a 70 µm cell strainer. The filtered suspensions were next centrifuged at

1,000 *x g* for 10 minutes at 4 °C. After discarding the supernatants, the nuclei and cell debris pellets were re-suspended gently with a paintbrush in 1 mL of ice-cold NPB buffer containing 2 µg/mL DAPI and incubated on ice for 3 minutes. These resulting suspensions were again centrifuged at 1,000 *x g* for 10 minutes at 4 °C with the supernatants discarded. Nuclei pellets were re-suspended in 1 mL of ice-cold NPB buffer. 25 µL M-280 streptavidin-coated Dynabeads per each nuclei sample (Thermo Fisher Scientific product number 11205D) were washed in 1 mL of ice- cold NPB buffer, pelleted by centrifugation at 3,500 *x g* for 2 minutes at 4 °C, and re-suspended in 25 µL of NPB buffer. All 25 µL of the M-280 streptavidin-coated Dynabead re-suspension was then added to each nuclei suspension and samples were placed on a rotating mixer for 30 minutes at 4 °C. Pipet tips for use in the column-flow capture procedure were treated as described (Deal & Henikoff, 2010) with NPB buffer containing 1% vol/vol BSA. 1 mL mixtures of nuclei samples and M280 Dynabeads were added to 9 mL of ice-cold NPB buffer containing 0.1% vol/vol Triton X-100. The resulting 10 mL suspensions were then used for 2 rounds of the column magnet-capture nuclei enrichment procedure exactly as described by Deal & Henikoff (2010). Following this flow-purification procedure, the resulting 1 mL presumably enriched samples were pelleted by centrifugation at 1,000 *x g* for 5 minutes at 4 °C. Samples were then re-suspended in 100 µL ice-cold NPB. However, nuclei samples obtained with this protocol showed no enrichment and exhibited a high degree of contamination with debris and chloroplasts. A heavily modified discontinuous density gradient protocol was then developed for purification of *Kalanchoe* nuclei prior to further enrichment attempts: 10 g of mature *Kalanchoe* leaves were collected from each transgenic line from tissue culture-grown

plants and cut in half with a scalpel. Protein-DNA crosslinking was achieved by adding 70 mL of nuclear purification buffer with formaldehyde (NPBf) (20 mM MOPS, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 1% v/v formaldehyde, pH 7.0) to the cut leaves followed by vacuum infiltration at RT for 15 minutes at -0.76 kPa. 2M glycine was then added to a final concentration of 125 mM to stop the crosslinking reactions and vacuum infiltration was continued for an additional 5 minutes. Infiltrated leaves were washed twice with water and blotted dry. Leaves were then ground in 60 mL of ice cold nuclear purification buffer (NPB) (20 mM MOPS, 40 mM NaCl, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine, 0.2 mM spermine, 0.01X Roche Complete Mini EDTA-free Protease Inhibitor Tablets, 0.5 mg/L E64, pH 7.0) with a tissue homogenizer (Cole-Parmer Hand-Held Homogenizer, 100-1000 mL, 2.5cm Stator; 110V) on low speed on ice. The homogenized tissue was then filtered through a 70 µm cell strainer. The filtered suspensions were next centrifuged at 1,000 *x g* for 10 minutes at 4 °C. After discarding the supernatants, the nuclei and cell debris pellets were re-suspended gently with a paintbrush in 10 mL of ice cold NPB buffer containing 2 µg/mL DAPI. These resulting suspensions were again centrifuged at 1,000 *x g* for 10 minutes at 4 °C with the supernatants discarded. Pellets were re-suspended in 12 mL of ice cold NPB buffer containing 0.7% v/v Triton X-100. These suspensions were next overlaid onto a 2-step Percoll gradient (25% and 75%) prepared in NPB. A 15 mL volume was used for each step in the gradient. Samples were then centrifuged at 2,000 *x g* for 20 minutes at 4 °C. Nuclei layers were present at the interface between the 25% and 75% Percoll layers after centrifugation. These nuclei bands were aspirated with a serological pipet and re-suspended once again in 12 mL of ice cold NPB buffer containing 0.7% v/v

Triton X-100. The Percoll gradient centrifugation was repeated. Nuclei layers appeared white and clean after this second round of gradient centrifugation and were transferred to a fresh tube and diluted with 3 volumes of ice cold NPB buffer. Samples were centrifuged at 1,000 *x g* for 10 minutes at 4 °C. Supernatants were removed and the nuclei were re-suspended in 20 mL of ice cold NPB buffer and centrifuged as described above. Supernatants were removed and nuclei were re-suspended in 1.5 mL ice cold NPB and frozen at -80 °C.

Whole-leaf sample nuclei purified with the protocol described above were used for another attempt at mesophyll enrichment with the flow column and magnet capture apparatus and protocol described by Deal & Henikoff (2010). These efforts were yet again unsuccessful and chromatin immunoprecipitation as described below commenced with non-enriched nuclei from whole-leaf tissue. Approximately 2 to 3 million nuclei per-sample were used for each chromatin prep described below. This estimation was obtained by counting the DAPI-stained nuclei on a Thoma Haemocytometer Counting Chamber using a Keyence BZ-X710 microscope. For DAPI imaging, a Keyence model OP-87762 filter was used with an excitation wavelength of 360/40 nm and an emission wavelength of 460/50 nm.

Chromatin immunoprecipitation

Chromatin immunoprecipitation for ChIP-seq was performed in a similar manner to that described by Deal & Henikoff (2010). 6 volumes of room temperature nuclear lysis buffer (50 mM Tris, pH 8.0, 10 mM EDTA, 1% wt/vol SDS, 2X Roche Complete Mini EDTA-free Protease Inhibitor Tablets, 0.5 µg/mL E64 protease inhibitor) was added

to each purified nuclear pellet and samples were vortexed vigorously for 2 minutes. Lysed nuclei were next sonicated with a Fisher Scientific Sonic Dismembrator Model 60 set to high-power in pulses of 30 s on and 30 s off for 5 minutes on ice. Samples were centrifuged at 16,000 *x g* at room temperature for 2 minutes to pellet debris. Supernatants were transferred to new tubes and diluted 10-fold with ice-cold ChIP dilution buffer (16.7 mM Tris, pH 8.0, 167 mM NaCl, 1.1% vol/vol Triton X-100, 1.2 mM EDTA). 50 µL Protein G Agarose beads/Salmon Sperm DNA (50 % slurry) (Sigma Aldrich product number 16-201) per chromatin sample was diluted in 20 volumes of ChIP dilution buffer. Beads were washed by inverting the tube several times and centrifuging at 3,500 *x g* for 2 minutes at 4 °C to pellet the beads. Supernatants were discarded and beads were restored to their original volume in ChIP dilution buffer. 2 mL of each cleared diluted chromatin sample described above was then incubated with 50 µL of the washed bead slurry for 1 hour at 4 °C with rotation on a rotating mixer to clear chromatin samples of proteins or nucleic acids which bind the beads non-specifically. Samples were then centrifuged at 3,500 *x g* for 2 minutes at 4 °C to pellet the beads. Chromatin supernatants were transferred to fresh tubes. For capture of TF-DNA complexes, the pre-cleared chromatin samples were incubated with Anti-HA High Affinity monoclonal clone 3F10 antibody (Roche product number 11867423001) at a concentration of 0.5 µg/mL chromatin. After addition of the antibody, samples were placed on a rotating mixer and mixed for 18 h at 4 °C. More (60 µL per 2 mL chromatin sample) washed Protein G Agarose beads/Salmon Sperm DNA (50 % slurry) was prepared in ChIP dilution buffer as described above. To collect antibody-transcription factor-DNA complexes, 60 µL of the washed bead slurry was added to each

chromatin/antibody sample at the end of the 18 h incubation. Samples were then rotated on a rotating mixer for 1.5 h at 4 °C. Resulting immune complexes were then pelleted by centrifugation at 3,500 *x g* for 2 minutes at 4 °C. Supernatants were discarded and the pelleted complexes were washed sequentially with low-salt wash buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 0.1% wt/vol SDS, 1% vol/vol Triton X-100, 2 mM EDTA), high-salt wash buffer (20 mM Tris, pH 8.0, 500 mM NaCl, 0.1% wt/vol SDS, 1% vol/vol Triton X-100, 2 mM EDTA), LiCl wash buffer (10 mM Tris, pH 8.0, 250 mM LiCl, 1% wt/vol sodium deoxycholate, 1% vol/vol IGEPAL CA-630, 1 mM EDTA), and TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA). Samples were washed once with 1 mL of each buffer for 5 minutes with rotation on a rotating mixer at 4 °C. After each wash, bead/immune complexes were pelleted by centrifugation at 3,500 *x g* for 2 minutes at 4 °C. Following pelleting and supernatant removal after the TE buffer wash, 200 µL of ChIP elution buffer (100 mM sodium bicarbonate, 1% wt/vol SDS) was added to each sample. Samples were then vigorously vortexed at room temperature for 5 minutes and centrifuged at 3,500 *x g* for 2 minutes at room temperature. Supernatants containing eluted chromatin were then treated with 20 µL of 5.0 M NaCl and heated to 100 °C for 15 minutes. Samples were centrifuged at 16,000 *x g* briefly at room temperature to collect condensation. 1 µL RNase A was added to each sample followed by incubation for 15 minutes at 37 °C. 1 µL of Proteinase K was then added to each sample with subsequent incubation at 50 °C for 30 minutes. ChIP'ed DNA was then purified with the Qiagen PCR Purification Kit beginning with the addition of the kit PB buffer. Final DNA elution was performed with nuclease-free water (50 µL per sample) which was heated to 70 °C prior to elution. Elution was performed twice with the same 50 µL portion of nuclease-

free water. Purity and DNA approximate quantities were assessed with a Thermo Fisher Scientific NanoDrop 2000c spectrophotometer and precise quantities assessed with a Thermo Fisher Scientific Qubit 4 Fluorometer using the Qubit dsDNA BR (Broad Range) Assay Kit (Thermo Fisher Scientific product number Q32853).

ChIP-sequencing and bioinformatics processing of data

Immunoprecipitated DNA samples were sequenced on an Illumina HiSeq PE150 platform with library preparation and sequencing services provided by Novogene Co., Ltd., Beijing, China. Raw read quality was assessed with FastQC (Brown et al., 2017). Adapter trimming was performed using Trim Galore (Brown et al., 2017). Alignment to the *Kalanchoe fedtschenkoi* genome was performed with the Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009). Duplicate, un-mappable, ambiguously mapped, and high-mismatch containing reads were discarded using deepTools (Ramírez et al., 2014). Peak calling was then performed with MACS2 (Zhang et al., 2008). Motif discovery, comparison to known motifs, and discovery of occurrences of predicted motifs in the *Kalanchoe fedtschenkoi* genome was performed with the STREME, Tomtom, and FIMO algorithms, respectively (Bailey et al., 2015).

Results

NF-YB3, HD-like, and MYB59 are differentially expressed in C_3 photosynthesis vs. CAM

Mean increases in mRNA transcript abundance of *KlNF-YB3, KlHD-like,* and *KlMYB59* obtained through C_3 photosynthesis *vs.* CAM comparative RNA-seq in *Kalanchoe laxiflora* are presented (Figure 1). *KlNF-YB3* possessed a mean FPKM value

of 883.7 when transcript abundance levels in C_3 photosynthesis and CAM were averaged and displayed a 4.5 log_2 fold- increase in mRNA abundance in leaf-pair 6 (CAM) in comparison to leaf pair 1 (C_3). *KlHD-like* possessed an average FPKM value of 760.6 between C_3 photosynthesis and CAM with a 4.5 log2 fold-induction in leaf-pair 6 relative to leaf pair 1. *KlMYB59* with an average CAM/C_3 transcript abundance of 209.9 FPKM and a 22 log_2 fold-induction in CAM photosynthesis displayed the greatest induction in leaf-pair 6 relative to leaf pair 1 of any high-abundance transcription factor identified through RNA-seq analysis (those with a minimum mean transcript level of 100 FPKM when C_3 photosynthesis and CAM values were averaged). 24 h diel time course data with a 12 h photoperiod and a 12 h dark period for *Kalanchoe fedtschenkoi NF-YB3, HD-like,* and *MYB59* in CAM-performing leaf pair 6 tissue are presented (Figure 2). These plots clearly appear to show a circadian-phase dependent regulation of the induction of these transcription factors with obvious peak expression levels at dusk and just before dawn for *KfNF-YB3* and *KfHD-like*, respectively. *KfMYB59* appears to exhibit perhaps the most robust oscillation across both light and dark periods with peak expression at midday.

KfNF-YB3, KfHD-like, and KfMYB59 are nuclear localized in Kalanchoe fedtschenkoi

Fluorescent microscopy analysis of *Agrobacterium*-infiltrated mature leaf tissue from *Kalanchoe fedtschenkoi* transiently expressing *KfNF-YB3, KfHD-like*, and *KfMYB59* as separate N and C-terminus sGFP fusions is presented (Figure 3). All three transcription factors displayed specific nuclear localization as both N- and C-terminus sGFP fusions with the resulting signal overlapping with DAPI fluorescence. No sGFP

signal from any other cellular location or structure was visible. As compared with transient protein expression in common C_3 photosynthesis model plants, such as *N. benthamiana*, we have observed transient expression in *Kalanchoe* species to be comparatively low-efficiency. Experiments such as that described here typically result in only a very small proportion of cells in infiltrated tissue displaying signs of recombinant protein expression. Hence, only a few isolated cells displaying a nuclear sGFP signal are seen in Figure 3.

The INTACT system was successfully expressed in Kalanchoe fedtschenkoi

While mesophyll cell enrichment of nuclei failed using both whole leaf tissue and purified nuclear samples as input for the INTACT approach, strong expression of the *NTF* transgene under the control of the 2.8 kb *Kalanchoe laxiflora Ppc1* promoter was confirmed in the mesophyll cells of *Kalanchoe fedtschenkoi* stably transformed with the *Kl*p*Ppc*-INTACT construct (Figure 4). The particular independent transformed line depicted in Figure 4 appeared to have the strongest and most consistent expression of the eGFP-labeled NTF protein as indicated by microscopy analysis of multiple leaf samples. Hence, this line was chosen for stable transformation with the several 3xHA-tagged transcription factor constructs to be used for ChIP-seq experiments.

ChIP-seq analysis of KfNF-YB3, KfHD-like, and KfMYB59 failed to identify enriched loci

Phred score distributions of Illumina paired-end reads for all experimental and control samples are shown in Figure 5. Although the majority of these reads were high-quality with a per-base accuracy level of 99.9% subsequent bioinformatics processing

failed to predict enriched transcription factor genomic binding sites. As seen in Figure 6, the percentage of reads which could be uniquely mapped to the *Kalanchoe fedtschenkoi* genome was low, approximately 20% across all samples and controls. Although a fairly low mapped read percentage might be expected in a species such as *Kalanchoe fedtschenkoi* with a genome which is not as well annotated as other model species, unsatisfactory results were also obtained with regard to peak calling and correlating these peaks with gene bodies. Figure 7 displays the total peak counts for each experimental sample after the subtraction of control sample data. Only approximately two to four thousand peaks were able to be called for each sample. A genome-wide distribution of these peaks in relation to annotated *K. fedtschenkoi* gene bodies is shown in Figure 8. Notably, the results of this plot shows poor enrichment of peaks in locations which would normally be expected to contain a large proportion of peaks in transcription factor-based ChIP-seq data, mainly promoter regions and transcriptional start sites. As can be seen in this figure, a large number of peaks occur within coding regions or in the areas downstream of genes.

Predicted binding motifs lack significance and are composed of repetitive elements

The top 25 most significant motifs from peak regions for *KfNF-YB3, KfHD-like,* and *KfMYB59* are shown in Figures 9, 10, and 11. However, these elicited motifs possessed low statistical significance and were comprised, to a large degree, of repetitive sequence elements with poor homology to known preferred binding motifs of these classes of transcription factors upon analysis with Tomtom. Moreover, many of the top scoring motifs across all three transcription factors exhibit a similar structure to each

other characterized by low-complexity A/G or A/T-rich repeats. As a result of the repetitive low-significance nature of the motifs elicited for *KfNF-YB3, KfHD-like,* and *KfMYB59,* a FIMO analysis of the *Kalanchoe fedtschenkoi* genome performed to identify the locations of these motifs did not generate data that were useful in predicting putative target genes.

Discussion

Much consideration was given in this study to the particular *Kalanchoe* transcription factors ultimately selected for ChIP-seq analysis. From C_3 photosynthesis *vs*. CAM RNA-seq experiments (Cushman, unpublished data; Hartwell, unpublished data) performed on *Kalanchoe fedtschenkoi* and *Kalanchoe laxiflora,* we have observed several hundred transcription factors to exhibit differential regulation between these two types of photosynthesis. However, many of these transcription factors possess very low mRNA abundances in both C_3 photosynthesis and CAM and/or show only a slight increase in *Kalanchoe* leaf pair 6 as compared with leaf pair 1. Through setting a minimum mean FPKM of 100 when mRNA abundances in C_3 and CAM were averaged along with a minimum C_3 to CAM $\log_2$ fold-induction of 2 this large set of transcription factors was narrowed to a list of 19 genes. The set of *Kalanchoe laxiflora* transcription factors meeting these criteria, including the three selected for ChIP-seq in this study, are presented in Supplemental Table 2. Although the possibility cannot be ruled out that important regulatory factors might be missed using these cut-off criteria, other gene expression profiling studies have made use of similar criteria to identify highly-expressed transcripts with significant expression changes among different conditions (McCarthy &

Smyth, 2009; Peart et al., 2005; Sabharwal et al., 2019). Notably, this study makes use of data obtained in both *Kalanchoe laxiflora* and *Kalanchoe fedtschenkoi*. Although the majority of the experiments described in this study were performed in *K. fedtschenkoi*, the attempted mesophyll nuclei enrichment work makes use of the 2.8 kb *Ppc* promoter from *K. laxiflora*. Initially, the tetraploid species *K. laxiflora* was used as the model *Kalanchoe* species by the authors in this study, but upon the availability of the genome and transcriptome for the diploid *K. fedtschenkoi*, a switch was made to this species due to its reduced gene copy number and allelic variation.

The ultimate failure in this study to successfully enrich mesophyll nuclei using the approach described by Deal and Henikoff (2010) may have stemmed from the nuclear isolation process for *Kalanchoe fedtschenkoi*. Using the nuclear isolation protocol followed by the cell type-specific nuclear capture technique as originally described by Deal and Henikoff (2010) resulted in no detectable captured nuclei. Examination of the flow-through revealed that these nuclear preps were heavily contaminated with cellular debris and a large number of intact chloroplasts. eGFP labeled nuclei were visible in this flow through suggesting that the NTF-protein was still present on the surface of mesophyll nuclei. However, the large amounts of contamination present might have physically hindered these tagged nuclei from interacting with the streptavidin-coated magnetic beads necessary for column-capture using this protocol. Following the Percoll discontinuous density gradient purification procedure, nuclei were largely free of contaminating debris and chloroplasts. Optimization of the concentration of Triton X-100 in the nuclear purification buffers used in this density gradient protocol was important to obtaining a satisfactory balance between the purity and integrity of isolated

nuclei. A range of Triton X-100 concentrations was tested with the ultimate choice of 0.7% being made as this resulted in nuclear extracts which showed only slight damage to nuclei upon microscopic examination yet were free of most contaminating debris and chloroplasts. However, a concentration of 0.7% Triton X-100 is still very high compared with the concentrations of this nonionic surfactant that are typically used in nuclear extraction protocols. Triton X-100 concentrations of 0.1% are frequently reported for nuclear isolation protocols performed on *Arabidopsis* and tomato (Deal and Henikoff, 2010; Reynoso et al., 2018). Nuclear preps of mammalian cells often also report 0.1% (Steiner et al., 2012). Although higher concentrations of 0.5% to 1.0 % have been reported for plants such as tobacco, potato, and apple, these nuclei were not subsequently evaluated in INTACT protocol-based cell type-specific nuclear enrichment procedures. Indeed, exposure of nuclei to these high concentrations of Triton X-100 is known to damage the nuclear envelope (Sikorskaite et al., 2013). Furthermore, due to the succulent nature of *Kalanchoe*, we found it necessary to scale up the amount of tissue and volumes of the buffers used to obtain suitable yields of nuclei. As a result of this, protease inhibitors were used at lower–than recommended concentrations due to cost. This factor might have facilitated any proteolytic degradation of NTF. Although eGFP fluorescence was still largely present in nuclear samples that were processed with the density gradient purification protocol used in this study, biotin might have been lost from the NTF peptide. This would explain the failure of these nuclei to bind to streptavidin beads. However, this possibility was not investigated through blotting experiments on nuclear extracts.

The failure of the subsequent ChIP-seq experiments described in this chapter to provide useful data regarding genomic binding sites and preferred DNA sequence motifs might have resulted from any number of problems in the experimental design as detailed above. Although the specific cause(s) are unclear, one of the most frequent reasons for unsatisfactory ChIP-seq data is poor antibody performance. Although many of the commercially available antibodies, including the one used in this study, are advertised as "ChIP-grade", many of these have failed rigorous specificity tests and/or display substantial lot-to-lot variations in performance (Furey, 2012; Park, 2009). Also, failure of the initial crosslinking of protein-DNA interactions in *Kalanchoe fedtschenkoi* leaf tissue might have contributed to the poor ChIP-seq results. Due to the thick, waxy, succulent nature of *Kalanchoe* leaves, crosslinking with formaldehyde was performed by cutting each leaf in half followed by vacuum infiltration of the formaldehyde solution. Although a visual inspection of the leaves after this step did appear to show efficient infiltration, it remains a possibility that this step was not properly optimized for *Kalanchoe*. Indeed, several other transcription factor ChIP-seq experiments in *Kalanchoe fedtschenkoi* have failed similarly when leaf tissue crosslinked in similar manner to that described here was submitted for ChIP-seq services through the epigenetics profiling company Diagenode (Diagenode, a Hologic company, Denville, NJ) (J. Hartwell, personal communication, October 2021).

In recent years, new approaches have been described for the genome-wide discovery of transcription factor binding sites. If the failure of the ChIP-seq experiments described in this chapter were due to unknown causes resulting from the immunoprecipitation process or initial tissue crosslinking procedure, these approaches

may prove to be superior. CUT&RUN is one such of these techniques and utilizes whole permeabilized nuclei immobilized on lectin-coated beads. An antibody against the particular transcription factor, or against an epitope tag, which is bound to micrococcal nuclease is then introduced into the mixture. This complex diffuses into the nuclei and, as a result of antibody binding and MNase cleavage, small chromatin fragments corresponding to the sites bound by the particular transcription factor are liberated. These fragments are then released into the supernatant through a centrifugation step and used for library preparation and DNA sequencing. Crosslinking and immunoprecipitation are not required with this protocol. Furthermore, because specific chromatin fragments are generated by the action of the antibody-MNase complex, contaminating non-specific background chromatin is reduced (Skene & Henikoff, 2017). A modification of this approach, CUT&Tag, is similar but utilizes a hyperactive Tn5 transposase fusion protein to simultaneously generate chromatin fragments and ligate sequencing adapters (Kaya-Okur et al., 2019). While these approaches, like conventional ChIP-seq, both share the need for high-quality antibody performance, they do circumvent some of the potential points for error in standard chromatin immunoprecipitation and may be worth attempting in *Kalanchoe*. In conclusion, this study attempted genome-wide binding site profiling of three transcription factors which display a marked induction upon the transition from C_3 to CAM in *Kalanchoe.* Ultimately, the failure of the final ChIP-sequencing data to provide biological insights might be the result of many potential causes arising from the challenges of working with a non-model organism.

Acknowledgements

The authors would like to thank Dr. James Hartwell at the Institute of Systems, Molecular and Integrative Biology, University of Liverpool, United Kingdom for providing much of the RNA-seq data referenced in this study. We are grateful for the tissue culture assistance provided by Kristen Straight, Bethany Townley, Bailey Watkins, and Daisy Tanner. We thank our current laboratory manager Lisa Petrusa and previous laboratory manager Rebecca Albion for their contributions in facilitating the research described in this study. We would like to acknowledge the VIB-Ugent Center for Plant Systems Biology Plasmid Repository at Ghent University, Ghent, Belgium for furnishing plasmid materials used in this study. This work was supported by the Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834, the Nevada Agricultural Experiment Station, and the UNR Foundation.

References

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, *43*(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

Borland, A. M., Griffiths, H., Hartwell, J., & Smith, J. A. C. (2009). Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *Journal of Experimental Botany*, *60*(10), 2879–2896. https://doi.org/10.1093/jxb/erp118

Boyer, J. S. (1982). Plant Productivity and Environment. *Science*, *218*(4571), 443–448. https://doi.org/10.1126/science.218.4571.443

Brilhaus, D., Bräutigam, A., Mettler-Altmann, T., Winter, K., & Weber, A. P. M. (2016). Reversible Burst of Transcriptional Changes during Induction of Crassulacean Acid Metabolism in *Talinum triangulare*. *Plant Physiology*, *170*(1), 102–122. https://doi.org/10.1104/pp.15.01076

Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, *33*(19), 3137–3139. https://doi.org/10.1093/bioinformatics/btx373

Chew, W., Hrmova, M., & Lopato, S. (2013). Role of Homeodomain Leucine Zipper (HD-Zip) IV Transcription Factors in Plant Development and Plant Protection from Deleterious Environmental Factors. *International Journal of Molecular Sciences*, *14*(4), 8122–8147. https://doi.org/10.3390/ijms14048122

Cushman, J. C., Tillett, R. L., Wood, J. A., Branco, J. M., & Schlauch, K. A. (2008). Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C3 photosynthesis and Crassulacean acid metabolism (CAM). *Journal of Experimental Botany*, *59*(7), 1875–1894. https://doi.org/10.1093/jxb/ern008

Deal, R. B., & Henikoff, S. (2010). A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. *Developmental Cell*, *18*(6), 1030–1040. https://doi.org/10.1016/j.devcel.2010.05.013

Fasani, E., DalCorso, G., Costa, A., Zenoni, S., & Furini, A. (2019). The Arabidopsis thaliana transcription factor MYB59 regulates calcium signalling during plant growth and stress response. *Plant Molecular Biology*, *99*(6), 517–534. https://doi.org/10.1007/s11103-019-00833-x

Furey, T. S. (2012). ChIP–seq and beyond: New and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics*, *13*(12), 840–852. https://doi.org/10.1038/nrg3306

Gong, S., Ding, Y., Hu, S., Ding, L., Chen, Z., & Zhu, C. (2019). The role of HD-Zip class I transcription factors in plant response to abiotic stresses. *Physiologia Plantarum*, *167*(4), 516–525. https://doi.org/10.1111/ppl.12965

Hickman, R., Van Verk, M. C., Van Dijken, A. J. H., Mendes, M. P., Vroegop-Vos, I. A., Caarls, L., Steenbergen, M., Van der Nagel, I., Wesselink, G. J., Jironkin, A., Talbot, A., Rhodes, J., De Vries, M., Schuurink, R. C., Denby, K., Pieterse, C. M. J., & Van Wees, S. C. M. (2017). Architecture and Dynamics of the Jasmonic Acid Gene Regulatory Network. *The Plant Cell*, *29*(9), 2086–2105. https://doi.org/10.1105/tpc.16.00958

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*, *10*(1), 1930. https://doi.org/10.1038/s41467-019-09982-5

Kellogg, E. A. (2013). C4 photosynthesis. *Current Biology*, *23*(14), R594–R599. https://doi.org/10.1016/j.cub.2013.04.066

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

McCarthy, D. J., & Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*, *25*(6), 765–771. https://doi.org/10.1093/bioinformatics/btp053

Mu, R.-L., Cao, Y.-R., Liu, Y.-F., Lei, G., Zou, H.-F., Liao, Y., Wang, H.-W., Zhang, W.-K., Ma, B., Du, J.-Z., Yuan, M., Zhang, J.-S., & Chen, S.-Y. (2009). An R2R3-type transcription factor gene AtMYB59 regulates root growth and cell cycle progression in Arabidopsis. *Cell Research*, *19*(11), 1291–1304. https://doi.org/10.1038/cr.2009.83

Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., Watanabe, Y., Nakamura, K., Kimura, T., & Ishiguro, S. (2007). Improved Gateway Binary Vectors: High-Performance Vectors for Creation of Fusion Constructs in Transgenic Analysis of Plants. *Bioscience, Biotechnology, and Biochemistry*, *71*(8), 2095–2100. https://doi.org/10.1271/bbb.70216

Park, P. J. (2009). ChIP–seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*, *10*(10), 669–680. https://doi.org/10.1038/nrg2641

Peart, M. J., Smyth, G. K., van Laar, R. K., Bowtell, D. D., Richon, V. M., Marks, P. A., Holloway, A. J., & Johnstone, R. W. (2005). Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. *Proceedings of the National Academy of Sciences*, *102*(10), 3697–3702. https://doi.org/10.1073/pnas.0500369102

Quan, R., Hu, S., Zhang, Z., Zhang, H., Zhang, Z., & Huang, R. (2010). Overexpression of an ERF transcription factor *TSRF1* improves rice drought tolerance. *Plant Biotechnology Journal*, *8*(4), 476–488. https://doi.org/10.1111/j.1467-7652.2009.00492.x

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*(W1), W187–W191. https://doi.org/10.1093/nar/gku365

Reynoso, M. A., Pauluzzi, G. C., Kajala, K., Cabanlit, S., Velasco, J., Bazin, J., Deal, R., Sinha, N. R., Brady, S. M., & Bailey-Serres, J. (2018). Nuclear Transcriptomes at High Resolution Using Retooled INTACT. *Plant Physiology*, *176*(1), 270–281. https://doi.org/10.1104/pp.17.00688

Sabharwal, A., Sharma, D., Vellarikkal, S. K., Jayarajan, R., Verma, A., Senthivel, V., Scaria, V., & Sivasubbu, S. (2019). Organellar transcriptome sequencing reveals mitochondrial localization of nuclear encoded transcripts. *Mitochondrion*, *46*, 59–68. https://doi.org/10.1016/j.mito.2018.02.007

Sato, H., Suzuki, T., Takahashi, F., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2019). NF-YB2 and NF-YB3 Have Functionally Diverged and Differentially Induce Drought and Heat Stress-Specific Genes. *Plant Physiology*, *180*(3), 1677–1690. https://doi.org/10.1104/pp.19.00391

Sikorskaite, S., Rajamäki, M.-L., Baniulis, D., Stanys, V., & Valkonen, J. P. (2013). Protocol: Optimised methodology for isolation of nuclei from leaves of species in the Solanaceae and Rosaceae families. *Plant Methods*, *9*(1), 31. https://doi.org/10.1186/1746-4811-9-31

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *ELife*, *6*, e21856. https://doi.org/10.7554/eLife.21856

Somerville, C., Youngs, H., Taylor, C., Davis, S. (2010). SP L: Feedstocks for lignocellulosic biofuels. *Science, (329)*, 790-792. https://doi.org/10.12691/jaem-6-2-4

Steiner, F. A., Talbert, P. B., Kasinathan, S., Deal, R. B., & Henikoff, S. (2012). Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Research*, *22*(4), 766–777. https://doi.org/10.1101/gr.131748.111

Yang, X., Hu, R., Yin, H., Jenkins, J., Shu, S., Tang, H., Liu, D., Weighill, D. A., Cheol Yim, W., Ha, J., Heyduk, K., Goodstein, D. M., Guo, H.-B., Moseley, R. C., Fitzek, E., Jawdy, S., Zhang, Z., Xie, M., Hartwell, J., … Tuskan, G. A. (2017). The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications*, *8*(1), 1899. https://doi.org/10.1038/s41467-017-01491-7

Zhang, J., Hu, R., Sreedasyam, A., Garcia, T. M., Lipzen, A., Wang, M., Yerramsetty, P., Liu, D., Ng, V., Schmutz, J., Cushman, J. C., Borland, A. M., Pasha, A., Provart, N. J., Chen, J.-G., Muchero, W., Tuskan, G. A., & Yang, X. (2020). Light-responsive expression atlas reveals the effects of light quality and intensity in *Kalanchoë fedtschenkoi*, a plant with crassulacean acid metabolism. *GigaScience*, *9*(3), giaa018. https://doi.org/10.1093/gigascience/giaa018

Zhang, T., Zhang, D., Liu, Y., Luo, C., Zhou, Y., & Zhang, L. (2015). Overexpression of a NF-YB3 transcription factor from Picea wilsonii confers tolerance to salinity and drought stress in transformed *Arabidopsis thaliana*. *Plant Physiology and Biochemistry*, *94*, 153–164. https://doi.org/10.1016/j.plaphy.2015.05.001

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Figures and Tables

Figure 1. Fold-induction of *NF-YB3*, *homeodomain-like superfamily protein*, and *MYB59* from C_3 to CAM in *K. laxiflora* as determined by comparative transcriptomics RNA-seq analysis between leaf pair 1 (C_3) and leaf pair 6 (CAM). These candidate transcription factors selected for ChIP-seq analysis were chosen on the basis of at least a $\log_2$ fold-induction of 2 between C_3 and CAM and a mean expression level of at least 100 FPKM when transcript abundance in C_3 and CAM were averaged.

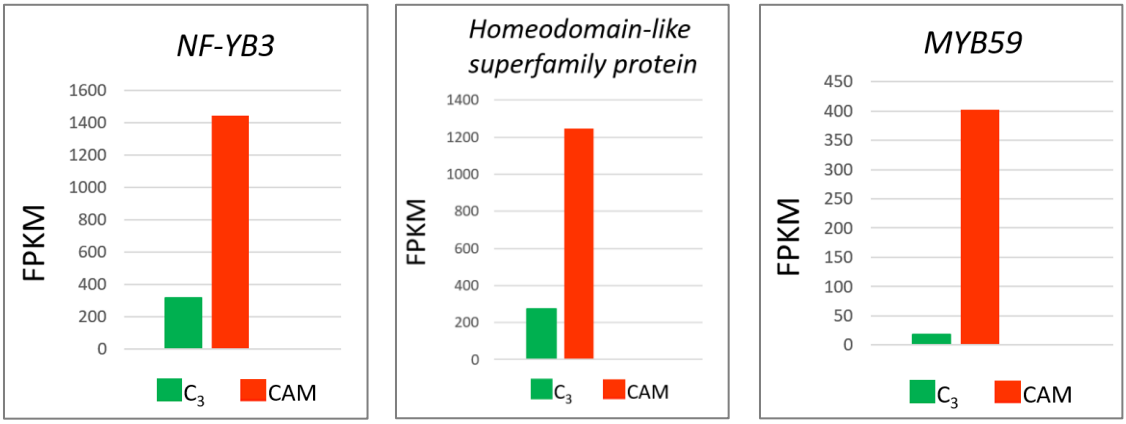


Figure 2. 24 h diel time course of mRNA abundance of *K. fedtschenkoi NF-YB3, homeodomain-like superfamily protein*, and *MYB59*. RNA-seq transcript profiling was undertaken in CAM-performing tissue of *K. fedtschenkoi* (leaf pair 6). Day periods are indicated by a white background and night periods by gray shading. Error bars represent the standard error of the mean (SEM).

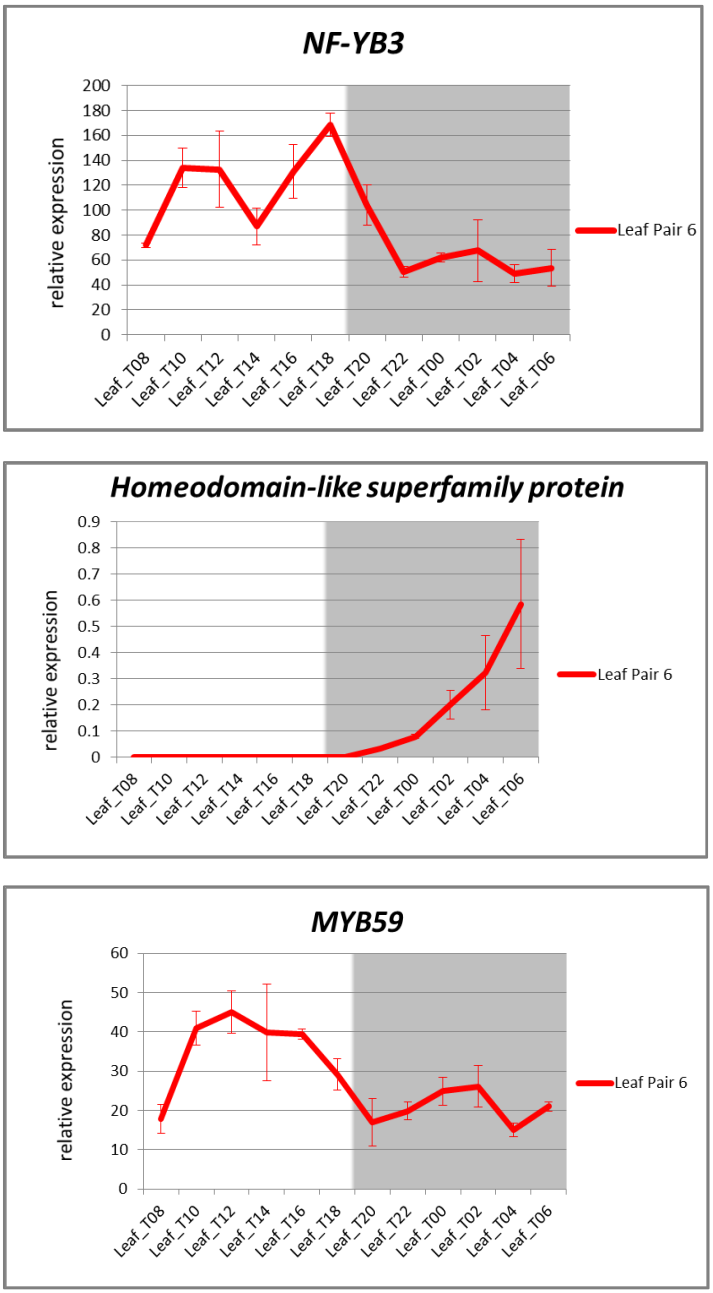


Figure 3. NF-YB3, homeodomain-like superfamily protein, and MYB59 are nuclear localized in *K. fedtschenkoi*. *Agrobacterium* cultures harboring NF-YB3, homeodomain-like superfamily protein, and MYB59 as sGFP fusion constructs were used for transient expression infiltration assays in *K. fedtschenkoi* leaf tissue. Sectioned tissue was then examined with fluorescence microscopy. Consistent with the expected localization of transcription factors, nuclear sGFP signals were observed. DAPI co-localization confirmed nuclear location. Images were taken at 200X magnification using a Keyence BZ-X710 microscope.

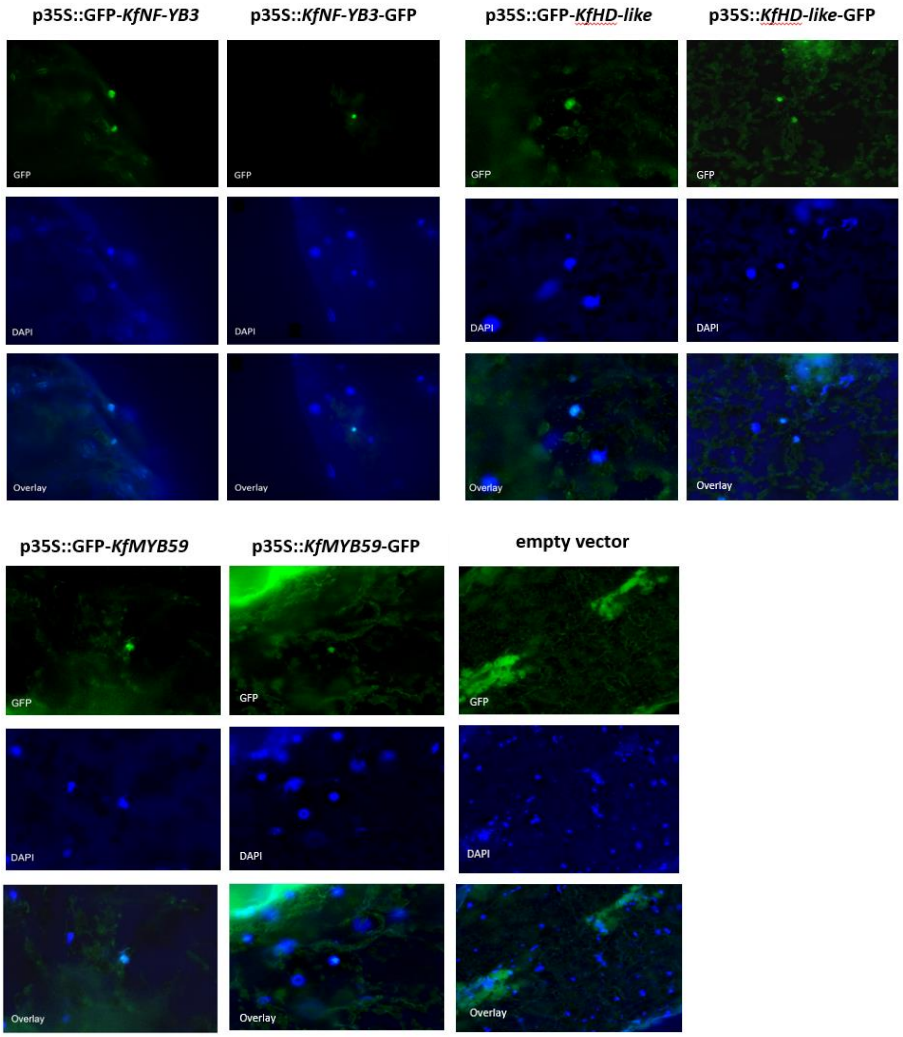


Figure 4. The NTF protein is expressed in *K. fedtschenkoi* mesophyll cells under the control of the 2.8 kb *KlPpc* promoter. eGFP labeled nuclei are evident in the *Kl*p*Ppc*-INTACT transformed sample and indicate expression of the NTF protein under the control of the *Kalanchoe laxiflora* phospho*enol*pyruvate carboxylase (*Ppc*) promoter. DAPI co-localization confirmed nuclear location. Images were taken at 200X magnification using a Keyence BZ-X710 microscope. All samples were obtained from CAM-performing leaf pair 6 tissue.

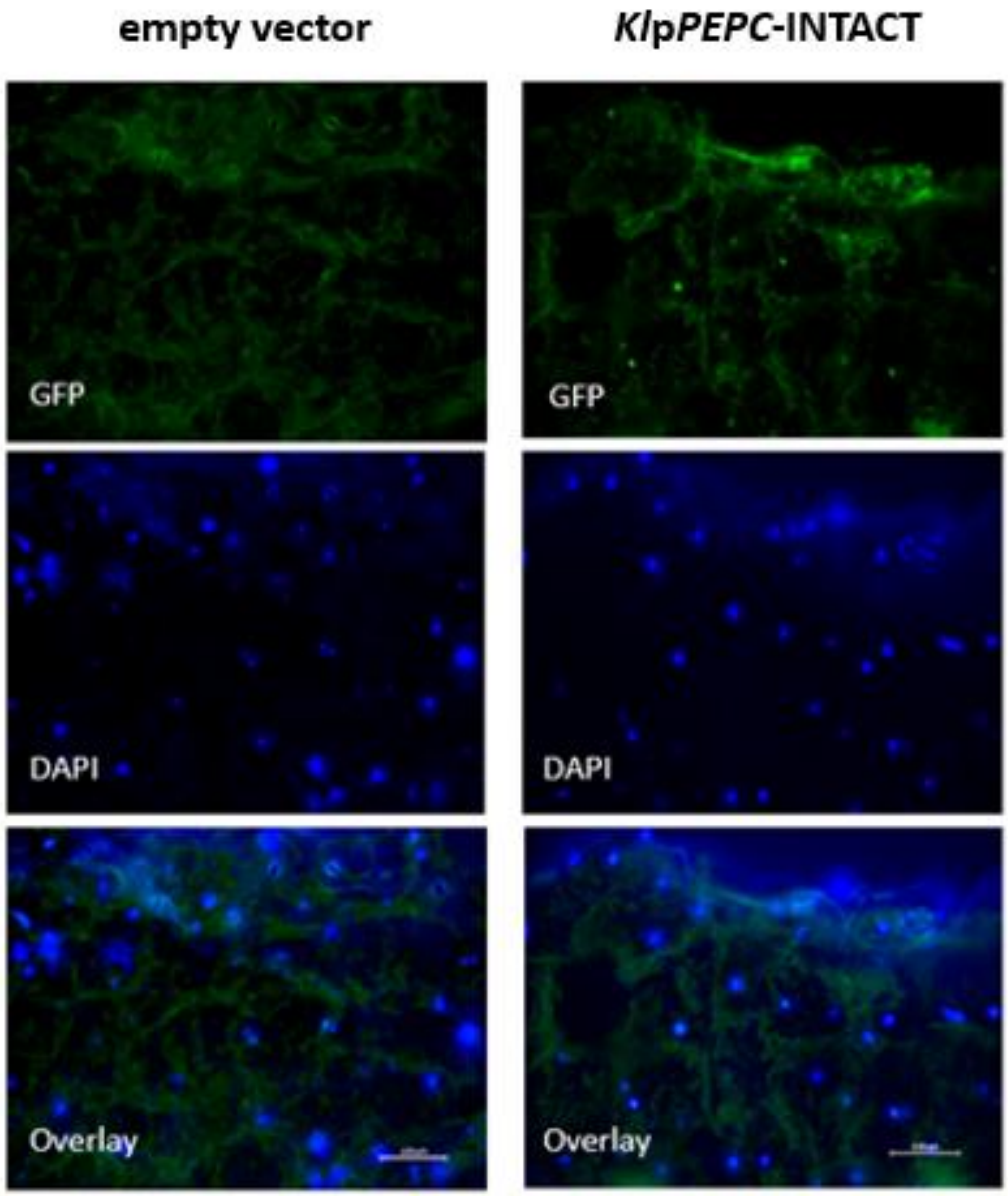


Figure 5. Phred quality score distribution of each base position from Illumina PE reads of ChIP-seq sample replicates and controls. FastQC was used to calculate mean Phred scores from Illumina paired-end read data. The majority of Phred scores were above 30, corresponding to base-call confidence levels of over 99.9%.

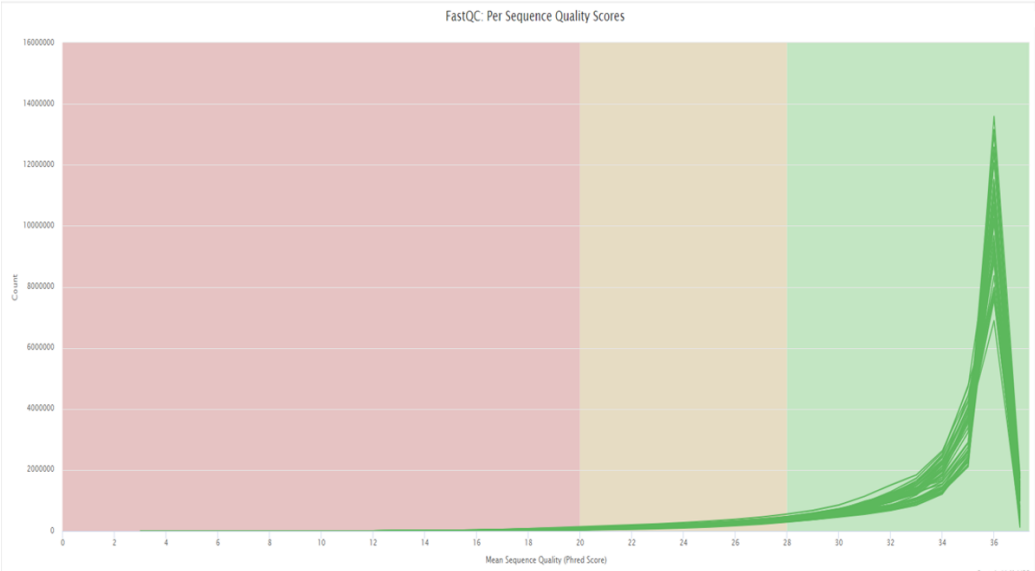


Figure 6. Percentages of mapped *vs.* unmapped reads for each ChIP-seq sample replicate and controls. Illumina reads were mapped to the *K. fedtschenkoi* genome using the Burrows-Wheeler Aligner (BWA). SAMtools was then used to calculate the percentage of mapped reads for each sample or control replicate. Only approximately 20% of reads per-sample were uniquely mappable, below that of what is generally observed for successful ChIP-seq experiments.

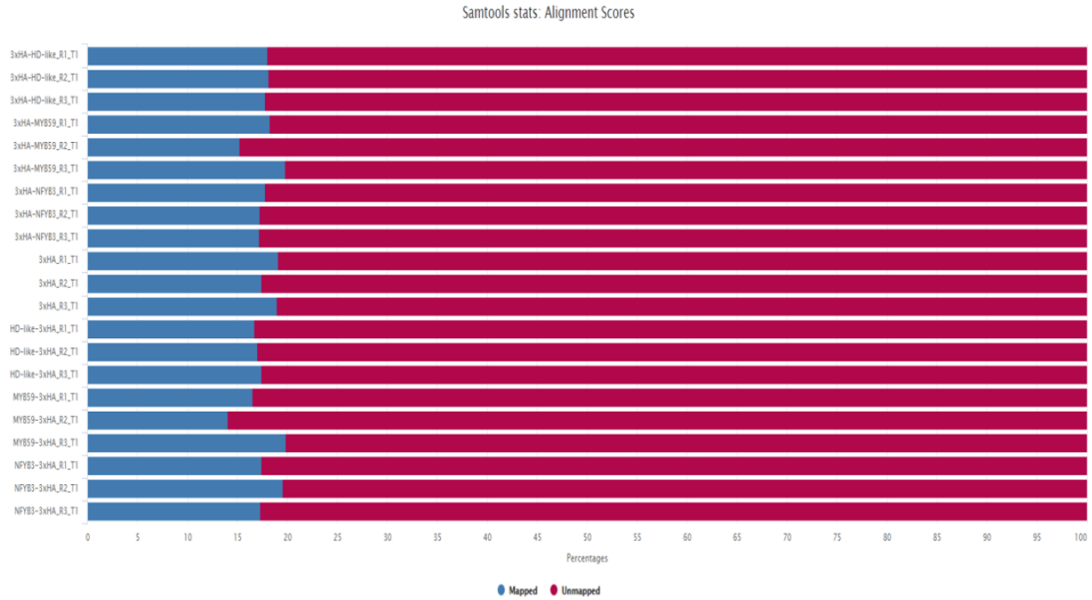


Figure 7. Total called peak count for each *KfNF-YB3*, *KfHD-like* and *KfMYB59* sample replicate. The MACS2 peak caller was used to discover peaks based on genomic sites enriched for mapped reads**.** 3xHA empty vector control data has been subtracted from these counts. Although the number of peaks in ChIP-seq experiments has been reported to be highly variable, the number of called peaks for these samples are very low. This may be at least partially due to the less-extensive annotation of the *K. fedtschenkoi* genome in comparison with other species. This led to difficulties in assigning target genes and in determining functions or gene ontologies enriched near peak regions.

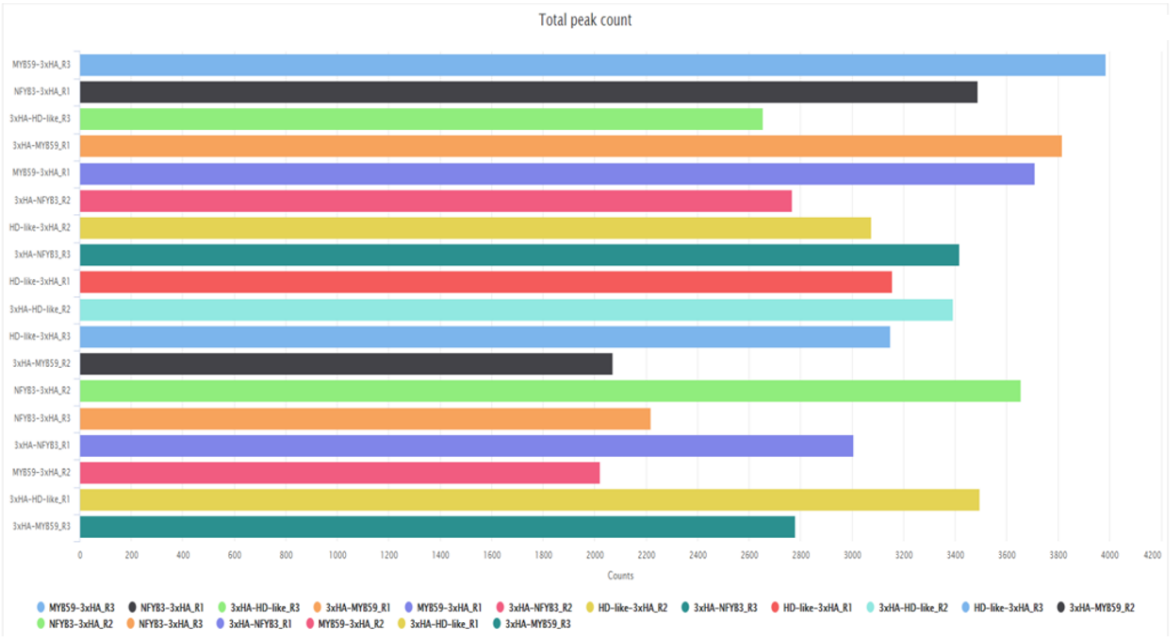


Figure 8. Genome-wide distribution of ChIP-seq sample peaks in relation to gene bodies. DeepTools was used to generate a distribution profile of peak locations for *KfNF-YB3*, *KfHD-like*, and *KfMYB59*. 3xHA control data has been subtracted from this plot. In contrast to what is typically observed in successful ChIP-seq experiments, this plot does not show enrichment of peaks around promoters and transcriptional start sites *vs*. adjacent regions. All annotated *K. fedtschenkoi* genes have been normalized to the same size for the generation of this plot.

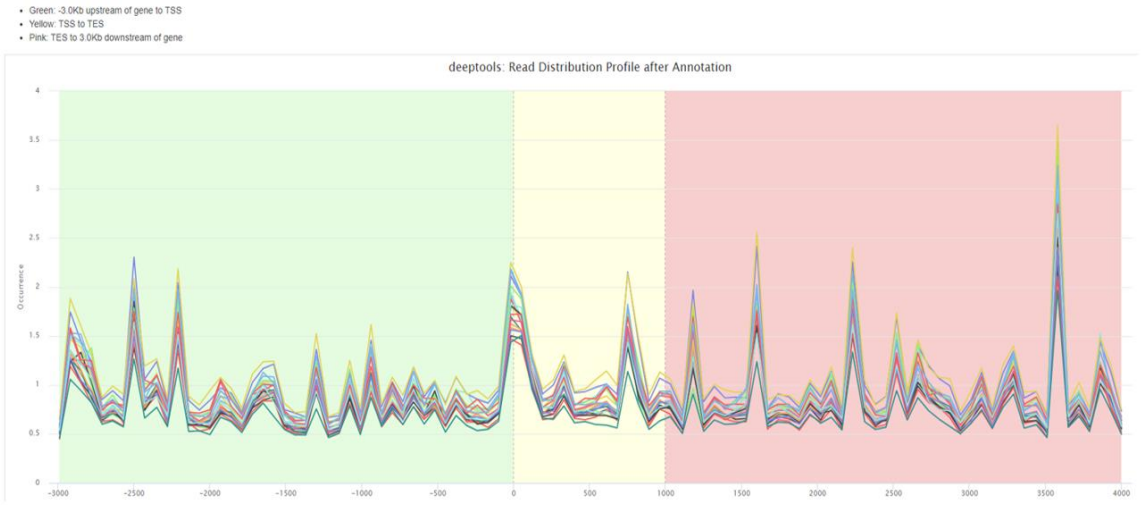


Figure 9. Top 25 most significant motifs at peak regions for *KfNF-YB3*. STREME (Simple, Thorough, Rapid, Enriched Motif Elicitation) analysis was performed on *KfNF-YB3* peak locations in an attempt to elucidate potential binding site preferences. However, the statistical significance of these motifs was low and many contain low-complexity repetitive sequences with little similarity to known binding site preferences of transcription factors from this family.

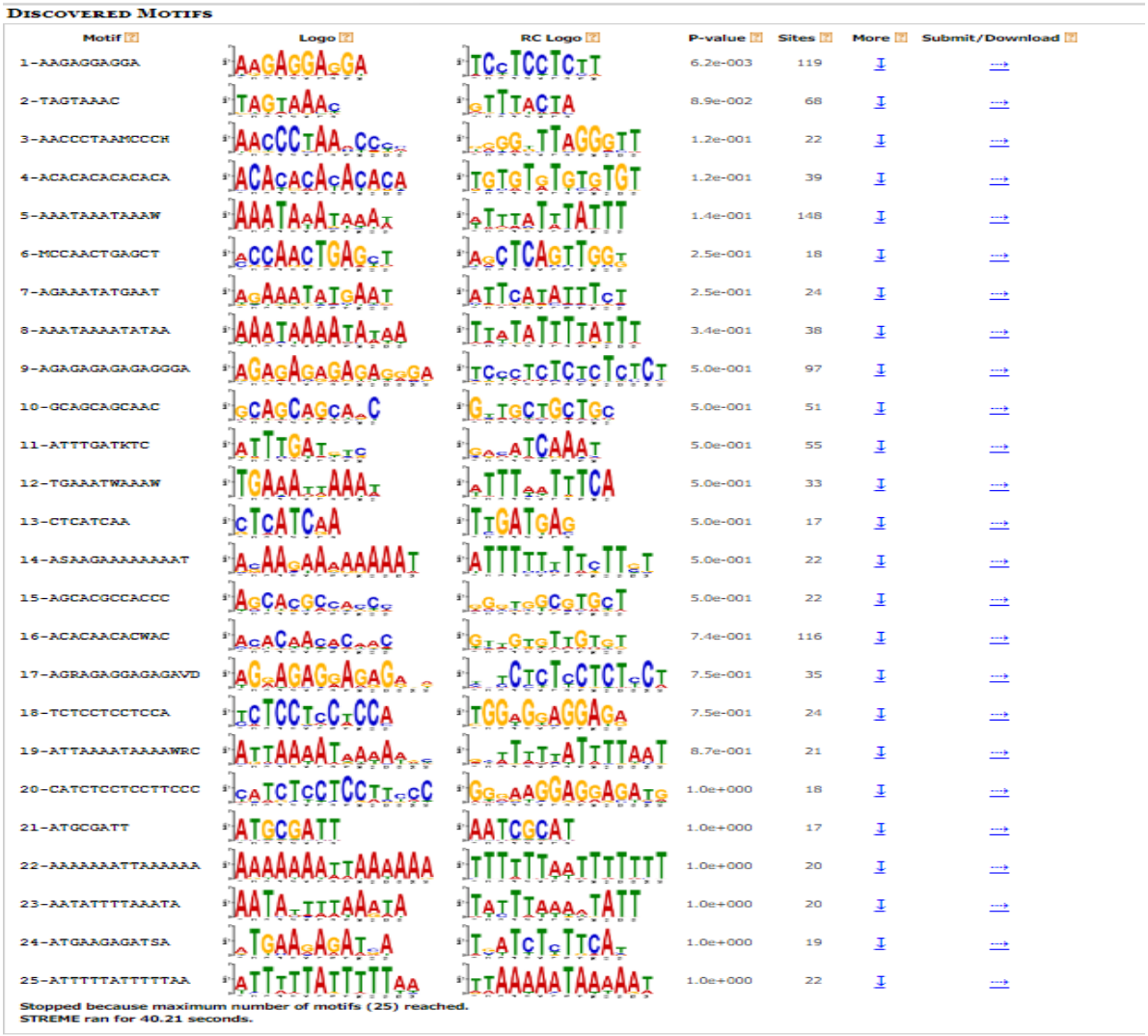


Figure 10. Top 25 most significant motifs at peak regions for *KfHD-like*. STREME (Simple, Thorough, Rapid, Enriched Motif Elicitation) analysis was performed on *KfHD-like* peak locations in an attempt to elucidate potential binding site preferences. However, the statistical significance of these motifs was low and many contain low-complexity repetitive sequences with little similarity to known binding site preferences of transcription factors from this family.

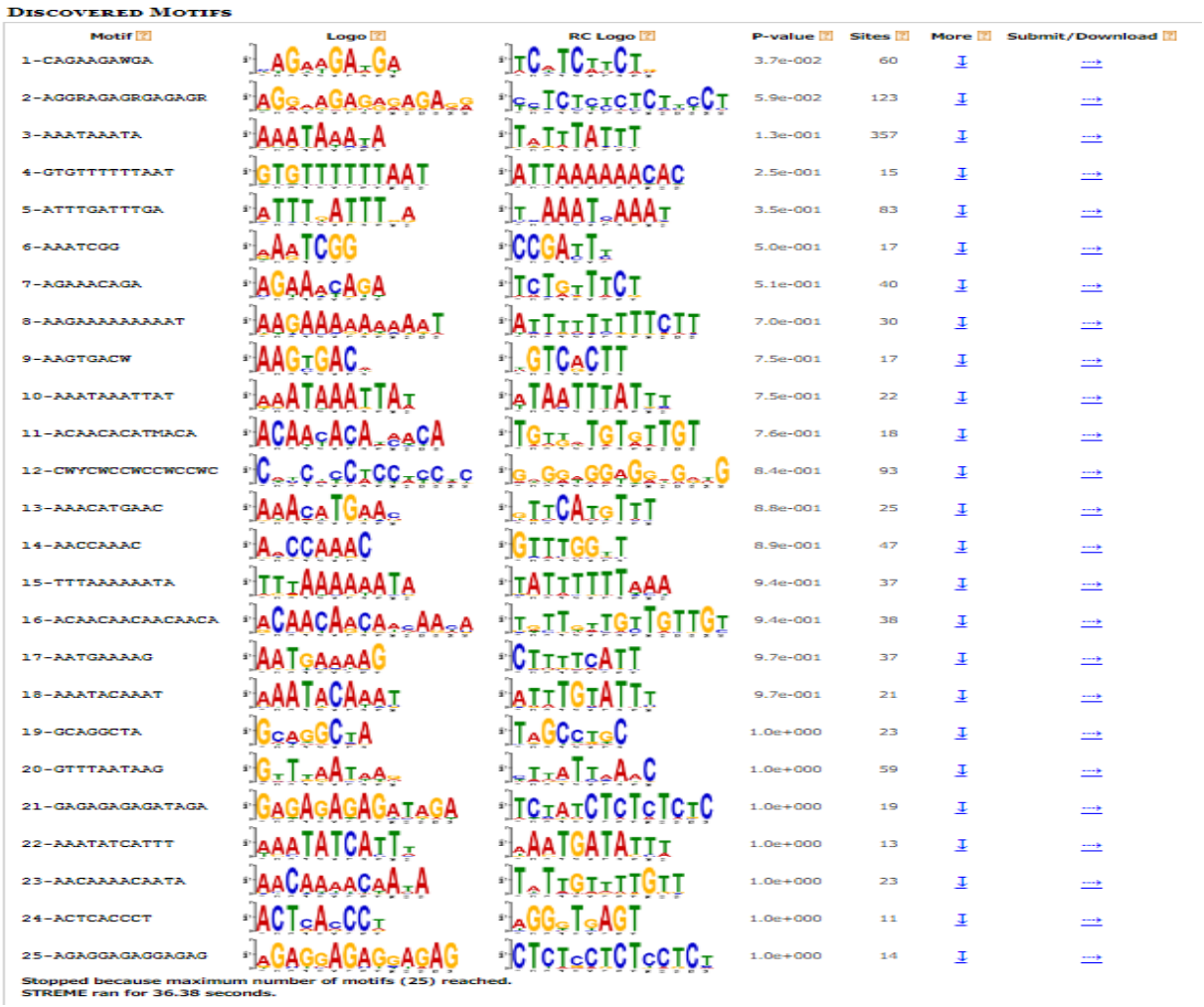
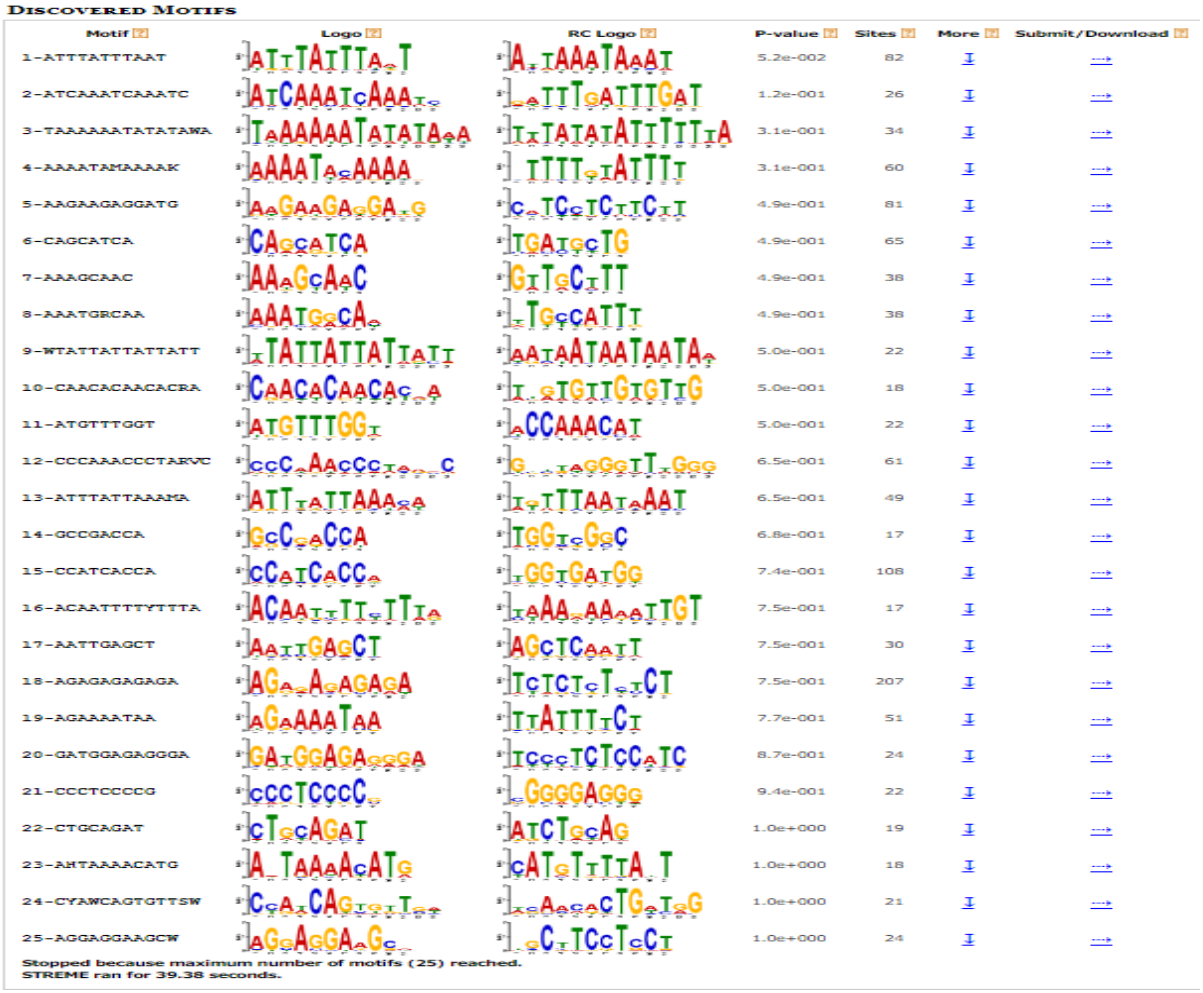


Figure 11. Top 25 most significant motifs at peak regions for *KfMYB59*. STREME (Simple, Thorough, Rapid, Enriched Motif Elicitation) analysis was performed on *KfMYB59* peak locations in an attempt to elucidate potential binding site preferences. However, the statistical significance of these motifs was low and many contain low-complexity repetitive sequences with little similarity to known binding site preferences of transcription factors from this family.



Supplemental Table 1. Primers used in this study.

Primer Name	Sequence 5' to 3'
KfNF-YB3 D-TOPO Fwd	CACCATGGCGGAGTCGGATAATGAG
KfNF-YB3 Rev	CTACCGGCCGGATCTTCCA
KfNF-YB3 Rev no stop codon	CCGGCCGGATCTTCCA
KfHD-like D-TOPO Fwd	CACCATGGATCTGTGTTCATCTGGG
KfHD-like Rev	TCAAGCTGAAGCCTCTCCTC
KfHD-like Rev no stop codon	AGCTGAAGCCTCTCCTCCTT
KfMYB59 D-TOPO Fwd	CACCATGGACAGAACAGGAAGACTT
KfMYB59 Rev	TCATGAGCCTCTTAACCTGG
KfMYB59 Rev no stop codon	TGAGCCTCTTAACCTGGATT
NTF Fwd	AATCATTCAGCGAAAACCAC
NTF Rev	TCGTTAAGTCCACCAGCC
BirA Fwd	TTTGTGTGTTTGCAGCTC
BirA Rev	CCCCCTGTTTGTCTATTCC
hpt Fwd	AGTACTTCTACACAGCCATC
hpt Rev	ATCTCGTGCTTTCAGCTTC
KfUBQ10 Fwd	ACACCATCGACAACGTCAAA
KfUBQ10 Rev	CACCTCAAGGGTGATGGTCT
3xHA tag Fwd	CCCATACGATGTTCCTGAC
3xHA tag Rev	CTGAGCAGCGTAATCTGGAA

Supplemental Table 2. Highly abundant (C_3/CAM mean > 100 FPKM) *Kalanchoe laxiflora* transcription factors displaying a log2 fold-change of at least 2 in CAM *vs.* C_3. Data was obtained through RNA-seq analysis of leaf pair 1 (C_3) and leaf pair 6 (CAM). Transcription factor descriptions are based upon the corresponding annotation of the closest *Arabidopsis thaliana* ortholog by nucleotide sequence homology of the full-length coding sequence (CDS). BLAST comparisons were made with TAIR BLAST 2.9.0+.

K. laxiflora transcript ID	Amino acid length	Arabidopsis gene ID	Arabidopsis description	Transcription factor family	C3/CAM mean FPKM	C3 to CAM log2 fold-change
kl_tr_010148t1	389	AT5G57660.1	CONSTANS-like 5	Zinc Finger Protein	2064.6	3.9
kl_tr_001634t1	849	AT5G62000.3	auxin response factor 2	Auxin Response Factor	2046	2.8
kl_tr_009804t1	397	AT1G53910.2	related to AP2 12	AP2/ERF	1639.9	2.2
kl_tr_009914t1	394	AT5G57660.1	CONSTANS-like 5	Zinc Finger Protein	1067.3	4.1
kl_tr_038313t1	167	AT4G14540.1	nuclear factor Y, subunit B3	NF-Y	883.7	4.5
kl_tr_002765t1	714	AT1G01060.4	Homeodomain-like superfamily protein	homeodomain-like superfamily	760.6	4.5
kl_tr_021707t1	170	AT4G14540.1	nuclear factor Y, subunit B3	NF-Y	743.6	3.5
kl_tr_005117t1	554	AT4G38900.3	Basic-leucine zipper (bZIP) transcription factor family protein	bZIP	726.4	2.2
kl_tr_015570t2	233	AT5G13180.1	NAC domain containing protein 83	NAC	705.3	8.8
kl_tr_016696t1	247	AT3G16770.1	ethylene-responsive element binding protein	AP2/ERF	611.3	5.5
kl_tr_013400t1	314	AT4G37260.1	myb domain protein 73	MYB	499.1	3.3
kl_tr_005117t2	554	AT4G38900.3	Basic-leucine zipper (bZIP) transcription factor family protein	bZIP	490.8	2.3
kl_tr_019584t1	200	AT3G30530.1	basic leucine-zipper 42	bZIP	457.4	2.2
kl_tr_038072t2	285	AT1G01720.1	NAC (No Apical Meristem) domain transcriptional regulator	NAC	271.6	2.1
kl_tr_018287t1	219	AT4G22950.1	AGAMOUS-like 19	MADS-box protein	216	2.2
kl_tr_016467t1	251	AT5G59780.2	myb domain protein 59	MYB	209.9	22.8
kl_tr_018748t1	212	AT1G30500.2	nuclear factor Y, subunit A7	NF-Y	198.8	2.1
kl_tr_020405t1	187	AT5G59780.2	myb domain protein 59	MYB	131.8	4.4
kl_tr_015570t1	268	AT5G13180.1	NAC domain containing protein 83	NAC	117.4	13.5

Supplemental Figure 1. Complete coding sequence and amino acid sequence of *K. fedtschenkoi NF-YB3*. The complete coding sequence of *KfNF-YB3* was used for construction of 3xHA and GFP fusion constructs described in this study. For construction of C-terminus fusions to 3xHA and GFP, final stop codons were omitted. Sequences were retrieved from JGI Phytozomev13 (https://phytozome-next.jgi.doe.gov/).

>K.fedtschenkoi v1.1|Kaladp0059s0310.1 CDS
ATGGCGGAGTCGGATAATGAGTCTGGGGGGCAGAGCAGCAACATGGCGACG
AGGGAGCAGGACAGGCTGCTGCCGATAGCCAACGTGAGCAGGATCATGAAG
AAGGCGCTGCCGGCGAACGCAAAGATATCGAAGGAGGCAAAGGAGACGGTG
CAGGAGTGCGTTTCGGAGTTCATAAGCTTCATAACGGGGGAGGCGTCGGAGA
AGTGCCAGAGGGAGAAGCGGAAGACGATCAACGGAGATGATCTGCTGTGGG
CGATGACCACGCTGGGGTTCGAGGACTACGTGGAGCCTCTGAAGATATACTT
GAGTAAGTTTAGGGAGATGGAGGGGGAGAAGATGAACCTGGGGAGAGTGCC
GCAGCAGGGTGATAGCGGAGAGTATGGAGGTGGTGGGTATGGCGGCGGCGG
CGGTGGGATGGGGTACGGTATGATGCAGCAGCAGCAGCAGCAGATGTATGA
CGGTGGAGGCTATGATCACGGGGCTGGTCGAAGAGTGTCCACTGGAAGATCC
GGCCGGTAG

>K.fedtschenkoi v1.1|Kaladp0059s0310.1.p
MAESDNESGGQSSNMATREQDRLLPIANVSRIMKKALPANAKISKEAKETVQEC
VSEFISFITGEASEKCQREKRKTINGDDLLWAMTTLGFEDYVEPLKIYLSKFREME
GEKMNLGRVPQQGDSGEYGGGGYGGGGGGMGYGMMQQQQQQMYDGGGYD
HGAGRRVSTGRSGR*

Supplemental Figure 2. Complete coding sequence and amino acid sequence of *K. fedtschenkoi HD-like*. The complete coding sequence of *KfHD-like* was used for construction of 3xHA and GFP fusion constructs described in this study. For construction of C-terminus fusions to 3xHA and GFP, final stop codons were omitted. Sequences were retrieved from JGI Phytozomev13 (https://phytozome-next.jgi.doe.gov/).

>K.fedtschenkoi v1.1|Kaladp0496s0018.1 CDS
ATGGATCTGTGTTCATCTGGGGAAGAGCTGGTTATCAAGTCAAGAAAACCTT
ATACAATTTCCAAACAGAGAGAGAGGTGGACAGAGGAGGAACATGGTAGAT
TTCTCGAGGCATTGAAACTCTATGGAAGAGCTTGGCAGCGAATTGAAGAACA
TATTGGAACTAAGACAGCTGTGCAGATCAGAAGTCATGCACAAAAGTTCTTT
TTAAAGCTGGAGAAGGATGCATCCGGCAAAGGGGTCCTGATCAGTGAAGTTC
ATCATATAGATATTCCACCTCCACGCCCTAAAAAGAAACCCAACAATCCTTA
TCCAAGGAAGATCAGTGTTGTTGTACCAACTCTGACAGTGGAAGCCAAAGAT
GGAAAAAAATCTATTCCAATTTCCTCATGCCCTGTTAAACAAGTACAAGAGC
CGCATAATGGAGTCCTTCAGGAGAAGCCGGGGGGGAAAAGAAACCCTTGATA
CTTCTTGCACTTCAGTTTCCCTTTCACATCGTGTAGAGAAGATGCACGAGGAC
GAACCAAATCAACACCCTGATCAAGATGGCACTGAACTTACAACTATGAGAA
AAGATTCGAGTGAGAACAGAAACTTGGAAATGTTGCAGGCTTCACCTGATAG
CGTTGTTCAATGTGTGAATTTGAAATGGGATAAATCACAGACATGTTCTACTC
AAGCCAGTACTGAAGCTACTCAGAGCTACCCTGGTAAAGTACCTGTAAATGC
TGTAGGAAGCAGCCCAGAACCTTTCAGACTAAGTTCACACCCTTTTTTTTACC
AATTTTTACCTGACCAGCTTGGTGGCACTGCTTTTAGCCCAAATGACTTCACA
AGTTCAGATGCATCAGCAACTAGAGAGCCCAAATGTCCAAGTGAAATGATTG
GTCCTTCGAATAGTCCCATCACTCCATTCTTTCTCAACCATGAAGATGCCTAT
AGGGCCTATCTTCAATTTTCTTCAGCTTTCTCAAATTTAGTTGTATCTACCTTG
CAGCAGAACCCAGCAGCCTATGCTGCCGCCGGTGTTGCTGCAGCTTCATTCTG
GCCAAATATTAACGCTGAAGCTCCTGCAGGATATTGTGGAGGGTTTCCAATG
AAAGTAATGGACACTGGTGGAGCTTCTCCGAGTATTACTGCAATCACTGCTG
CCACCGTGGCTGCAGCAACTGCATGGTGGGCTTCCAATGGATTGATTCCTTTC
TGTGCCTCTCAACAAAGTCTCACTCACAGCCAGAGCTGCTTTACCACAAACA
ATGTTGTTCCTCATGCAGCAGCAAGGACCAGGACTGTGACAGACGAGAAGAG
CCATCAGGATACAACTGTCAGTCAAGATCAGCAAGCAGAAGTGGTTAACTCT
CCAACCTTTCCTGGGAGAAACTCAGCTTTTAAGCCGCTTACTACAGTTCTAGA
GAAATCAGAAGAAAGTCAAGTTATGGTCATGGACAAACAAGTGCCTCCAACT
GCAGTGATCAGTAAAAATAAATCTGCAGCAATGGATGTTGAGATTCAAGGTT
CTAACACTGATTTGAAAGATGGAAAAAAGTTGGACCGTTCCTCTTCTGGTTCC
AATACAACCTCTAGCACAGAATTAGAAGAAACAGATGTTTTAGAGAAGCATG
AAAGTGGAAATGAGGAGTCAAATGCACCAGATGATGCACATTCTATTGGTGA
TTCTTCAAACCGACGTGGCAGAAATGTACCAAATGCCAACGAGTCCTGGAAA
GAGGTTTCGGAAGAGGGACGTGTGGCTTTTCAAGCACTTTTCTCCAGACAAG
TGCTGCCTCAAAAATTTTCATCATCGGCTGGTCTAGTTACTTTAGAGACTCAT

GCCAACACTGATCAAGAAAGCAAGCTGAAGCTGGGTGGAGTCGCATCACCC
ATCGACCTTAATAGTGATATACTTGTGTCATGCTCTGCTGATGGAGTTGTCAA
AACCAACGGAAAGGTAAGGGCTGAATGTGAAGGACAAGAAATACTACATTC
TATAGGTGTCGGAAAGCTGAGAACTCATAAAACTGGATTTAAGCCTTACAAG
CGAAGTTCGGTTGAGCCAAAGGACATCAGTGTGCCAAATTGTTGTAGCCAAG
GAAAAGACAAGAGGCTCAAGAGGATGTGCTTGGAAGGAGGAGAGGCTTCAG
CTTGA

>K.fedtschenkoi v1.1|Kaladp0496s0018.1.p
MDLCSSGEELVIKSRKPYTISKQRERWTEEEHGRFLEALKLYGRAWQRIEEHIGT
KTAVQIRSHAQKFFLKLEKDASGKGVLISEVHHIDIPPPRPKKKPNNPYPRKISVV
VPTLTVEAKDGKKSIPISSCPVKQVQEPHNGVLQEKPGGKETLDTSCTSVSLSHR
VEKMHEDEPNQHPDQDGTELTTMRKDSSENRNLEMLQASPDSVVQCVNLKWD
KSQTCSTQASTEATQSYPGKVPVNAVGSSPEPFRLSSHPFFYQFLPDQLGGTAFSP
NDFTSSDASATREPKCPSEMIGPSNSPITPFFLNHEDAYRAYLQFSSAFSNLVVSTL
QQNPAAYAAAGVAAASFWPNINAEAPAGYCGGFPMKVMDTGGASPSITAITAA
TVAAATAWWASNGLIPFCASQQSLTHSQSCFTTNNVVPHAAARTRTVTDEKSHQ
DTTVSQDQQAEVVNSPTFPGRNSAFKPLTTVLEKSEESQVMVMDKQVPPTAVIS
KNKSAAMDVEIQGSNTDLKDGKKLDRSSSGSNTTSSTELEETDVLEKHESGNEES
NAPDDAHSIGDSSNRRGRNVPNANESWKEVSEEGRVAFQALFSRQVLPQKFSSS
AGLVTLETHANTDQESKLKLGGVASPIDLNSDILVSCSADGVVKTNGKVRAECE
GQEILHSIGVGKLRTHKTGFKPYKRSSVEPKDISVPNCCSQGKDKRLKRMCLEGG
EASA*

Supplemental Figure 3. Complete coding sequence and amino acid sequence of *K. fedtschenkoi MYB59*. The complete coding sequence of *KfMYB59* was used for construction of 3xHA and GFP fusion constructs described in this study. For construction of C-terminus fusions to 3xHA and GFP, final stop codons were omitted. Sequences were retrieved from JGI Phytozomev13 (https://phytozome-next.jgi.doe.gov/).

>K.fedtschenkoi v1.1|Kaladp0033s0028.1 CDS
ATGAACCAAAGGCTGCATGCATCATTCTTTTGGGTTTGGAAGACAGGTGATA
TCAATAATTGGGAAGAACAAAGCAAAGACAGGCGAGGAAGCCAATATGATG
CAACAAGATACCAGAAAGGGTCCATGGACAGAACAAGAAGACTTTCTGCTG
GCAAATGTTGTAGGCTTATTTGGAGACCGCCGATGGGATTTTGTAGCAAAAG
TTTCAGGTTTGAAGGTGGCGGGAGACAAGTTCTAGGTTTAAACAGAACAGGA
AAGAGTTGCAGGTTACGTTGGCTTAATTATCTTCACCCTGGTCTCAAAAGAGG
AAAGATGACATCCACAGAGGAAAGGATTGTGCTTGAACTTCATGCGAGGTGG
GGAAACAGATGGTCGAGAATTGCTAGGAGATTACCAGGACGTACAGATAAT
GAGATCAAGAACTATTGGAGAACCCACATGAGGAAAAAAGCTCAAGAGAGA
AAGCGGGTTAATTATCTTTCAATCACATCTTCTGAATCTTACTCATCATCAAG
CAACCATACACAAGACTCTCTAAGCTGTAGCGGAAAAGAAAGCTTTTATGAT
ACAGGGGGGCTTGCAGCTATTGCAGTGACAGGAAAGAAGCAAAAAGAGGAA
GAAGGTGAACGGATAAAAATGGATGAAATATGGAATGATATAGACATATCA
AAAGACACAATCACCCCACTGAATGATAGTTACAACCAAGGGTTATGCACTT
ACTCTTGTGCTCCCTTAGTTTCACCAATATGGGAGTACTGCCCAGACCCGCTA
TGGAGGATGGATGAAGATGATAGCAAGATGCCACCCCCTTCAAACGACTATT
TCCAAACTTGTTGTGAAGCAGGGACATCTACTCTAAAAGGCCCATGA

>K.fedtschenkoi v1.1|Kaladp0033s0028.1.p
MNQRLHASFFWVVWKTGDINNWEEQSKDRRGSQYDATRYQKGSMDRTRRLSAG
KCCRLIWRPPMGFCSKSFRFEGGGRQVLGLNRTGKSCRLRWLNYLHPGLKRGK
MTSTEERIVLELHARWGNRWSRIARRLPGRTDNEIKNYWRTHMRKKAQERKRV
NYLSITSSESYSSSSNHTQDSLSCSGKESFYDTGGLAAIAVTGKKQKEEEGERIKM
DEIWNDIDISKDTITPLNDSYNQGLCTYSCAPLVSPIWEYCPDPLWRMDEDDSKM
PPPSNDYFQTCCEAGTSTLKGP*

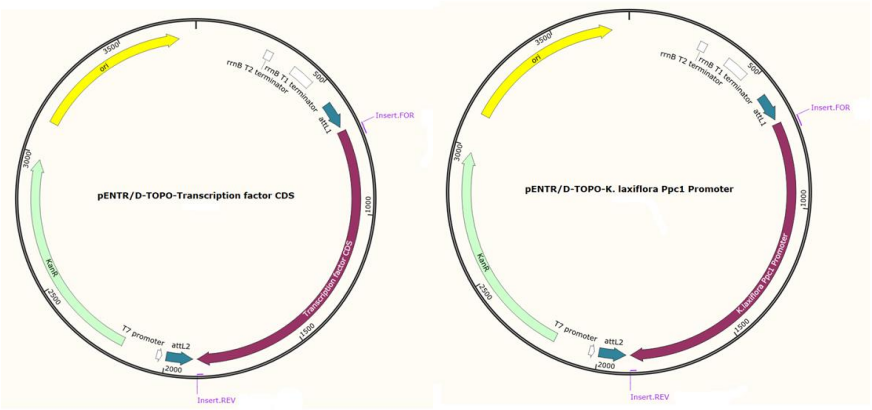
Supplemental Figure 4. Proximal *K. laxiflora Ppc1* promoter region used to drive expression of the NTF peptide in the *Kl*p*Ppc*-INTACT construct. The sequence below corresponds to the region directly upstream of the transcriptional start site of the CAM-type *K. laxiflora Ppc* gene (*Ppc1*) (Kalax.0018s0056.1) and was retrieved from JGI Phytozomev13 (https://phytozome-next.jgi.doe.gov/).

>Kalax.0018s0056.1 (*KlPpc1*)
GTTTGGCAGATGAGAACCGGGCCTGCAACTGTTAACTGATCAGTGTGGCAAA
GGCAGTTTGCAGGTTTGGAACCTCTACAGGGTCTTCTTCCTTATACCCTACCT
CTGGAGGTTTTTTCATATTGAATGGCGACCCTGGTCAATTATATAATATATGC
TAGACAAAAGCCTGAGATGATAAGAGCTCTCAGGAAGAAGCGATAAGCTTTC
ACTGTCTTAGAAATTGAAATTGAAATTGACTCTGCCCCACTATTCCCTCCGAC
ATCTCCATGATACTCGCTAAATTATTGAACTGGGAGACGCACCTGTGTCCATG
TTTCTTCCACAAAAATGGGATTTTTATCCATTTGGATTATTCATGACAACTTA
GTCGTAGAGAGTGGACCACAATTCGCCAAGGACCCGTTGTCAGTACTCATCG
ACGGCCACAATCCGAGTGGGATCGTGACAAGGGACGTGCTGAATACGTGTTG
CCACCTGGAGACACTTGCACGGTATGAGTCGAGGATGGAGCAACTTGACTCT
TTGCCTACTGTAAGAGGCCTCTTGTTCTTTTCCTCTCATTGGGTCTATGGTTCT
TTTCCACTAGGGATCAGGTGTTCGAAAAACAGTGATGCTGTCCCCACTGATGT
CTATTCAGAGCTAGGAGACAGACAGCAACAAAACGGGTACTGTTGTTTTTGC
ATGGTTTATGAACCAATCTTACAACTCAATCATGACATGCAGCTACAAAGGT
GAGAAATGATTGATTGATTTCGAGCTTCAATTATCTGCATCTGATAAGGATCA
GTTTAAGATTGTTATCAGCCTTGTGGCGCTCGTCTTATTGATAATTAGAACTC
CCAAGTGAGTTCTTAGCAAGACAGCTTTGGGTTTACGGCATTATTGCTCTTCA
AGTGAGGTACCCACCCTATGTTTTTAGAATTTAGTAACACAGCTCCTAGTAGG
ACAACATAATGAAGACACCAACAAGGAATCACAACGGAATATTCCCTCCTTT
CAATACTCACCCAACACACTAATGCTTGGACTTCTAACTCGGCCAGAAGTTTT
ATTAAAATTGGACCGGTCAAATCAGTCTGACGTCTGGTTCTCAATTCATCGGT
TAGATTGATTGGTTTAAACTTGTTTGATCACCAATCCAATCAGGTGGATAATC
TATATATATAACATAATAGGATGTCCGCCGCTCCCTGCGGTTCGGTTTGGTTC
GGTCCGGTTCAGTTTGATTCGGTTTGGTTCGGTTTGATTCAGTTCGGTTTCAAA
TTTTAAAAAATTATTTGGTTTACAGTTTTATATTTTAAAAAGAGATGACTAAT
AATTTGGTTCAATTTTTAATTTTTTTCGAACCAAACCTTCGGGACGACAATTG
CCTTAATTGTTCGGTTTGCAGTTTCATATTTTAAAAAGGGATTAATAATTTGG
TTTGATTCTTAATTTTTTTCGAACCGAACCATCGGGACGACGGATGCCTTTAC
TCGTTATAAATATTTGATTATAATTTTTTTAGGTAACGTGGTGGAGAAGTCTT
GTTTATCACTAACTTTTATTATACTGGTTATTAATTATTAAAATAAGTCTTTAA
TACTTAATTAATCACTTAATATGTGTAATAAATCAGTTTATAATACAACCGAC
CAGTTTGGGCGAATTTTCTATGAACTGGATGGTTCGGTCCGGTTACACCTAGT
TTGATGTGCAGACCGATTAATACACGTAAGTCAGACCAGAACCAAATCCGAT
TCCCAACTAATTGAAATTGGTCGGCTAGACCGATTGACCATTGATCTAATTAG
ATAGATAATGTAAATTCTCATATCGTAAATACTAAATAATAATAACTTTTTCG
GTAGTGTAGTTAAAAAAGTCCTTGCTGGGACAAGCCAGCTCGCACTCTCTGG
TTAATTTTTTTACTACTTTTTCTTTAATCAATTAGTTAAAATGTTTTAATGCTTA
ATTAATCAGATAATGCGTATAACAAACCCGTTTATATTGAAATCGACCGGTGT

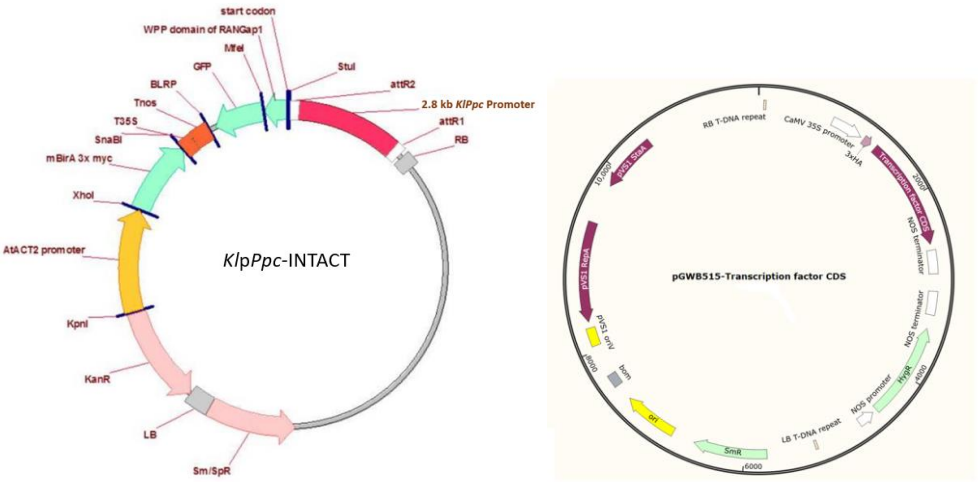
GAGTGATTTTTCTTTGAATTGGCTTGTTCAATCCGATTATATCCGATTTGATAT
GCAGATCATATATGTAAATCATATCAAAAGCAAGTCTGATTCCTGTTCCAACC
GATTAAACCAATTAATCTGATTTGTATAATTATGTGTAAGAGTATTCTCAGGT
GACATGCTCTACGTGTACCATGTATCAACTATTGAGCAATTTTAACAAACCAT
AACAATAGTAACATTTTGAAAAATAAAAAATATTAGGAATCTAATCTCATGC
CAAAGAGATTATTTCTCTCATAATTAATTAATAGTTTATACGTGTCCCAATAA
ATGTGAATACTAATAAAAATATTAAATATTTTTCAGATGTTAGATCGATATTA
ATCATAAAAATAGTAACTTTTCTCGCAACAAATCGGATAAGATATGAGTCCA
AAAAATTATTAATGTCCACCCAAAATTCCCTGTTATGCGTAACACTGTACTTG
CTTTAATGACTGAAAATCACACTCGCCCACACGTAATCAACCTAGCAAAATC
CTCATTAGCTTACCTAAACAAGCAGGCTCTCATCTTGCAGTGCTTCGGATATT
TACTTCTGGTAAAAAGCTTGCATCTTTGATCAACAATGTCATCTGGGTCATTG
TCAAAATCTCAATCCAATCGGTCGGGAGCCTATCCTCTATTATGGGCTGACAT
GTGTCGTTTGATTATTGCCTTCTCTCAGTCTCCATAGCTTATTTATTGGGTCAT
TGTGTAAGCTTGAGGTCATTGTGCTCTTGGGTTGAGTCTAATCCCAAGGTTGA
GAAATACTGTTAGGGGTCAAAATGCAG

Supplemental Figure 5. Plasmid constructs used in this study. The proximal 2.8 kb *K.laxiflora Ppc1* promoter and transcription factor coding sequences were cloned into the Thermo Fisher Scientific pENTR/D-TOPO vector to generate Gateway-compatible entry constructs (**a**, **b**). Gateway LR cloning was then conducted to generate plant expression constructs for the INTACT approach (**c**) and to produce transcription factor N and C-terminus 3xHA (**d**, **e**) and GFP fusions (**f**, **g**). Transcription factor stop codons were omitted for construction of C-terminus fusions. Transcription factor plasmids illustrated below are not necessarily to scale as a generic insert is used to illustrate the particular fusion produced by each expression construct.

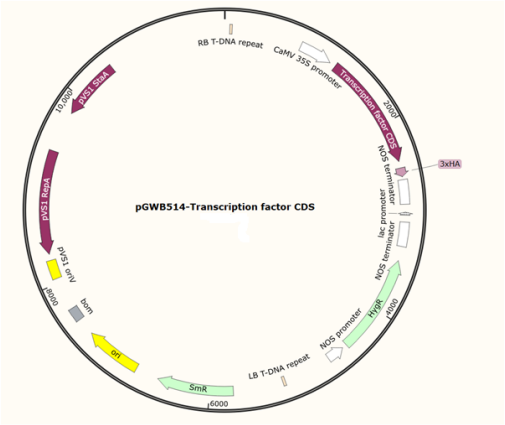
a **b**



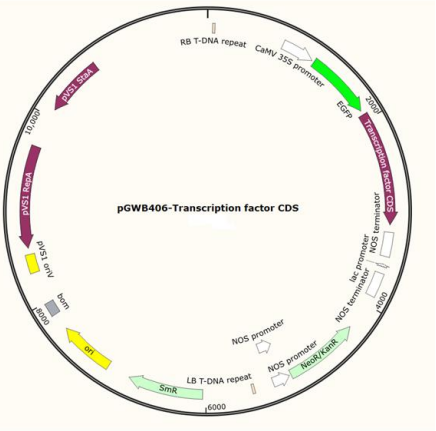
c **d**



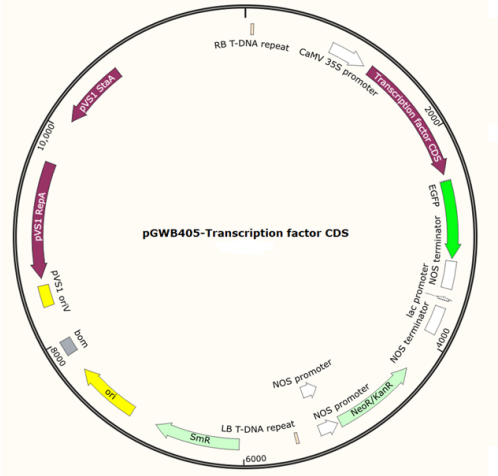
e


pGWB514-Transcription factor CDS
RB T-DNA repeat
CaMV 35S promoter
Transcription factor CDS
3xHA
NOS terminator
lac promoter
NOS terminator
hygR
NOS promoter
LB T-DNA repeat
SmR
ori
bom
 Avi 1Sid
pVS1 RepA
pVS1 StbA
2000
4000
6000
8000
10,000

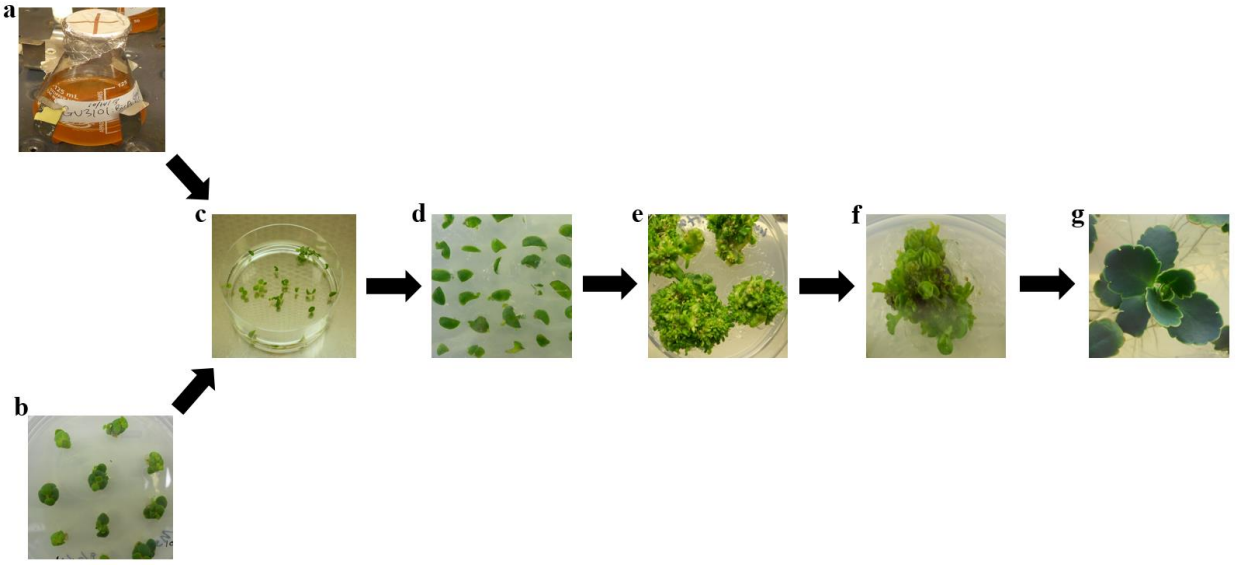
f


pGWB406-Transcription factor CDS
RB T-DNA repeat
CaMV 35S promoter
EGFP
Transcription factor CDS
NOS terminator
lac promoter
NOS terminator
NOS promoter
NOS promoter
NeoR/KanR
LB T-DNA repeat
SmR
ori
bom
pVS1 oriV
pVS1 RepA
pVS1 StbA
2000
4000
6000
8000
10,000

g

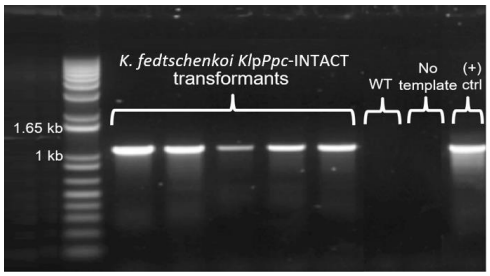

pGWB405-Transcription factor CDS
RB T-DNA repeat
CaMV 35S promoter
Transcription factor CDS
EGFP
NOS terminator
lac promoter
NOS terminator
NOS promoter
NOS promoter
NeoR/KanR
LB T-DNA repeat
SmR
ori
bom
pVS1 oriV
pVS1 RepA
pVS1 StbA
2000
4000
6000
8000
10,000

Supplemental Figure 6. *Agrobacterium*-mediated stable transformation protocol for *K. fedtschenkoi*. *Agrobacterium tumefaciens* strain GV3101 was used to introduce the *Klp*Ppc-INTACT construct and 3xHA epitope-tagged transcription factor constructs stably into *Kalanchoe fedtschenkoi* for ChIP-seq studies. Small leaves taken from young *K. fedtschenkoi* plantlets originally started from leaf explants were sliced in half and incubated in *Agrobacterium* suspensions for 1 h (**a**, **b**, **c**). Leaves were then placed on co-cultivation media for 4 d. After co-cultivation, explants were then transferred to regeneration media (**d**) until shoot initials of approximately 1 to 2 mm in length became apparent (**e**). At this stage explants were transferred to shoot induction media. Once shoots reached approximately 8 to 10 mm in height and exhibited some signs of forming leaf structures (**f**), they were excised from the underlying mother explant and transferred to rooting media. Plants which survived and formed roots on rooting media (**g**) were considered putative stable transformants pending PCR confirmation.

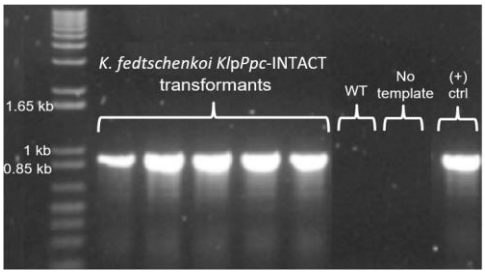


Supplemental Figure 7. Validation of *K. fedtschenkoi* stable transformants and semi-quantitative expression analysis of TF-3xHA fusion transgenic lines. PCR screening was conducted to confirm the presence of *NTF* (**a**) and *BirA* (**b**) in *Klp**Ppc*-INTACT-transformed *K. fedtschenkoi* lines. Confirmation of expression of the nuclear envelope-localized NTF peptide was confirmed by fluorescence microscopy (figure 4). Stable transformation with TF-3xHA fusion constructs was indicated by the PCR amplification of the *hpt* hygromycin resistance gene (**c**). Plasmid DNA samples were utilized as positive controls for PCR screens. Semi-quantitative RT-PCR was then conducted (**d**) using primers targeting the 3xHA tag and the constitutively expressed *KfUBQ10* transcript. PCR was performed using 30 cycles as the expression of *KfUBQ10* was consistently saturated at this cycle number but clear differences in the expression levels of the 3xHA tag among the different independent transgenic lines were apparent. Transgenic lines (3 for each construct) exhibiting an intermediate expression of each TF-3xHA fusion by visual inspection of the amplified products were selected for further experiments except in those cases where only 3 independent transformed lines in total were recovered (pGWB514-*KfNF-YB3*-3xHA and empty vector control). Primer sequences utilized in the screens depicted in this figure are provided in Supplemental Table 1.

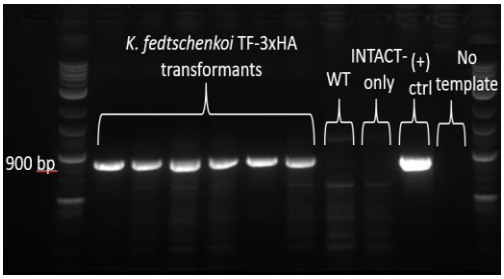
a



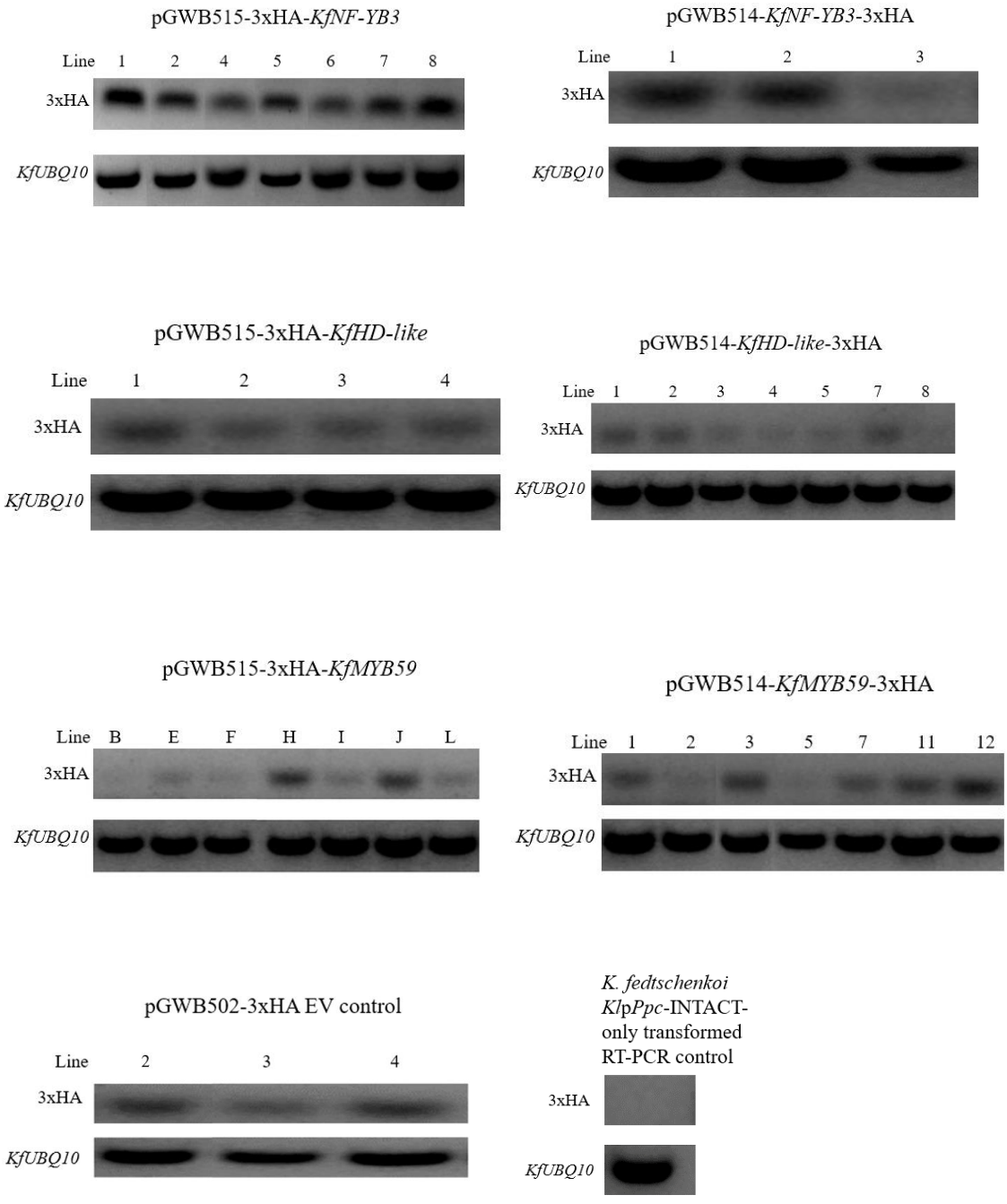
b



c



d


pGWB515-3xHA-*KfNF-YB3*
Line 1 2 4 5 6 7 8
3xHA
KfUBQ10
pGWB514-*KfNF-YB3*-3xHA
Line 1 2 3
3xHA
KfUBQ10
pGWB515-3xHA-*KfHD-like*
Line 1 2 3 4
3xHA
KfUBQ10
pGWB514-*KfHD-like*-3xHA
Line 1 2 3 4 5 7 8
3xHA
KfUBQ10
pGWB515-3xHA-*KfMYB59*
Line B E F H I J L
3xHA
KfUBQ10
pGWB514-*KfMYB59*-3xHA
Line 1 2 3 5 7 11 12
3xHA
KfUBQ10
pGWB502-3xHA EV control
Line 2 3 4
3xHA
KfUBQ10
K. fedtschenkoi
*Kl*p*Ppc*-INTACT-
only transformed
RT-PCR control
3xHA
KfUBQ10

Chapter 3. Identification of putative transcriptional regulators using the *Kalanchoe* phospho*enol*pyruvate carboxylase 1 (*Ppc1*) promoter

Authors:

Travis M. Garcia[1], Jungmin Ha[2], Won C. Yim[1], John C. Cushman[1]

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557-0330 USA

[2]Department of Plant Science, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826 South Korea

Abstract

In view of the alarming trends in global warming and the threat this poses to sustainable agricultural production, the biological mechanisms underlying plant water-use efficiency have emerged as an important area of research. Owing to their unique ability to temporally separate carboxylation and decarboxylation plant species that perform crassulacean acid metabolism (CAM) photosynthesis have evolved to possess a high degree of water-use efficiency. Following the initial conversion of CO_2 to HCO_3- by carbonic anhydrase, p*hospho**enol*pyruvate carboxylase (PEPC) catalyzes the nocturnal assimilation of carbon into oxaloacetate (OAA), followed by malate dehydrogenase, which converts the OAA to malate. Despite PEPC's importance as one of the core components of the CAM carboxylation module, the complex regulatory mechanisms underlying its expression are not well-studied. We examined the *Ppc1* promoter from the model CAM species *Kalanchoe* and discovered putative *cis*-regulatory elements similar to sequence motifs previously reported to be associated with evening-phased expression and mesophyll tissue-specificity. Additionally, several GATA box-like motifs and putative enhancer sequences were also identified. These individual elements along with five overlapping 200-bp segments of the proximal 880 bp region of the *Ppc1* promoter were used as bait in yeast one-hybrid (Y1H) screens against a *Kalanchoe* cDNA library to identify *trans*-acting factors for further study. From this analysis, we identified the transcription factors *ERF9*, *ERF106, TCP4*, and *PIF1*. In addition to a high-confidence of interaction with the *Kalanchoe Ppc1* promoter from our Y1H analysis, *in-silico* examination of the *Ppc1* promoter revealed likely binding sites for these factors based on homology to validated preferred binding sequences in *Arabidopsis*. These candidate

Ppc1 transcriptional regulators are currently the focus of additional experiments to confirm their interaction with the *Kalanchoe Ppc1* promoter and elucidate the nature of their regulatory effects.

Introduction

Climate change-driven agricultural challenges along with expanding desertification in many regions of the world underscores the increasingly pressing need to understand the molecular basis of plant water-use efficiency. The photosynthetic pathway of crassulacean acid metabolism (CAM) is a key evolutionary adaptation of plants to water-limited environments and is now present in at least 343 plant genera across 35 families (Silvera et al., 2010; Edwards, 2019). An essential defining feature of CAM is the nocturnal opening of stomata and concurrent recruitment of photosynthetic phospho*enol*pyruvate carboxylase (PEPC) to catalyze the primary fixation of CO_2, a process which minimizes water loss due to transpiration. Decarboxylation and fixation of CO_2 by Rubisco then occurs during the next photoperiod behind closed stomata. In this manner, CAM achieves a water-use efficiency several fold higher than C_3 or C_4 photosynthesis (Borland et al., 2011; Cushman & Borland, 2002).

In vascular plants, *Ppc* genes are encoded by a multigene family and are associated with a variety of both photosynthetic and non-photosynthetic roles. "Housekeeping" forms of *Ppc* are involved in the replenishment of Krebs cycle enzymes, whereas other non-photosynthetic forms of *Ppc* have been found to have roles in nutrient absorption and transport, fruit maturation, germination, maintenance of cellular pH, and

stomatal control (Ping et al., 2018; Silvera et al., 2014). The available molecular evidence suggests that CAM-specific isoforms of *Ppc* are not unique to these species, but rather originated from non-CAM isoforms as a result of gene duplication events followed by divergences in amino acid sequence and the acquisition of novel *cis*-regulatory transcriptional control elements (Silvera et al., 2014). The transcriptional control of photosynthetic *Ppc* genes along with the magnitude of their expression appear to be important determinants of the genotypic capacity of a plant for CAM and due to the central role of this enzyme in CAM photosynthesis a detailed understanding of its regulation may have important bioengineering implications (Deng et al., 2016; Taybi et al., 2004). Indeed, the overexpression of an *Agave* CAM-specific *Ppc* isoform in tobacco resulted in improved abiotic stress tolerance, enhanced biomass productivity, and the increased expression of CAM gene orthologs in transgenic lines (Liu et al., 2021). In this study, we seek to improve understanding of the *cis*-regulatory basis of the spatiotemporal expression patterns of the CAM photosynthetic *Ppc1* gene through a detailed examination of promoter elements in the developmental CAM model *Kalanchoe* using Y1H experiments combined with *in silico* modeling.

Previous research into the temporal phasing of CAM metabolic genes has established a clear link to the core circadian clock (Chen et al., 2020; Ming et al., 2015; Zhang et al., 2020). In addition, CAM-type *Ppc1* not only appears to be circadian-regulated, but also functions in the maintenance of normal expression patterns of core circadian clock and stomatal guard cell-associated genes (Boxall et al., 2020).

Detailed studies in *Arabidopsis* on time-of-day transcriptional regulatory modules has revealed the presence of a particular promoter element, the evening element (EE),

which is both necessary and sufficient to confer evening-specific rhythmicity. This regulatory element, comprised of the sequence AAAATATCT, possesses a GATA box-like TATC core and also bears strong similarity to the CIRCADIAN CLOCK ASSOCIATED 1 (*CCA1*) binding site (Michael & McClung, 2002; Michael et al., 2008). Inspection of the *Kalanchoe* CAM photosynthetic *Ppc1* promoter indicates the presence of several motifs with similar sequences to the evening element and GATA box.

Cis-regulatory modules conferring mesophyll tissue specificity have also been reported. As in the case of CAM, photosynthetic *Ppc1* expression in C_4-performing species is also restricted to mesophyll cells. Comparative *Ppc* gene promoter analysis of closely related C_3 and C_4 *Flaveria* species by Akyildiz et al. (2007) identified the presence of a *cis*-regulatory module necessary to confer mesophyll-specific expression of *Ppc* genes in the C_4-performing *Flaveria trinervia*. This module, termed mesophyll expression module1 (MEM1), is itself composed of an 11-bp A submodule and a 30-bp B submodule. A G-to-A exchange at the start of the A submodule and the insertion of the tetranucleotide sequence CACT within the B submodule appear to be responsible for the *Ppc* genes mesophyll specificity conferred by MEM1 in C_4 *Flaveria* species in contrast to C_3 MEM1 variants (Akyildiz et al., 2007). In *Kalanchoe*, both MEM1a-like and MEM1b-like sequences were identified within the *Ppc* gene promoter and were selected for inclusion in this study.

Our inspection of the *Kalanchoe Ppc1* promoter also revealed the presence of several overrepresented enhancer-like motifs. These putative regulatory elements, detected through a MEME (Bailey et al., 2015) analysis, were concatenated and along with evening element, GATA, MEM1a, and MEM1b-like sequences used to design bait

probes for Y1H screens against a *Kalanchoe* cDNA library. The results of these screens in conjunction with five additional Y1H screens of discrete sections of the proximal *Kalanchoe Ppc1* promoter identified several specific transcription factors with a high-confidence of interaction with these probes. Subsequent detailed *in-silico* analysis of the *Kalanchoe Ppc1* promoter using tools provided by the MEME suite (Bailey et al., 2015) combined with leveraging the available binding data on the *Arabidopsis* orthologs of the transcription factors identified in our screens indicated *Kalanchoe ERF9, ERF106, TCP4*, and *PIF1* as particularly likely direct regulators of the *Ppc1* promoter.

As CAM *Ppc* isoforms have been reported to be induced by abiotic stress (Cushman & Bohnert, 1992; Schaeffer et al., 1995) and exhibit circadian-phase dependent expression patterns (Zhang et al., 2020), direct interactions between promoter *cis*-elements and *ERF9, ERF106, TCP4*, and *PIF1* would be consistent with the reported roles of these transcription factors as reported in *Arabidopsis*. *AtERF9* was previously reported to be upregulated in response to dehydration stress in *Arabidopsis* in which it was elucidated as one of the "non-memory" transcription factors induced in response to this stress (Ding et al., 2013). More recently, *AtERF9* was found to be one of the central "hub" components of osmotic stress-induced transcriptome changes (Van den Broek et al., 2017). Analysis of transcriptional networks in developing *Arabidopsis* seedlings in response to light identified *AtERF106* as one of the transcription factors induced by *phyA* (Mazzella et al., 2005). Additional findings on circadian-associated genes in *Arabidopsis* revealed *AtERF106* to be a particularly short-lived unstable transcript, a feature which appears to be associated with transcription factors with critical roles in circadian-directed processes (Gutierrez et al., 2002).

TCP4 has been linked previously to photoperiod-associated transcriptional networks. Y1H analysis of the *Arabidopsis CONSTANS* promoter confirmed the binding of this transcription factor to predicted TCP binding elements. Additional experiments in this study revealed that *AtTCP4* acts as a transcriptional activator of *CONSTANS* at dusk (Kubota et al., 2017). *AtTCP4* has also been found to act as a transcriptional activator of the auxin biosynthetic gene *YUCCA5* and to integrate auxin signaling with brassinosteroid transcriptional networks to promote cell expansion in response to environmental cues such as light (Challa et al., 2016).

PIF1 as one of the members of the phytochrome interacting factors within the bHLH superfamily has well-established roles in photomorphogenesis. Primarily acting as a negative regulator of *phyA* and *phyB*-associated processes, PIF1 promotes gravitropism and is also known to inhibit excess protochlorophyllide accumulation prior to light activation (Duek and Fankhauser, 2005). PIF1 also functions as an important component of seed dormancy by activating ABA biosynthesis while repressing GA biosynthesis in the absence of light through direct targeting of GA biosynthetic gene promoter elements (Castillon et al., 2007).

In summary, the work presented in this study highlights several transcription factors that will likely prove to be important direct regulators of CAM photosynthetic *Ppc1* through *in silico* promoter analysis and Y1H experiments utilizing the model CAM species *Kalanchoe laxiflora* and *Kalanchoe fedtschenkoi*.

Materials and Methods

Plant material, maintenance, and growth conditions

Kalanchoe fedtschenkoi diploid M2 accession plants and *Kalanchoe laxiflora* accession 1982-6028 were sterilized by washing for 30 seconds in 70% v/v ethanol followed by a 10-minute wash in 10% v/v NaOCl containing 1% v/v Tween 20. Plants were then washed 5 times in sterile water and transferred to 1X Murashige and Skoog media with Gamborg vitamins, 30 g/L sucrose, 0.7% Plant TC agar, pH 5.8 (MS30 media). All plants used for the experiments described in this chapter unless otherwise stated were maintained on this media in a growth chamber under a 16 h light 8 h dark schedule at a constant 20 °C with a photon flux density of approximately 280 μmol m^{-2} s^{-1}.

Kalanchoe Ppc1 promoter analysis and design of yeast one-hybrid bait probes

Sequence analysis was performed on the 2,824 bp proximal promoter region of the CAM-associated *Kalanchoe laxiflora Ppc1* gene (JGI Phytozome accession number Kalax.0018s0056.1). This *Kalanchoe laxiflora* transcript represents the closest ortholog to the previously characterized CAM-associated *Kalanchoe fedtschenkoi Ppc1* gene (JGI Phytozome accession number Kaladp0095s0055.1). Inspection of the above described *K. laxiflora Ppc1* promoter identified a single occurrence of three of the eleven variants of the evening element (EE) as described by Michael et al. (2008). These identified EEs (GATATT, AAGATAT, and GGATATTTA) along with 2 to 3 bp 5’ and 3’ flanking sequences were placed in tandem (Supplemental Table 1). This construct was then repeated three times to synthesize the full-length EE probe for Y1H assays. Six GATA

box sequences (Michael et al., 2008) were discovered in the *K. laxiflora Ppc1* promoter. These elements, each composed of the sequence GGATA, along with 5 bp 5' and 3' flanking sequences were placed in tandem and repeated three times to synthesize the full-length GATA box Y1H probe (Supplemental Table 1). As previously described, the MEM1 module consists of two submodules, MEM1a followed by MEM1b, which act together to confer mesophyll specificity. Six pairs of these elements were found within the *K. laxiflora Ppc1* promoter. Each individual submodule hereafter referred to as G-mesophyll (MEM1a) and CACT-mesophyll (MEM1b) were used to design separate probes (Supplemental Table 1). Each probe was synthesized in similar fashion to those previously described. Fifteen overrepresented enhancer-like elements identified with MEME (Bailey et al., 2015) were identified in the *K. laxiflora Ppc1* promoter. These sequences including with 2 bp 5' and 3' flanking sequences were concatenated, but not repeated to construct the full-length Y1H probe (Supplemental Table 1). All bait probes described above were synthesized with 5' and 3' *SfiI* restriction sites for integration into the pJ204 vector and were constructed by DNA2.0 Gene Design & Synthesis (Newark, CA). For construction of the *Kalanchoe fedtschenkoi* promoter region bait probes total genomic DNA was isolated from *K. fedtschenkoi* leaf tissue using the Qiagen DNeasy Plant Mini Kit according to product instructions. The 880 bp upstream region directly proximal of the *K. fedtschenkoi Ppc1* (Kaladp0095s0055.1) transcriptional start site was amplified as a series of 5 separate 200 bp fragments with 29 bp overlap between each sequential segment. Each 200 bp fragment was then used as a separate probe in Y1H assays (Supplemental Table 1). Sequences of all ten bait probes described in this section

are presented in Supplemental Table 1. An alignment of the highly-conserved 3 kb *K. laxiflora and K. fedtschenkoi Ppc1* promoter is provided in Supplemental Figure 1.

Yeast one-hybrid experiments

Total RNA was isolated from frozen ground leaf tissue from tissue culture-grown *Kalanchoe fedtschenkoi* plants using the Qiagen RNeasy Plant Mini Kit following the manufacturer's instructions with the addition of 9 mg polyethylene glycol 20,000 (PEG-20,000) to the 450 µL of RLT buffer used for extraction. On-column DNA digestion was performed according to RNeasy kit instructions with the Qiagen RNase-Free DNase Set. Final RNA elution was performed with RNase-free water. RNA purity and approximate quantity was assessed with a Thermo Fisher Scientific NanoDrop 2000c spectrophotometer and precise quantity assessed using Thermo Fisher Scientific Quant-iT RiboGreen fluorescence. RNA integrity was evaluated by electrophoresis on a 1% w/v agarose gel using 250 ng RNA. 1 mg total RNA along with the ten bait probes in the pJ204 vector described above were submitted for cDNA library preparation and the ULTImate Y1H assay service provided by Hybrigenics Services (Cambridge, MA). Sequences of interacting clones identified by the Y1H experiments were used in a BLAST search against the *Kalanchoe fedtschenkoi* transcriptome deposited in Phytozome (https://phytozome-next.jgi.doe.gov/) to identify the full-length transcripts. Transcription factor names and descriptions in this study were based upon the corresponding annotation of the closest *Arabidopsis* ortholog by nucleotide sequence homology of the full-length coding sequence (CDS). BLAST comparisons were made with TAIR BLAST 2.9.0+.

In silico analysis of the Kalanchoe fedtschenkoi Ppc1 promoter

A 3 kb segment of the proximal *Kalanchoe fedtschenkoi Ppc1* (Kaladp0095s0055.1) promoter directly upstream of the transcriptional start site was used as input for the STREME algorithm. The list of MEME motif IDs generated from this analysis was then used to perform a Tomtom search to produce a list of transcription factors known to interact with each unique motif. The entire set of Tomtom transcription factor names for each MEME motif ID was then searched for matches to the transcription factors identified as "very high confidence" or "high confidence" from the collective set of all ten Y1H screens. Identification of the location(s) of each MEME motif within the 3 kb *K. fedtschenkoi Ppc1* promoter was performed with the FIMO algorithm. STREME, Tomtom, and FIMO were accessed through the MEME Suite (Bailey et al., 2015) using default parameters.

Additionally, the 3 kb *K. fedtschenkoi Ppc1* promoter was searched for complete or partial matches to the *Arabidopsis* consensus binding sequences of the "very high confidence" or "high confidence" transcription factors from the 10 screens. If available, *Arabidopsis* consensus binding sequences were retrieved from Cistrome Data Browser or JASPAR. As direct experimental binding site data were not available for the three transcription factors *HSI2-LIKE 1* (AT4G32010.1), *HAT9* (AT2G22800.1), and *ERF106* (AT5G07580.1), consensus binding site sequences were predicted from Protein-Binding Microarray (PBM) data of closely related members of the same transcription factor family (Franco-Zorrilla et al., 2014).

Transient expression subcellular localization assays in Kalanchoe fedtschenkoi

Total RNA isolation from *Kalanchoe fedtschenkoi* plants was undertaken as described above. 10 ng RNA was then used for cDNA synthesis with the Bio-Rad Iscript cDNA Synthesis Kit according to product instructions. The full-length coding sequences of *KfERF9*, *KfERF106*, *KfTCP4*, and *KfPIF1* were amplified without stop codons from the resulting cDNA using New England Biolabs High Fidelity Phusion DNA polymerase. Primer sequences are provided in Supplemental Table 2. The amplified PCR products were run on a 1% w/v agarose gel and purified with the Qiagen QIAquick Gel Extraction Kit following the manufacturer's recommendations although nuclease-free water was used for final elution of the PCR products.

Purified transcription factor cDNAs were then used to set up cloning reactions with the Thermo Fisher Scientific pENTR D-TOPO directional cloning kit according to product instructions with the exception that the incubation time of the ligation reaction was increased to overnight to improve efficiency. Successful entry clones were confirmed by Sanger sequencing and used to generate plant expression constructs using the Thermo Fisher Scientific Gateway LR Clonase II Enzyme Mix according to manufacturer's instructions. After numerous attempts, *KfPIF1* entry cloning failed and this transcription factor was not used subsequently in the experiments described in this section. CaMV 35S promoter-driven C-terminal sGFP fusion expression constructs were made using the improved Gateway binary vector ImpGWB405 (Nakagawa et al., 2007). Final expression clones were verified by Sanger sequencing. CDS and amino acid sequences of *KfERF9*, *KfERF106*, and *KfTCP4* are provided in Supplemental Figures 2, 3, and 4. Construct diagrams are provided in Supplemental Figure 5.

Agrobacterium tumefaciens GV3101 cells containing *KfERF9*, *KfERF106*, and *KfTCP4* as sGFP fusion constructs and a culture carrying empty vector ImpGWB405 were grown for 2 d at 28 °C in YEP containing 25 mg/L gentamycin, 50 mg/L rifampicin, and 100 mg/L spectinomycin. Cells were pelleted by centrifugation at 3,200 *x g* for 10 minutes at 4 °C and re-suspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES, 150 µM acetosyringone, pH 5.6 in sterile water). Re-suspended cells were adjusted to an OD_{600} of 0.5 and incubated at 25 °C for 3 h. Mature *Kalanchoe fedtschenkoi* leaves from tissue culture-grown plants were detached and gently punctured several times on their abaxial leaf surfaces with an 18 gauge needle. Care was taken not to puncture the leaves all the way through. *Agrobacterium* suspensions described above were then used to syringe-infiltrate the punctured locations on the detached leaves. Several sites were infiltrated on each leaf. Infiltrated leaves were then placed on wet filter paper in petri plates and returned to the growth chamber for 4 days. After this time leaves were thinly sectioned by hand with a razor blade and examined for sGFP fluorescence using a Keyence BZ-X710 microscope. Fluoroshield with DAPI (Sigma Aldrich product no. F6057) was used for microscope slide mounting and visualization of nuclei. For visualization of GFP, a Keyence model OP-87763 filter was used with an excitation wavelength of 470/40 nm and an emission wavelength of 525/50 nm. For DAPI imaging, a Keyence model OP-87762 filter was used with an excitation wavelength of 360/40 nm and an emission wavelength of 460/50 nm.

Results

Y1H assays reveal several high-confidence interactions with Ppc1 promoter elements

Numerous prey clone fragments were identified across all bait probes used for Y1H screens. A full list of the predicted *K. fedtschenkoi* transcript IDs based on the sequences of these prey clones sorted by bait probe is provided in Appendix Table 1. The data provided in this table also contains the corresponding closest *Arabidopsis* ortholog by CDS sequence homology, related functional annotations, number of redundant positive colonies per prey clone from the screen, and binding confidence scores. The majority of these predicted interactions were assigned a "D" Predicted Biological Score (PBS) by Hybrigenics Services. This category of interactions represents binding events that are barely detectable by the Y1H technique and/or carry a risk of being a false positive. A detailed explanation of PBS scores can be found at the address www.hybrigenics-services.com/files/medias/ourservices/how_to_read_your_results.pdf. However, several very high-confidence (PBS score of "A") and high-confidence (PBS score of "B") transcription factor interactions were detected by the screens. The ABI family transcription factor *HSI2-LIKE1*, bHLH transcription factor *PIF1*, and AP2/ERF member *ERF9* were all predicted to interact with the G-mesophyll (MEM1a) bait probe with a PBS score of "B". Results from the CACT-mesophyll (MEM1b) screen indicate "A" PBS score binding interactions with the homeobox domain containing transcription factors *HAT9*, *HB-2*, and *HB-4*. GATA box results predict interactions with the AP2/ERF transcription factor *ERF106* and the MYB family member *MYBS1* with a PBS score of "A" and "B", respectively. *ERF106* was also identified as interacting with the -

171 to -370 *K. fedtschenkoi Ppc1* promoter region with a PBS score of "B". *TCP4* was detected in screens of both the -341 to -540 and -681 to -880 *K. fedtschenkoi Ppc1* promoter regions with a PBS score of "A" in both assays. These "A" and "B"-scoring interactions described above, which were often detected by multiple prey clones corresponding to the same *K. fedtschenkoi* transcript, are presented in Table 1. Although numerous lower- scoring interactions were detected in screens of the evening element, enhancer-like elements, -1 to -200 *Ppc1* promoter, and -511 to -710 *Ppc1* promoter, these bait probes did not interact with any prey clones to produce "A" or "B" PBS scores. As can be seen in Appendix Table 1, the vast majority of prey transcripts producing "A" and "B" PBS scores are predicted to encode proteins possessing transcription factor functions. In comparison, the annotations of prey transcripts with lower PBS scores are enriched for terms relating to enzymatic processes, photosystem-related roles, membrane functions, and ribosomal proteins.

The Ppc1 promoter contains likely binding sites for Y1H-predicted interacting factors

In-silico inspection of the proximal 3 kb *KfPpc1* promoter for un-gapped nucleotide sequences similar to the reported *Arabidopsis* consensus binding sequences of the Y1H-identified "A" and "B" scoring transcription factors provides additional evidence for the binding of some of these factors. The closest *KfPpc1* promoter match to each of the "A" and "B" scoring transcription factors' *Arabidopsis* consensus binding site is presented in Table 1 along with the *KfPpc1* promoter MEME motif sequence(s) and ID(s) in which these matches to the *Arabidopsis* binding site occur. A schematic of

the 3 kb *KfPpc1* promoter with the locations of these putative binding sites labeled is presented in Figure 1 and was generated from data derived, in part, from Table 1. Text searches of the Tomtom transcription factor annotations from the list of MEME IDs identified in the *KfPpc1* promoter produced matches to three of the high-scoring ("A" or "B" PBS score) transcription factors from the Y1H screens, *ERF9*, *PIF1*, and *TCP4*. The MEME IDs whose Tomtom annotations include these transcription factors and the location of these MEME motifs are included in Figure 1. Although no text-based Tomtom matches to *ERF106* were found in any of the MEME IDs identified in the *KfPpc1* promoter, we observed the presence of one occurrence of the sequence CCGCCG (MEME ID-2) (Table 1 and Figure 1). This sequence bears strong similarity to the consensus *Arabidopsis* binding site, CGCCGCC, of closely-related AP2/ERF family members from the results of PBM assays by Franco-Zorrilla et al. (2014). *KfPpc1* promoter locations of the individual putative *cis*-elements used in Y1H screens are provided in Figure 2, whereas the overlap of these elements with putative transcription factor binding sites (MEME IDs) is shown by Figure 3.

KfERF9, KfERF106, and KfTCP4 are nuclear localized in Kalanchoe fedtschenkoi

Due to the strong evidence in support of KfERF9, KfERF106, KfTCP4, and KfPIF1 interactions with the *Kalanchoe Ppc1* promoter, we sought to confirm their nuclear subcellular localization. Although *KfPIF1* was originally planned to be included in these transient expression assays, repeated attempts at cloning this transcript failed. Hence, no data are reported here for *KfPIF1*. Fluorescent microscopy analysis of *Agrobacterium*-infiltrated mature leaf tissue from *Kalanchoe fedtschenkoi* transiently

expressing *KfERF9, KfERF106,* and *KfTCP4* as C-terminus sGFP fusions showed that all three transcription factors displayed specific nuclear localization with the resulting signal overlapping with DAPI fluorescence (Figure 4). No sGFP signal from any other cellular location or structure was visible. As compared with transient protein expression in common C_3 model plants, such as *N. benthamiana*, we have observed transient expression in *Kalanchoe* species to be comparatively low-efficiency. Experiments such as that described here typically result in only a very small proportion of cells in infiltrated tissue displaying signs of recombinant protein expression. Hence, only a few isolated cells displaying a nuclear sGFP signal are seen in Figure 4.

Discussion

In this study we have taken a multipronged approach to identifying transcription factors that interact with the *Kalanchoe Ppc1* promoter. Results from Y1H screens using bait probes designed from discrete *KfPpc1* promoter regions or concatenated putative *cis*-regulatory elements elicited nine transcription factors with either "A" or "B" PBS binding scores. *In silico* analysis based on correlating identified *Ppc1* promoter motifs with preferred transcription factor binding sites resulted in the prediction of numerous transcription factors. Four of these, *KfERF9, KfERF106, KfTCP4,* and *KfPIF1*, overlapped with the list of nine high-scoring Y1H transcription factors. As PBS scores have been validated to correlate positively with actual biological significance (Ruffin et al., 2019) and the vast majority of high-scoring transcripts identified through the Y1H screens possess transcription factor-related annotations, the set of *trans*-acting factors identified in this study appear to be promising candidates for further functional analysis.

To this end, additional experiments in the form of electrophoretic mobility shift assays (EMSAs) followed by effector-reporter studies designed to further confirm the interaction of several high-scoring transcription factors with discreet regions of the *KfPpc1* promoter and elucidate the nature of their regulatory effects were planned. However, attempts at recombinant protein production for use in EMSA assays first in *N. benthamiana*, then in *E. coli* have not been successful to date.

A comparison of the *in-silico* predictions generated in this study with the experimentally-derived Y1H data illustrates both the validity and limitations of bioinformatics-based transcription factor binding site modeling. Indeed, the majority of high-scoring interactions (Table 1) detected through Y1H screens were not predicted with the bioinformatics workflow used here. Even in model plant systems, the limited reliability with which transcription factor DNA interactions can be predicted with traditional bioinformatics approaches (e.g., MEME followed by FIMO) is well-known (Peng et al., 2016). An additional challenge to *in-silico* modeling is posed when working with a non-model organism, such as *Kalanchoe*, as the vast majority of available transcription factor binding site information in databases such as JASPAR, Cistrome, and AtTFDB is obtained from *Arabidopsis*. Although *cis*-regulatory elements within promoters generally exhibit higher conservation than regions comparatively lacking in regulatory functions, this is not universally true. In some cases, functional regulatory sites have even been found to possess more variation than background sequences (Peng et al., 2016; Shen et al., 2020). This divergence in preferred transcription factor binding sites might be illustrated by the observation in the current study that the majority of STREME-elicited motifs for the *Kalanchoe Ppc1* promoter did not contain associated

Tomtom transcription factor annotations that matched the list of high-scoring Y1H-detected interactions (Table 1 and Figure 1). This uncertainty in attempting to correlate *Arabidopsis* bioinformatics data to *Kalanchoe* is unavoidable and will continue to be an issue faced by other *in-silico* experiments in other CAM and non-model plant species at least for the foreseeable future.

An interesting question is raised by the relative lack of binding interactions detected by the Y1H screen of the evening element and enhancer elements bait probe (Appendix Table 1). As can be seen in this table, only three interactions were detected with the evening element, two of which are annotated transcription factors, *AHDP* and *BZIP29*, with a PBS score of "C". In the case of the enhancer elements bait probe, only eight interactions were detected, all of which were low-confidence possible false positive interactions. *MYB59* with a PBS score of "D" represents the only annotated transcription factor from the enhancer elements screen. The lack of binding events to the evening element is quite interesting given its sequence similarity to the GATA box (Supplemental Table 1), which generated the largest set of interactions of all bait probes evaluated. This seems to suggest that the evening element's *trans*-activation or repression is largely dependent upon the regulatory binding events of nearby regions and that although the EE possesses a GATA core, this alone is insufficient to recruit interacting factors. Given the circadian-associated nature of the evening element, this seems to be a likely possibility. Enhancers have been reported to exhibit more sequence variation between species compared to other CREs (Peng et al., 2016), which raises the question if enhancer sequences in the *Kalanchoe Ppc1* promoter were correctly identified based upon predictions from *Arabidopsis* considering the low number of Y1H interactions. Because

enhancers are often far upstream of genes, downstream, or in intronic regions (Li et al., 2020) it is also a possibility that the most biologically significant *Ppc1* enhancers were not identified in this study as only the 2.8 kb promoter region of *Ppc1* was considered when designing Y1H bait probes. Another possibility is that the putative enhancers utilized in this study are in fact biologically significant, but are only bound and activated by relevant factors in a tightly regulated manner that is very much interdependent upon binding events in other enhancer regions present within the *Ppc1* promoter. In conclusion the work presented in this study has identified several transcription factors that likely interact with predicted *cis*-regulatory elements of the *Kalanchoe* CAM-type *Ppc* promoter. These transcription factors appear to be promising candidates for further characterization as to their role in the regulation of *Ppc1* expression in CAM photosynthesis. Therefore, future work should focus on additional steps to confirm the interaction of these high-scoring transcription factors with predicted *KfPpc1* promoter binding sites. Electrophoretic mobility shift assays (EMSAs) (Hellman & Fried, 2007) using wild type and mutated putative *KfPpc1* promoter binding sites for these factors would likely be an effective *in vitro* technique for accomplishing this. Regulatory effects of these transcription factors on *KfPpc1* expression could be investigated through transient co-transfection assays in *Kalanchoe* leaf tissue using a *KfPpc1* promoter-GFP reporter and transcription factor overexpression constructs similar to Chen et al. (2006) where putative promoter binding regions are either examined with their native sequence, mutated, or deleted. Although Chen et al. utilized *O. sativa* protoplasts for co-transfection assays, protoplast generation has not been optimized for *Kalanchoe* species.

Hence, a leaf co-infiltration approach using similar methodology to that described above for transcription factor subcellular localization assays may prove more successful.

Acknowledgements

The authors would like to thank Dr. James Hartwell at the Institute of Systems, Molecular and Integrative Biology, University of Liverpool, United Kingdom for providing much of the *Kalanchoe* sequencing data referenced in this study. We are grateful for the tissue culture and greenhouse assistance provided by Bailey Watkins. We thank our current laboratory manager Lisa Petrusa and previous laboratory manager Rebecca Albion for their contributions in facilitating the research described in this study. We would like to acknowledge The National Institutes of Health INBRE and the Nevada INBRE for their support. This work was supported by the Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834, the Nevada Agricultural Experiment Station, and the UNR Foundation.

References

Akyildiz, M., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., & Westhoff, P. (2007). Evolution and Function of a *cis* -Regulatory Module for Mesophyll-Specific Gene Expression in the C4 Dicot *Flaveria trinervia*. *The Plant Cell*, *19*(11), 3391–3402. https://doi.org/10.1105/tpc.107.053322

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, *43*(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

Borland, A. M., Barrera Zambrano, V. A., Ceusters, J., & Shorrock, K. (2011). The photosynthetic plasticity of crassulacean acid metabolism: An evolutionary innovation for sustainable productivity in a changing world. *New Phytologist*, *191*(3), 619–633. https://doi.org/10.1111/j.1469-8137.2011.03781.x

Boxall, S. F., Kadu, N., Dever, L. V., Kneřová, J., Waller, J. L., Gould, P. J. D., & Hartwell, J. (2020). *Kalanchoë* PPC1 Is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and Guard Cell Signaling Genes. *The Plant Cell*, *32*(4), 1136–1160. https://doi.org/10.1105/tpc.19.00481

Castillon, A., Shen, H., & Huq, E. (2007). Phytochrome Interacting Factors: Central players in phytochrome-mediated light signaling networks. *Trends in Plant Science*, *12*(11), 514–521. https://doi.org/10.1016/j.tplants.2007.10.001

Challa, K. R., Aggarwal, P., & Nath, U. (2016). Activation of *YUCCA5* by the Transcription Factor TCP4 Integrates Developmental and Environmental Signals to Promote Hypocotyl Elongation in Arabidopsis. *The Plant Cell*, *28*(9), 2117–2130. https://doi.org/10.1105/tpc.16.00360

Chen, L.-Y., Xin, Y., Wai, C. M., Liu, J., & Ming, R. (2020). The role of cis-elements in the evolution of crassulacean acid metabolism photosynthesis. *Horticulture Research*, *7*(1), 5. https://doi.org/10.1038/s41438-019-0229-0

Chen, S., Tao, L., Zeng, L., Vega-Sanchez, M. E., Umemura, K., & Wang, G.-L. (2006). A highly efficient transient protoplast system for analyzing defence gene expression and protein-protein interactions in rice. *Molecular Plant Pathology*, *7*(5), 417–427. https://doi.org/10.1111/j.1364-3703.2006.00346.x

Cushman, J. C., & Bohnert, H. J. (1992). Salt stress alters A/T-rich DNA-binding factor interactions within the phosphoenolpyruvate carboxylase promoter from *Mesembryanthemum crystallinum*. *Plant Molecular Biology*, *20*(3), 411–424. https://doi.org/10.1007/BF00040601

Cushman, J. C., & Borland, A. M. (2002). Induction of Crassulacean acid metabolism by water limitation: CAM induction by water limitation. *Plant, Cell & Environment*, *25*(2), 295–310. https://doi.org/10.1046/j.0016-8025.2001.00760.x

Deng, H., Zhang, L.-S., Zhang, G.-Q., Zheng, B.-Q., Liu, Z.-J., & Wang, Y. (2016). Evolutionary history of PEPC genes in green plants: Implications for the evolution of CAM in orchids. *Molecular Phylogenetics and Evolution*, *94*, 559–564. https://doi.org/10.1016/j.ympev.2015.10.007

Ding, Y., Liu, N., Virlouvet, L., Riethoven, J.-J., Fromm, M., & Avramova, Z. (2013). Four distinct types of dehydration stress memory genes in Arabidopsis thaliana. *BMC Plant Biology*, *13*(1), 229. https://doi.org/10.1186/1471-2229-13-229

Duek, P. D., & Fankhauser, C. (2005). BHLH class transcription factors take centre stage in phytochrome signalling. *Trends in Plant Science*, *10*(2), 51–54. https://doi.org/10.1016/j.tplants.2004.12.005

Edwards, E. J. (2019). Evolutionary trajectories, accessibility and other metaphors: The case of C $_4$ and CAM photosynthesis. *New Phytologist*, *223*(4), 1742–1755. https://doi.org/10.1111/nph.15851

Franco-Zorrilla, J. M., & Solano, R. (2017). Identification of plant transcription factor target sequences. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1860*(1), 21–30. https://doi.org/10.1016/j.bbagrm.2016.05.001

Gutierrez, R. A., Ewing, R. M., Cherry, J. M., & Green, P. J. (2002). Identification of unstable transcripts in Arabidopsis by cDNA microarray analysis: Rapid decay is associated with a group of touch- and specific clock-controlled genes. *Proceedings of the National Academy of Sciences*, *99*(17), 11513–11518. https://doi.org/10.1073/pnas.152204099

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols*, *2*(8), 1849–1861. https://doi.org/10.1038/nprot.2007.249

Kubota, A., Ito, S., Shim, J. S., Johnson, R. S., Song, Y. H., Breton, G., Goralogia, G. S., Kwon, M. S., Laboy Cintrón, D., Koyama, T., Ohme-Takagi, M., Pruneda-Paz, J. L., Kay, S. A., MacCoss, M. J., & Imaizumi, T. (2017). TCP4-dependent induction of CONSTANS transcription requires GIGANTEA in photoperiodic flowering in Arabidopsis. *PLOS Genetics*, *13*(6), e1006856. https://doi.org/10.1371/journal.pgen.1006856

Li, K., Liu, Y., Cao, H., Zhang, Y., Gu, Z., Liu, X., Yu, A., Kaphle, P., Dickerson, K. E., Ni, M., & Xu, J. (2020). Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nature Communications*, *11*(1), 485. https://doi.org/10.1038/s41467-020-14362-5

Liu, D., Hu, R., Zhang, J., Guo, H.-B., Cheng, H., Li, L., Borland, A. M., Qin, H., Chen, J.-G., Muchero, W., Tuskan, G. A., & Yang, X. (2021). Overexpression of an *Agave* Phosphoenolpyruvate Carboxylase Improves Plant Growth and Stress Tolerance. *Cells*, *10*(3), 582. https://doi.org/10.3390/cells10030582

Mazzella, M. A., Arana, M. V., Staneloni, R. J., Perelman, S., Rodriguez Batiller, M. J., Muschietti, J., Cerdán, P. D., Chen, K., Sánchez, R. A., Zhu, T., Chory, J., & Casal, J. J. (2005). Phytochrome Control of the *Arabidopsis* Transcriptome Anticipates Seedling Exposure to Light. *The Plant Cell*, *17*(9), 2507–2516. https://doi.org/10.1105/tpc.105.034322

Michael, T. P., & McClung, C. R. (2002). Phase-Specific Circadian Clock Regulatory Elements in Arabidopsis. *Plant Physiology*, *130*(2), 627–638. https://doi.org/10.1104/pp.004929

Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., Hazen, S. P., Shen, R., Priest, H. D., Sullivan, C. M., Givan, S. A., Yanovsky, M., Hong, F., Kay, S. A., & Chory, J. (2008). Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific cis-Regulatory Modules. *PLoS Genetics*, *4*(2), e14. https://doi.org/10.1371/journal.pgen.0040014

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., Lyons, E., Wang, M.-L., Chen, J., Biggers, E., Zhang, J., Huang, L., Zhang, L., Miao, W., Zhang, J., Ye, Z., Miao, C., Lin, Z., Wang, H., … Yu, Q. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics*, *47*(12), 1435–1442. https://doi.org/10.1038/ng.3435

Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., Watanabe, Y., Nakamura, K., Kimura, T., & Ishiguro, S. (2007). Improved Gateway Binary Vectors: High-Performance Vectors for Creation of Fusion Constructs in Transgenic Analysis of Plants. *Bioscience, Biotechnology, and Biochemistry*, *71*(8), 2095–2100. https://doi.org/10.1271/bbb.70216

Peng, F. Y., Hu, Z., & Yang, R.-C. (2016). Bioinformatic prediction of transcription factor binding sites at promoter regions of genes for photoperiod and vernalization responses in model and temperate cereal plants. *BMC Genomics*, *17*(1), 573. https://doi.org/10.1186/s12864-016-2916-7

Ping, C.-Y., Chen, F.-C., Cheng, T.-C., Lin, H.-L., Lin, T.-S., Yang, W.-J., & Lee, Y.-I. (2018). Expression Profiles of Phosphoenolpyruvate Carboxylase and Phosphoenolpyruvate Carboxylase Kinase Genes in Phalaenopsis, Implications for Regulating the Performance of Crassulacean Acid Metabolism. *Frontiers in Plant Science*, *9*, 1587. https://doi.org/10.3389/fpls.2018.01587

Ruffin, M., Thompson, K. E., Corvol, H., & Guillot, L. (2019). Two-hybrid screening of FAM13A protein partners in lung epithelial cells. *BMC Research Notes*, *12*(1), 804. https://doi.org/10.1186/s13104-019-4840-9

Schaeffer, H. J., Forsthoefel, N. R., & Cushman, J. C. (1995). Identification of enhancer and silencer regions involved in salt-responsive expression of Crassulacean acid metabolism (CAM) genes in the facultative halophyte *Mesembryanthemum crystallinum*. *Plant Molecular Biology*, *28*(2), 205–218. https://doi.org/10.1007/BF00020241

Shen, Z., Lin, Y., & Zou, Q. (2020). Transcription factors–DNA interactions in rice: Identification and verification. *Briefings in Bioinformatics*, *21*(3), 946–956. https://doi.org/10.1093/bib/bbz045

Silvera, K., Neubig, K. M., Whitten, W. M., Williams, N. H., Winter, K., & Cushman, J. C. (2010). Evolution along the crassulacean acid metabolism continuum. *Functional Plant Biology*, *37*(11), 995. https://doi.org/10.1071/FP10084

Silvera, K., Winter, K., Rodriguez, B. L., Albion, R. L., & Cushman, J. C. (2014). Multiple isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): Implications for the evolution of crassulacean acid metabolism. *Journal of Experimental Botany*, *65*(13), 3623–3636. https://doi.org/10.1093/jxb/eru234

Taybi, T., Nimmo, H. G., & Borland, A. M. (2004). Expression of Phospho *enol* pyruvate Carboxylase and Phospho *enol* pyruvate Carboxylase Kinase Genes. Implications for Genotypic

Van den Broeck, L., Dubois, M., Vermeersch, M., Storme, V., Matsui, M., & Inzé, D. (2017). From network to phenotype: The dynamic wiring of an Arabidopsis transcriptional network induced by osmotic stress. *Molecular Systems Biology*, *13*(12), 961. https://doi.org/10.15252/msb.20177840

Zhang, J., Hu, R., Sreedasyam, A., Garcia, T. M., Lipzen, A., Wang, M., Yerramsetty, P., Liu, D., Ng, V., Schmutz, J., Cushman, J. C., Borland, A. M., Pasha, A., Provart, N. J., Chen, J.-G., Muchero, W., Tuskan, G. A., & Yang, X. (2020). Light-responsive expression atlas reveals the effects of light quality and intensity in *Kalanchoë fedtschenkoi*, a plant with crassulacean acid metabolism. *GigaScience*, *9*(3), giaa018. https://doi.org/10.1093/gigascience/giaa018

Figures and Tables

Table 1. High-confidence interacting transcription factors from yeast one-hybrid screens. Summary table of transcription factors across all ten yeast one-hybrid screens with a very high-confidence (A) or high-confidence (B) global PBS score as determined by the Hybrigenics confidence scoring system. The proximal 3 kb *Kalanchoe fedtschenkoi Ppc1* promoter was searched for complete or partial matches to the experimentally confirmed or predicted consensus binding sequences of these transcription factors in *Arabidopsis*. These matching nucleotide sequences and their occurrence within motifs identified in the 3 kb *KfPpc1* promoter with the STREME algorithm are indicated in green.

Transcription factor description	Transcription factor family	Y1H screen bait probe interactions	Global PBS score from Y1H screen	*K. fedtschenkoi* transcript ID	*Arabidopsis* gene ID	*Arabidopsis* consensus binding motif(s)	Matches to consensus binding motif in *K. fedtschenkoi PEPC1* promoter (in green)	Motifs in *K. fedtschenkoi PEPC1* promoter containing the match	Location of motif in *K. fedtschenkoi PEPC1* promoter (start)	Location of motif in *K. fedtschenkoi PEPC1* promoter (end)	Strand (+ or -)	MEME ID associated with this motif
HSF2-LIKE 1	ABI	G-nucoplyll	B	Kalah|01066319.1	AT4G22810.1	CACCTG	CACCTG	CACCTC	511	516	plus	MEME-96
								ACCTCC	315	320	minus	MEME-22
PHY-INTERACTING FACTOR 1 (PIF1)	bHLH	G-nucoplyll	B	Kalah|05956837.1	AT2G20180.2	GGAAGGTCACGTGG	GGAAGGTCACGTGG	TCACCC	1222	1227	plus	MEME-75
								TCACTA	1349	1354	plus	MEME-23
												MEME-74
ATHB2-9	AP2-EREBP	G-nucoplyll	B	Kalah|01066954.1	AT5G44210.1	TCCTCCGCCCCCATTTT	TCCTCCGCCCCCATTTT	CCGCCG	1399	1404	plus	MEME-2
						CGGCGGCGGCTGGCGGCGGCGG	CGGCGGCGGCTGGCGGCGGCGG	GGTGGC	681	686	minus	MEME-16
												MEME-41
HAT9	HB	CACT-nucoplyll	A	Kalah|01254697.1	AT2G22430.1	AATNATT	AATGATT	GATGAAG	165	171	minus	MEME-58
							AATCATT	GGATCA	129	134	plus	MEME-44
									648	653	plus	
									798	803	plus	
									992	997	plus	
ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 2 (ATHB-2)	HB	CACT-nucoplyll	A	Kalah|01446811.1	AT4G16780.1	TAATGATTA	TAATGATTA	GATGAAG	165	171	minus	MEME-58
						CAATGATTG	CAATGATTG					
						CAATCATTG	CAATCATTG	GGATCA	129	134	plus	MEME-44
									648	653	plus	
									798	803	plus	
									992	997	plus	
ARABIDOPSIS THALIANA HOMEOBOX-LEUCINE ZIPPER PROTEIN 4 (ATHB-4)	HB	CACT-nucoplyll	A	Kalah|01866045.1	AT2G44910.1	TAATCATTGG	TAATCATTGG	TCATCC	167	172	plus	MEME-14
ERF106	AP2-EREBP	GATA box	A	Kalah|01036001.1	AT5G07580.1	CGCCGCC	CGCCGCC	CCGCCG	1399	1404	plus	MEME-2
MYBS1	MYB	GATA box	B	Kalah|01096366.1	AT1G09810.1	AATGTGGATAAGGTT	AATGTGGATAAGGTT	TGGATG	1884	1889	plus	MEME-77
									168	173	minus	
								TGGATC	120	133	plus	MEME-41
						ATTAAGCTTATCTTC	ATTAAGCTTATCTTC	CAGCTC	511	516	plus	MEME-96
								ACCTCC	315	320	minus	MEME-22
TCP FAMILY TRANSCRIPTION FACTOR 4 (TCP4)	TCP	(GPPC1 promoter (-543 to -538); GPPC1 promoter (+681 to +686)	A	Kalah|01161441.1	AT3G15030.1	GGGAATCAC	GGGAACCAC	AGAACCC	288	293	minus	MEME-43
									774	779	minus	

Figure 1. *KfPpc1* promoter locations of MEME IDs associated with putative binding sites of high-confidence interacting transcription factors from yeast one-hybrid screens. MEME IDs and locations are obtained from the data presented in Table 1. Additionally, MEME IDs 72 (CCCACA), 28 (CCCACC), 26 (ACGTGG), 93 (GACAAC), 46 (GCCKCC), 6 (GGTGGG), and 7 (GTCCACC) were included in this figure as the Tomtom annotations associated with these MEME IDs include the high-confidence interacting factors *TCP4* (MEME IDs 72, 28, 26, 93, 6, and 7), *PIF1* (MEME ID 26), and *ERF9* (MEME ID 46). This image was produced with CLC Genomics Workbench 8.0 using a proximal 3 kb *KfPpc1* promoter. Base position 3,000 is directly upstream of the transcriptional start site.

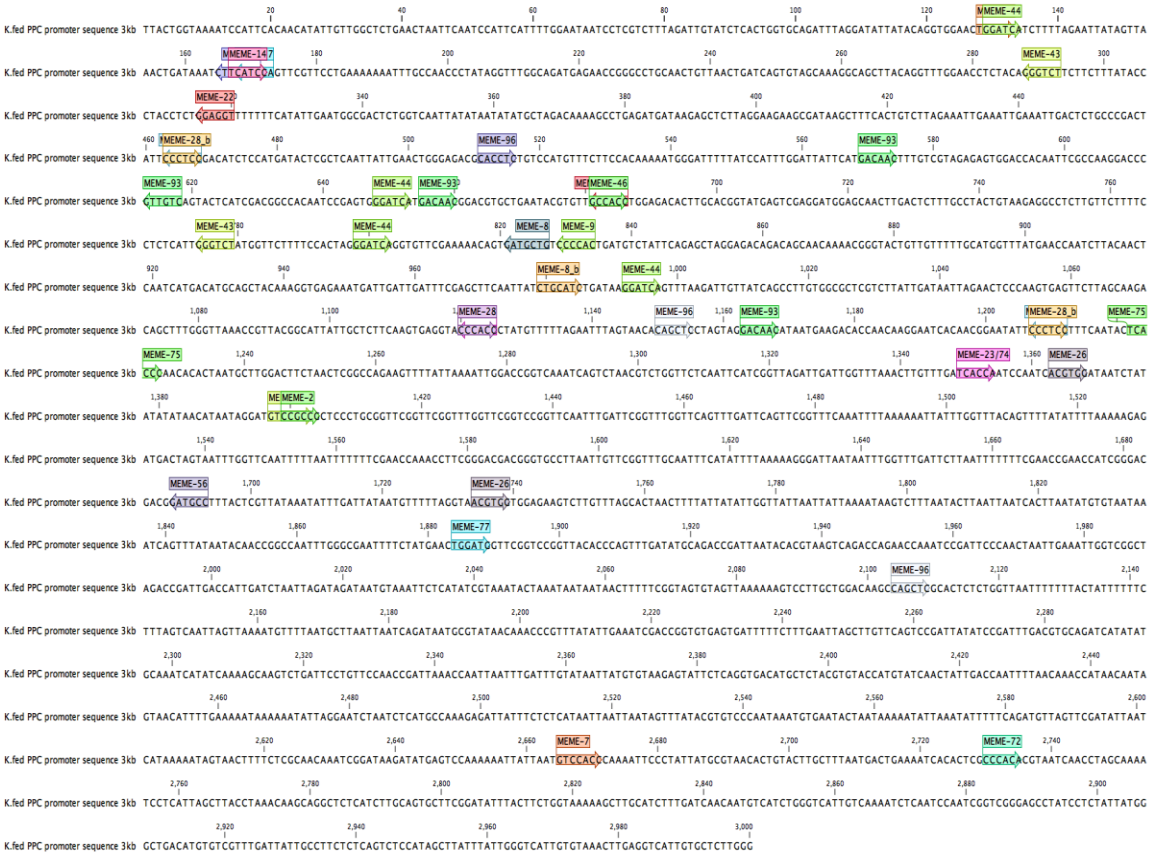
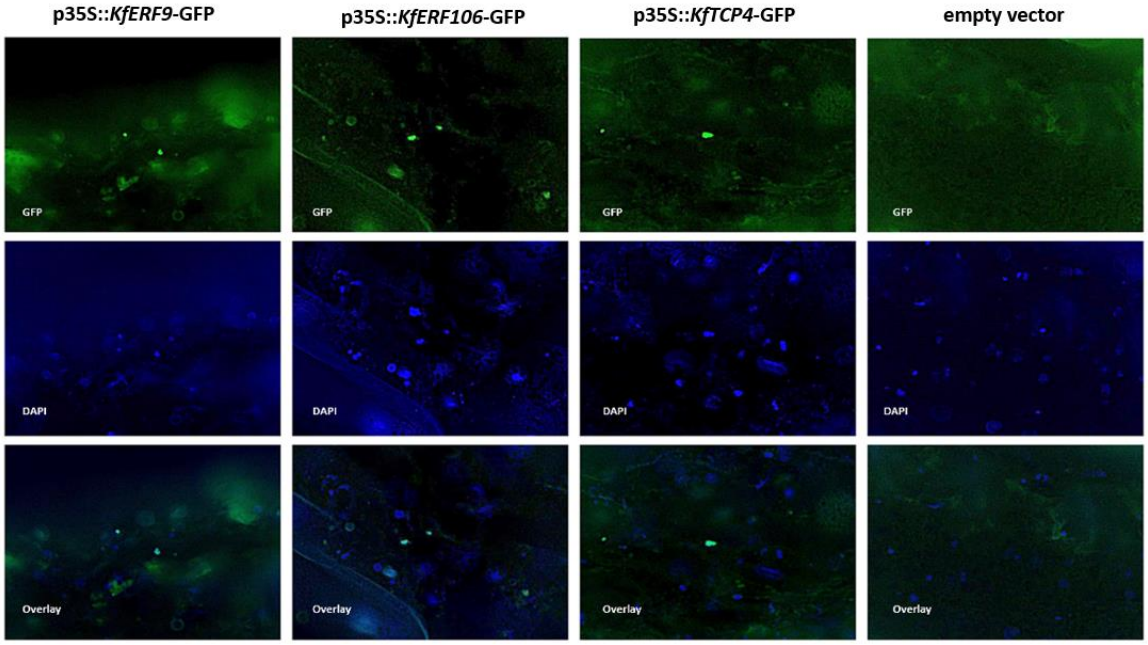


Figure 2. *KfPpc1* promoter locations of putative *cis*-regulatory elements used as yeast one-hybrid screen bait probes. Each annotation represents one instance of the core element of the particular CRE. This image was produced with CLC Genomics Workbench 8.0 using a proximal 3 kb *KfPpc1* promoter. Base position 3,000 is directly upstream of the transcriptional start site.

Figure 3. *KfPpc1* promoter locations of MEME IDs associated with putative binding sites of high-confidence interacting transcription factors from yeast one-hybrid screens and locations of putative *cis*-regulatory elements used as bait in these screens. This figure combines annotations from those of Figures 1 and 2. This image was produced with CLC Genomics Workbench 8.0 using a proximal 3 kb *KfPpc1* promoter. Base position 3,000 is directly upstream of the transcriptional start site.

Figure 4. ERF9, ERF106, and TCP4 are nuclear localized in *K. fedtschenkoi*. *Agrobacterium* cultures harboring ERF9, ERF106, and TCP4 as sGFP fusion constructs were used for transient expression infiltration assays in *K. fedtschenkoi* leaf tissue. Sectioned tissue was then examined with fluorescence microscopy. Consistent with the expected localization of transcription factors, nuclear sGFP signals were observed. DAPI co-localization confirmed nuclear location. Images were taken at 200X magnification using a Keyence BZ-X710 microscope.



Supplemental Table 1. Bait probes used in yeast one-hybrid screens. Each bait probe was screened against a *Kalanchoe fedtschenkoi* cDNA expression clone library with the ULTImate Y1H assay service through Hybrigenics Services (Cambridge, MA). 5' and 3' *SfiI* restriction sites are indicated by lowercase blue nucleotide sequences. Individual instances of repeated elements are indicated by red nucleotide sequences within the full bait probe sequences.

Yeast one-hybrid bait probe	*KlPEPC1* promoter elements	Full bait probe sequence
evening element (EE)	TCGATATTAA	ggccggacgggccTCGATATTAATAAGATATGACGGATATTTATCGATATTAATAAGATATGACGGATATTTATCG
	TAAGATATGA	ATATTAATAAGATATGACGGATATTTATCGATATTAATAAGATATGACGGATATTTATCGATATTAATAAGA
	CGGATATTTA	TATGACGGATATTTAggccactggggcc
GATA box	ATGGATAAAA	ggccggacgggccATGGATAAAAGTGGATAAATCTCGGATATAATCGGATAAGATCGGATATTTGAGGATAGGCATG
	GTGGATAATC	GATAAAAGTGGATAATCTCGGATATAATCGGATAAGATCGGATATTTGAGGATAGGCATGGATAAAAGTGG
	TCGGATATAA	ATAATCTCGGATATAATCGGATAAGATCGGATATTTGAGGATAGGCATGGATAAAAGTGGATAATCTCGGAT
	TCGGATAAGA	ATAATCGGATAAGATCGGATATTTGAGGATAGGCATGGATAAAGTGGATAATCTCGGATAATCGGATA
	TCGGATATTT	AGATCGGATATTTGAGGATAGGCggccactggggcc
	GAGGATAGGC	
G-mesophyll	TCCATGTTTC	ggccggacgggccTCCATGTTTCTGCATGTCATGTTGTGATTCTGATGTGCAGAGCATGTCACAATGTGAATATCCA
	TGCATGTCAT	TGTTTCTGCATGTCATGTTGTGATTCTGATGTGCAGAGCATGTCACAATGTGAATATCCATGTTTCTGCATGT
	GTTGTGATTC	CATGTTGTGATTCTGATGTGCAGAGCATGTCACAATGTGAATATCCATGTTTCTGCATGTCATGTTGTGATTC
	TGATGTGCAG	TGATGTGCAGAGCATGTCACAATGTGAATATCCATGTTTCTGCATGTCATGTTGTGATTCTGATGTGCAGAGC
	AGCATGTCAC	ATGTCACAATGTGAATAggccactggggcc
	AATGTGAATA	
CACT-mesophyll	AGACACTTGC	ggccggacgggccAGACACTTGCCATCACTGTTCCTCACTTGATCGCACTCTCAATCACTCACTAACACTGTAAGAC
	CATCACTGTT	ACTTGCCATCACTGTTCCTCACTTGATCGCACTCTCAATCACTCACTAACACTGTAAGACACTTGCCATCACT
	CCTCACTTGA	GTTCCTCACTTGATCGCACTCTCAATCACTCACTAACACTGTAAGACACTTGCCATCACTGTTCCTCACTTGA
	TCGCACTCTC	TCGCACTCTCAATCACTCACTAACACTGTAAGACACTTGCCATCACTGTTCCTCACTTGATCGCACTCTCAAT
	AATCACTCAC	CACTCACTAACACTGTAggccactggggcc
	TAACACTGTA	
enhancer-like elements	AAACCGAACA	ggccggacgggccAAACCGAACAGAACCGGACCAAACCAAATAGAACCGAACCGAACCGGACCAAACCAAATAA
	GAACCGGACC	AACCGAACCGAACCGGACCAAACCAAATAAACCGAACCTAACCGGACCGAACCAAATTAAACCGAACTTA
	AAACCAAATA	ACCGGACCGAACCAAATTggccactggggcc
	GAACCGAACC	
	GAACCGGACC	
	AAACCAAATA	
	AAACCGAACC	
	GAACCGGACC	
	AAACCAAATA	
	AAACCGAACC	
	TAACCGGACC	
	GAACCAAATT	
	AAACCGAACT	
	TAACCGGACC	
	GAACCAAATT	
KfPEPC 1 promoter (-1 to -200)		ggccggacgggccTATTATGGGCTGACATGTGTCGTTTGATTATTGCCTTCTCTCAGTCTCCATAGCTTATTTATTGG
		GTCATTGTGTAAACTTGAGGTCATTGTGCTCTTGGGTTGAGTCTAATCCCAAGGTTGAGAAATACTGTTAGG
		GGTCAAAATGCAGggccggacgggcc
KfPEPC 1 promoter (-171 to -370)		ggccggacgggccAGGCTCTCATCTTGCAGTGCTTCGGATATTTACTTCTGGTAAAAAGCTTGCATCTTTGATCAAC
		AATGTCATCTGGGTCATTGTCAAAATCTCAATCCAATCGGTCGGGAGCCTATCCTCTATTATGGGCTGACATG
		TGTCGTTTGATTAggccggacgggcc
KfPEPC 1 promoter (-341 to -540)		ggccggacgggccTTAATGTCCACCCAAAATTCCCTGTTATGCGTAACACTGTACTTGCTTTAATGACTGAAAATCA
		CACTCGCCCACACGTAATCAACCTAGCAAAATCCTCATTAGCTTACCTAAACAAGCAGGCTCTCATCTTGCA
		GTGCTTCGGATATTggccggacgggcc
KfPEPC 1 promoter (-511 to -710)		ggccggacgggccCCAATAAATGTGAATACTAATAAAAATATTAAATATTTTTCAGATGTTAGATCGATATTAATCA
		TAAAAATAGTAACTTTTCTCGCAACAAATCGGATAAGATATGAGTCCAAAAAATTATTAATGTCCACCCAAA
		ATTCCCTGTTATGCggccggacgggcc
KfPEPC 1 promoter (-681 to -880)		ggccggacgggccTTGAGCAATTTTAACAAACCATAACAATAGTAACATTTTGAAAAATAAAAAATATTAGGAATC
		TAATCTCATGCCAAAGAGATTATTTCTCTCATAATTAATTAATAGTTTATACGTGTCCCAATAAATGTGAATA
		CTAATAAAAATATTggccggacgggcc

Supplemental Table 2. Primers used in this study.

Primer Name	Sequence 5' to 3'
KfERF9 D-TOPO Fwd	CACCATGGCTCCGAAGACGGTG
KfERF9 Rev no stop codon	AGCGGCGGTGGTGTCAGGAG
KfERF106 D-TOPO Fwd	CACCATGGATGAAACGTCGACCTTG
KfERF106 Rev no stop codon	AACCAAAAGCATCTGATCCTCCGTC
KfTCP4 D-TOPO Fwd	CACCATGGGAGAAAGAAACAGCTATC
KfTCP4 Rev no stop codon	ATGGCGAGAATCAGAGGAAGCAG
KfPIF1 D-TOPO Fwd	CACCATGGCGAACCACTGCGTTC
KfPIF1 Rev no stop codon	ACACGTAAGGACTGAGGAGTGGC

Supplemental Figure 1. Local (Smith-Waterman) alignment of the *Kalanchoe fedtschenkoi* and *Kalanchoe laxiflora Ppc1* promoters. The *K. fedtschenkoi* (top strand) and the *K. laxiflora* (bottom strand) *Ppc1* promoters share 96.7% identity across the proximal 3 kb region with 1.9% gaps. Alignment was performed using SnapGene version 3.5.2. Nucleotides in red represent non-conserved bases while gaps are indicated by dashes.

Supplemental Figure 2. Complete coding sequence and amino acid sequence of *K. fedtschenkoi ERF9*. The coding sequence of *KfERF9* was used for construction of the p35S::*KfERF9*-GFP fusion construct for subcellular localization assays described in this study. The final stop codon was omitted. Sequences were retrieved from JGI Phytozomev13 (https://phytozome-next.jgi.doe.gov/).

>K.fedtschenkoi v1.1|Kaladp0010s0058.1 CDS
ATGGCTCCGAAGACGGTGGAGAAGAAGGCGGCGGCGGGAATCAAGGAGGCG
CATTTCAGAGGGGTAAGGAAGCGGCCGTGGGGGCGGTACGCGGCGGAGATC
AGAGATCCGACGAAGAAGAGTAGGGTGTGGCTTGGAACGTTTGACACGGCG
GAGGAGGCCGCGAGGGCGTACGACGCGGCGGCGATCGAGTTCCGCGGCACC
AAGGCGAAGACCAACTTCCCGATGCCAGGCGGCGCGTCTGCCGCAACGCCCC
ATCTGAGCCCCAGCCAGAGCAGCACCGTCGAGTCCTCCAGCCGAGAAAGCCA
CGCCCCGGCGGTTACCTCGATCCTCGATCTCAGCCTCAAGACCGGCTCACCCG
CCGTCACTCCCTCCGCAGCCTACGGCGGCCCGATCAGGTTCCCGTTCCAACAC
TTCCAACTCGCCGCCGCCTCCCCGAGCCTCCTTCCCGCCGCCAACTTGTTCTT
CTTCAACCCGGTTGCGCGCTCTGCCCGACCCGTGACTTTCAGATCCTACCCTA
ACATCCAAGCCGCCGTCACCGTCGCCGCCGCCACCGGTATGAACTTCAACGG
CCCGGCCCTAAGTGACTCCGACTCCTCCTCTGTCGTTGATGCGAGCAGTCCCG
CCTCCACTGCTGCCGTCCCTGACATAGACCTCAACTTCCCTCCTCCTGACACC
ACCGCCGCTTGA

>K.fedtschenkoi v1.1|Kaladp0010s0058.1.p
MAPKTVEKKAAAGIKEAHFRGVRKRPWGRYAAEIRDPTKKSRVWLGTFDTAEE
AARAYDAAAIEFRGTKAKTNFPMPGGASAATPHLSPSQSSTVESSSRESHAPAVT
SILDLSLKTGSPAVTPSAAYGGPIRFPFQHFQLAAASPSLLPAANLFFFNPVARSAR
PVTFRSYPNIQAAVTVAAATGMNFNGPALSDSDSSSVVDASSPASTAAVPDIDLN
FPPPDTTAA*

Supplemental Figure 3. Complete coding sequence and amino acid sequence of *K. fedtschenkoi ERF106*. The coding sequence of *KfERF106* was used for construction of the p35S::*KfERF106*-GFP fusion construct for subcellular localization assays described in this study. The final stop codon was omitted. Sequences were retrieved from JGI Phytozomev13 (https://phytozome-next.jgi.doe.gov/).

>K.fedtschenkoi v1.1|Kaladp0403s0001.1 CDS
ATGGATGAAACGTCGACCTTGGACCTCATCCGGAGCCACCTCCTGGCCGACT
TCGCCTCCACCGAGTCCTTCATCTCATCTCTCCACTCAGACCTACTGCTCGAC
TCCACCCCCAGCACCACCACCAGCCCCTACGACGCGCCTTTCTGGTCTCTTTT
CCCGGTCAAACGCGAGATGACAGACACCGATCACAACTACAACTTGCTGGAG
TTCATCAAGCCGGAGTTCTTCGACATAGCGACGTCCCAGAGGGAAGACACTG
GCAAGACCTCCTCGTCTGAGGAAACCTCCTCGCCGCCATTCGACAGGAAGCA
TTACAGGGGGGTCCGGAGGAGGCCGTGGGGGAAGTTCGCGGCGGAGATCCG
GGACCCGGCCAGGAAGGGCAGCCGGGTGTGGCTGGGCACGTTCGACTCGGA
TGTGGACGCAGCCAAGGCGTATGACTTTGCCGCTTTCAAGATGAGGGGGAGC
AAAGCGATTCTCAACTTCCCGCTTCAGGCTGGAAAAGAGGCTGGGCCGCCGG
AAAACAGCGGCCGGAGGAGGAGGAGGAGGAGGGACAGTCGAGCGGAGACC
ACCGCGACGGAGTCTGCAGCAGGCAGTAGCAGCAGCTGCTTCAATAATAATG
ATGATGACAATAATAATGAGCCATGGACCCCATCCAGTCACAATTTGACTTTC
ACCACCACCACCACGACGGAGGATCAGATGCTTTTGGTTTGA

>K.fedtschenkoi v1.1|Kaladp0403s0001.1.p
MDETSTLDLIRSHLLADFASTESFISSLHSDLLLDSTPSTTTSPYDAPFWSLFPVKR
EMTDTDHNYNLLEFIKPEFFDIATSQREDTGKTSSSEETSSPPFDRKHYRGVRRRP
WGKFAAEIRDPARKGSRVWLGTFDSDVDAAKAYDFAAFKMRGSKAILNFPLQA
GKEAGPPENSGRRRRRRRDSRAETTATESAAGSSSSCFNNNDDDNNNEPWTPSS
HNLTFTTTTTTEDQMLLV*

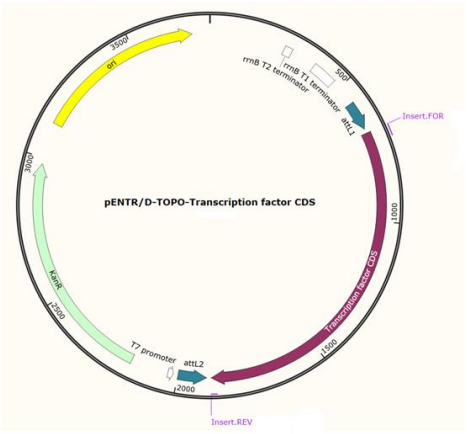
Supplemental Figure 4. Complete coding sequence and amino acid sequence of *K. fedtschenkoi TCP4*. The coding sequence of *KfTCP4* was used for construction of the p35S::*KfTCP4*-GFP fusion construct for subcellular localization assays described in this study. The final stop codon was omitted. Sequences were retrieved from JGI Phytozomev13 (https://phytozome-next.jgi.doe.gov/).

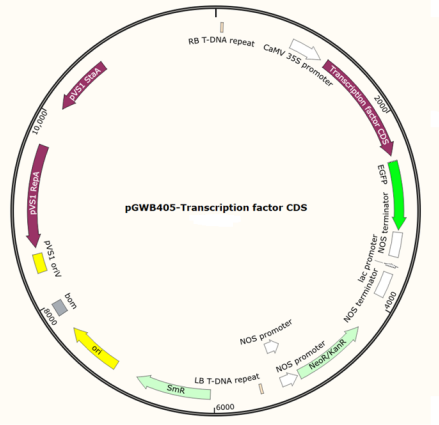
>K.fedtschenkoi v1.1|Kaladp0011s1344.1 CDS
ATGGGAGAAAGAAACAGCTATCGTCAGACAACGTCCAGATTGGGAATGAGG
AACTCAGGTGGAGGAGAGATCATAGAGGTGCAAGGAGGCCACATTGTTCGGT
CCACCGGAAGAAAAGACAGACACAGCAAGGTTTGCACGGCCAAGGGTCCAA
GAGACAGGAGGGTCAGGTTATCAGCCCACACTGCCATTCAATTCTACGATGT
TCAAGACCGATTAGGCTACGACAGGCCAAGCAAAGCGGTGGACTGGCTCATC
AAGAAGGCCAAAGCTGCTATAGACGAGCTTGCCGAGCTACCACCTTGGCATC
CCACCGCCACTACCACCATTCCAACTGCTACTACGACTGCAACAAATGCTAG
TTCCCAGCCAGAGCAAATGGTCCATGGTCATGTTGCTAACTCAATGGAAGAT
GCTAATGATTACACAATTCAGCACCAGCAACATGAGAGAGTTCAAAATGAGA
ACAACCCCAGTTTTCTACCTCCATCCATGGATTCAGACTCCATTGCTGATACA
ATCAAGACATTCTTTCCGATGGGTGCTTCAACTGAGCCAACTTCCGGGTCTGG
CATGCAGTTTTATCCACCTGATTTGCTCTCTAGAACTAGCAGTAGTAGTCAGA
ACAACCAAGATCTAAGGCTCTCACTTCAGTCGTTTCAGGATTCACAGTTTTTA
CATTCTCACCATCACCACCACCACCACCAGCAACATCCTGGTCAGACCCACC
AAGCTATTTTCTCAGGAAACCAGCTAGGTTTTGATGCTTCCTCAGGAGGAGG
GGGCTGGTCAGACCATCAATCAGCAGACATCAATCGATTCCATAGAATGATG
ACTTGGAACAATACTGGTGGAGACACTGCAGGTGGCAGTGGTGGTGGTGGTG
GGGGGAGTGGTGGTAGCAGTGGGTTTATGTTTAACTCATTGCCACCACCACC
CTTGTCTCAGCCATCTCTGCAGCAGTTGTTTGGCCAAAGTGGCCAGTTTTTCA
CTCAGAGGGGACCCCTTCAGTCCAGTAATACGTCTTCAGTTCGTGCTTGGATT
GACCAGCCAATTGCCGACCACCATCAACATCATCAAACACTTCATCACCACC
CAGCTACACTGTCCGGCATTGGCCACTTCCAGAATCCTGGTGTTTTCTCCGGG
TTTCGAATTCCAGCACGAATACAAGGCGACGATGAGGAGCAAGATGGCATTC
ACAACAAACCGTCATCTGCTTCCTCTGATTCTCGCCATTGA

>K.fedtschenkoi v1.1|Kaladp0011s1344.1.p
MGERNSYRQTTSRLGMRNSGGGEIIEVQGGHIVRSTGRKDRHSKVCTAKGPRDR
RVRLSAHTAIQFYDVQDRLGYDRPSKAVDWLIKKAKAAIDELAELPPWHPTATT
TIPTATTTATNASSQPEQMVHGHVANSMEDANDYTIQHQQHERVQNENNPSFLP
PSMDSDSIADTIKTFFPMGASTEPTSGSGMQFYPPDLLSRTSSSSQNNQDLRLSLQ
SFQDSQFLHSHHHHHHHQQHPGQTHQAIFSGNQLGFDASSGGGGWSDHQSADIN
RFHRMMTWNNTGGDTAGGSGGGGGSGGSSGFMFNSLPPPPLSQPSLQQLFGQ
SGQFFTQRGPLQSSNTSSVRAWIDQPIADHHQHHQTLHHHPATLSGIGHFQNPGV
FSGFRIPARIQGDDEEQDGIHNKPSSASSDSRH*

Supplemental Figure 5. Plasmid constructs used in this study. Transcription factor coding sequences were cloned into the Thermo Fisher Scientific pENTR/D-TOPO vector to generate Gateway-compatible entry constructs (**a**). Gateway LR cloning was then conducted into the ImpGWB405 vector to generate transcription factor C-terminus GFP fusions for subcellular localization assays (**b**). Transcription factor stop codons were omitted for construction of these fusions. Transcription factor plasmids illustrated below are not necessarily to scale as a generic insert is used to illustrate the fusion produced.

a **b**





Chapter 4. Concluding remarks

Authors:

Travis M. Garcia[1]

[1]Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV 89557-0330 USA

Concluding remarks

Taken together, the research presented in this dissertation attempts to advance the current understanding of the transcriptional regulation of crassulacean acid metabolism using the CAM model *Kalanchoe* with the aspiration that ultimately this research will aid in the engineering of the CAM pathway into important C_3 crops. Borland et al. (2014) and Yang et al. (2015) outlined a multi-step approach to accomplishing this goal through synthetic biology. Initially, a "parts library" consisting of individual genes, promoters, and other regulatory elements will need to be identified. Systems modeling will then provide CAM "circuit" blueprints through integration of multi-omics data sets. Computational modeling of CAM metabolics, fluxes, and energy balance such as that reported by Shameer et al. (2018) and Topfer et al. (2020) will be a key part of this step. Assembly of large complex multi-gene constructs followed by transformation into C_3 species can then be carried out. Lastly, extensive phenotypic evaluation will identify aspects of the engineering requiring optimization or reassessment.

The research presented here can be viewed within the context of contributing to the creation of a CAM parts library. Although ChIP-seq analysis of *KfNF-YB3*, *KfHD-like*, and *KfMYB59* was unsuccessful, the diel transcript profile of these transcription factors combined with their high induction in CAM suggests that these transcription factors warrant further research as to their potential functions in the transcriptional control of CAM-related processes. To our knowledge, ChIP-seq has never been conducted successfully on any *Kalanchoe* species and it seems likely that this lack of an optimized experimental workflow may be responsible for the failure of these experiments. In the coming years the increasingly lower costs of high-throughput

sequencing will likely motivate more ChIP-seq and other *in vivo* genome-wide protein-DNA interaction studies in *Kalanchoe* and related CAM species. Hence, CAM species-optimized protocols and techniques are likely to be developed. Alternatively, *in vitro* approaches such as Protein-Binding Microarrays (PBMs) (Berger & Bulyk, 2009; Franco- Zorrilla et al., 2014), DNA affinity purification sequencing (DAP-seq) (O' Malley et al., 2016; Gallie et al., 2018; Zhu et al., 2021) and systematic evolution of ligands by exponential enrichment followed by sequencing (SELEX-seq) (Jolma et al., 2010) are powerful and reliable techniques for transcription factor binding site profiling. These approaches both avoid some of the technical challenges of ChIP-seq and related *in vivo* techniques, while providing data that largely agrees with *in vivo*-derived binding site data (Bartlett et al., 2017; Orenstein & Shamir, 2014). The results of future *in vivo* and *in vitro* transcription factor binding studies will be critical in uncovering the gene regulatory networks necessary for introducing CAM into C_3 species.

The data presented in this dissertation provides an initial step in identifying the individual *cis*-regulatory elements and their associated interacting transcription factors necessary for expression of the CAM-type *Ppc1* gene. To further confirm the interactions of the Y1H high-scoring transcription factors with the *KfPpc1* promoter and investigate their regulatory effects additional experiments were initially planned. At this time, bait probes for EMSA assays have been successfully generated from PCR-amplified *K. fedtschenkoi* genomic DNA. Collectively covering the entire 3 kb *KfPpc1* promoter used for analysis in chapter 3, probes were produced as a series of 14 individual fragments each of approximately 250 bp with approximately 30 base pairs of overlap between sequential fragments. Following the results of EMSA assays utilizing these

native promoter *cis*-elements, mutant and deleted versions of the *cis*-elements were then to be interrogated using custom-designed oligos to confirm the necessity of the specific motif(s) sequences for protein-DNA interaction. Recombinant 6xHis-tagged proteins were to be used for said EMSA assays in a manner similar to Wu et al. (2020) which used recombinant polyhistidine-tagged ERF11 in EMSA assays with probes based on the sequences of peaks derived from ChIP-seq data. Production and purification of recombinant 6xHis-tagged KfERF9, KfERF106, and KfTCP4 was first attempted through transient expression in *N. benthamiana* following a technique similar to Popescu et al. (2007). However, no detectable recombinant protein for any of these transcription factors was detected in either crude protein extracts or Ni-NTA resin-purified samples processed under both native and denaturing conditions. Following this failure, 6xHis-tagged protein expression was then attempted in *E. coli* using a T7*lac* promoter IPTG-inducible system (Studier et al., 1990). Once again, no detectable recombinant protein for any of these transcription factors was detected in crude extracts or Ni-NTA resin-purified samples processed under both native and denaturing conditions even after repeated modifications to culture induction time and growth temperature. At this time, protein production for use in EMSA assays has not been achieved. We also propose *KfPpc1* promoter-GFP and transcription factor co-expression experiments to examine the regulatory effects of KfERF9, KfERF106, and KfTCP4. Protoplast-based systems using GFP as a reporter are frequently used for gene expression analysis as in the case of Chen et al. (2006) where *O. sativa* protoplasts were used to investigate the regulation of defense-related genes through the measurement of GFP fluorescence following co-transfection. Effects of KfERF9, KfERF106, and KfTCP4 on *KfPpc1* expression could

be investigated through transient co-transfection assays in *Kalanchoe* using a *KfPpc1* promoter-GFP reporter and transcription factor overexpression constructs where putative promoter binding regions are either examined with their native sequence, mutated, or deleted. Although Chen et al., like many other studies of the type, utilized protoplasts for co-transfection assays, protoplast generation has not been optimized for *Kalanchoe* species. Hence, a leaf co-infiltration approach may be the best option currently available even though the efficiency of transient leaf tissue expression in *Kalanchoe* is limited.

As one of the core components of the CAM carboxylation module whose transcript and protein abundance are correlated with the magnitude of CAM (Liu et al., 2018) identifying the "parts list" of *cis*-regulatory elements (CREs) for *Ppc1* and their *trans*-regulatory factors necessary for correct spatiotemporal expression will be of particular importance in CAM engineering. Indeed, the CAM-type *Ppc1* gene is required not only for nocturnal CO_2 fixation and accumulation of malate, but also for the normal function of the circadian clock and stomatal responses (Boxall et al., 2020). The approaches of *in-silico* predictive binding site modeling and Y1H assays applied to *KfPpc1* in this study should prove viable strategies for dissecting the CREs of other CAM pathway genes. Currently, the power of *in-silico* modeling for *Kalanchoe*, like other CAM species, lags behind that of popular C_3 plant models. However, this situation should improve as more omics datasets and protein-DNA interaction data becomes available for *Kalanchoe* and related CAM species. Recently, the concept of artificial intelligence-driven machine learning has been applied to *in-silico* transcription factor-DNA interaction prediction. These novel approaches appear promising as early validation results suggest that they offer improved prediction accuracy over traditional

algorithms and can more effectively model the effects of SNP variations at or near transcription factor binding sites (Chen et al., 2021; Jin et al., 2021; Kim et al., 2021; Zhang et al., 2021). Emerging *in-silico* techniques such as these might prove to be valuable tools in computer-based modeling of the transcriptional regulation of CAM and in understanding the differences in the use of *cis*-elements between the C_3 photosynthesis and CAM states. Y1H assays will likely continue to be a powerful tool for *in-silico* prediction validation and DNA-centered studies on gene regulation within the CAM research community. Although large-scale Y1H assays using a whole-genome cDNA library, such as that presented in this study, are still laborious and time consuming, the increasing availability and cost-effectiveness of automation and robotics should encourage more studies of this kind on other components of the CAM pathway. Hopefully, through extensive transcription factor-centered and DNA-centered studies and the integration of these results with genomic, transcriptomic and proteomic data a full picture of the genetic regulatory control of CAM may one day be achieved.

The future prospects for *Kalanchoe laxiflora* and, in particular, the diploid *Kalanchoe fedtschenkoi,* as model organisms for understanding CAM appear promising. At this time both of these species have a sequenced genome and transcriptome (Phytozome info: K.fedtschenkoi v1.1 (doe.gov); Phytozome info: K.laxiflora v1.1 (doe.gov)). Other omics resources are rapidly becoming available for *Kalanchoe fedtschenkoi*. Recently, Abraham et al. (2020) reported a comparison between the mesophyll and epidermal proteomes across a diel time course for *K. fedtschenkoi* and linked these differing proteomes to tissue-specific enzymatic processes. Studies such as this will provide an important bridge to understanding how the underlying transcriptional

circuits of CAM give rise to the divergent physiological responses of CAM *vs*. C_3 photosynthesis species. The ease with which stable transformation and regeneration can be achieved with *K. fedtschenkoi* is also facilitating its use in site-directed mutagenesis studies designed to functionally dissect components of the CAM pathway. Liu et al. (2019) reported a CRISPR/Cas9-based mutational analysis of the blue light photoreceptor *KfePHOT2*, which elucidated the importance of this receptor to stomatal physiology in CAM. In addition, this study established the first CRISPR/Cas9 "toolkit" in a CAM species in the form of a database of guide RNAs for targeting all known protein coding transcripts in *K. fedtschenkoi*. Resources such as this combined with the ever-improving technologies with respect to targeted gene editing and guide RNA multiplexing will prove valuable in the understanding of CAM from a functional genomics perspective and *K. fedtschenkoi* appears to be well-suited for these prospective studies.

References

Abraham, P. E., Hurtado Castano, N., Cowan-Turner, D., Barnes, J., Poudel, S., Hettich, R., Flütsch, S., Santelia, D., & Borland, A. M. (2020). Peeling back the layers of crassulacean acid metabolism: Functional differentiation between *Kalanchoë fedtschenkoi* epidermis and mesophyll proteomes. *The Plant Journal*, *103*(2), 869–888. https://doi.org/10.1111/tpj.14757

Bartlett, A., O'Malley, R. C., Huang, S. C., Galli, M., Nery, J. R., Gallavotti, A., & Ecker, J. R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols*, *12*(8), 1659–1672. https://doi.org/10.1038/nprot.2017.055

Berger, M. F., & Bulyk, M. L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols*, *4*(3), 393–411. https://doi.org/10.1038/nprot.2008.195

Borland, A. M., Hartwell, J., Weston, D. J., Schlauch, K. A., Tschaplinski, T. J., Tuskan, G. A., Yang, X., & Cushman, J. C. (2014). Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends in Plant Science*, *19*(5), 327–338. https://doi.org/10.1016/j.tplants.2014.01.006

Boxall, S. F., Kadu, N., Dever, L. V., Kneřová, J., Waller, J. L., Gould, P. J. D., & Hartwell, J. (2020). *Kalanchoë* PPC1 Is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and Guard Cell Signaling Genes. *The Plant Cell*, *32*(4), 1136–1160. https://doi.org/10.1105/tpc.19.00481

Chen, C., Hou, J., Shi, X., Yang, H., Birchler, J. A., & Cheng, J. (2021). DeepGRN: Prediction of transcription factor binding site across cell-types using attention-based deep neural networks. *BMC Bioinformatics*, *22*(1), 38. https://doi.org/10.1186/s12859-020-03952-1

Chen, S., Tao, L., Zeng, L., Vega-Sanchez, M. E., Umemura, K., & Wang, G.-L. (2006). A highly efficient transient protoplast system for analyzing defence gene expression and protein?protein interactions in rice. *Molecular Plant Pathology*, *7*(5), 417–427. https://doi.org/10.1111/j.1364-3703.2006.00346.x

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., & Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences*, *111*(6), 2367–2372. https://doi.org/10.1073/pnas.1316278111

Galli, M., Khakhar, A., Lu, Z., Chen, Z., Sen, S., Joshi, T., Nemhauser, J. L., Schmitz, R. J., & Gallavotti, A. (2018). The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nature Communications*, *9*(1), 4526. https://doi.org/10.1038/s41467-018-06977-6

Jin, Y., Jiang, J., Wang, R., & Qin, Z. S. (2021). Systematic Evaluation of DNA Sequence Variations on *in vivo* Transcription Factor Binding Affinity. *Frontiers in Genetics*, *12*, 667866. https://doi.org/10.3389/fgene.2021.667866

Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J. M., Yan, J., Sillanpaa, M. J., Bonke, M., Palin, K., Talukder, S., Hughes, T. R., Luscombe, N. M., Ukkonen, E., & Taipale, J. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Research*, *20*(6), 861–873. https://doi.org/10.1101/gr.100552.109

Kim, G. B., Gao, Y., Palsson, B. O., & Lee, S. Y. (2021). DeepTFactor: A deep learning-based tool for the prediction of transcription factors. *Proceedings of the National Academy of Sciences*, *118*(2), e2021171118. https://doi.org/10.1073/pnas.2021171118

Liu, D., Chen, M., Mendoza, B., Cheng, H., Hu, R., Li, L., Trinh, C. T., Tuskan, G. A., & Yang, X. (2019). CRISPR/Cas9-mediated targeted mutagenesis for functional genomics research of crassulacean acid metabolism plants. *Journal of Experimental Botany*, *70*(22), 6621–6629. https://doi.org/10.1093/jxb/erz415

Liu, D., Palla, K. J., Hu, R., Moseley, R. C., Mendoza, C., Chen, M., Abraham, P. E., Labbé, J. L., Kalluri, U. C., Tschaplinski, T. J., Cushman, J. C., Borland, A. M., Tuskan, G. A., & Yang, X. (2018). Perspectives on the basic and applied aspects of crassulacean acid metabolism (CAM) research. *Plant Science*, *274*, 394–401. https://doi.org/10.1016/j.plantsci.2018.06.012

O'Malley, R. C., Huang, S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., & Ecker, J. R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*, *165*(5), 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Orenstein, Y., & Shamir, R. (2014). A comparative analysis of transcription factor binding models learned from PBM, HT-SELEX and ChIP data. *Nucleic Acids Research*, *42*(8), e63–e63. https://doi.org/10.1093/nar/gku117

Popescu, S. C., Popescu, G. V., Bachan, S., Zhang, Z., Seay, M., Gerstein, M., Snyder, M., & Dinesh-Kumar, S. P. (2007). Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays. *Proceedings of the National Academy of Sciences*, *104*(11), 4730–4735. https://doi.org/10.1073/pnas.0611615104

Shameer, S., Baghalian, K., Cheung, C. Y. M., Ratcliffe, R. G., & Sweetlove, L. J. (2018). Computational analysis of the productivity potential of CAM. *Nature Plants*, *4*(3), 165–171. https://doi.org/10.1038/s41477-018-0112-2

Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in enzymology*, *185*, 60–89. https://doi.org/10.1016/0076-6879(90)85008-c

Töpfer, N., Braam, T., Shameer, S., Ratcliffe, R. G., & Sweetlove, L. J. (2020). Alternative Crassulacean Acid Metabolism Modes Provide Environment-Specific Water-Saving Benefits in a Leaf Metabolic Model. *The Plant Cell*, *32*(12), 3689–3705. https://doi.org/10.1105/tpc.20.00132

Wu, J., Deng, Y., Hu, J., Jin, C., Zhu, X., & Li, D. (2020). Genome-wide analyses of direct target genes of an ERF11 transcription factor involved in plant defense against bacterial pathogens. *Biochemical and Biophysical Research Communications*, *532*(1), 76–81. https://doi.org/10.1016/j.bbrc.2020.07.073

Yang, X., Cushman, J. C., Borland, A. M., Edwards, E. J., Wullschleger, S. D., Tuskan, G. A., Owen, N. A., Griffiths, H., Smith, J. A. C., De Paoli, H. C., Weston, D. J., Cottingham, R., Hartwell, J., Davis, S. C., Silvera, K., Ming, R., Schlauch, K., Abraham, P., Stewart, J. R., Guo H-B, Albion RA, Ha J, Lim SD, Wone BWM, Yim WC, Garcia T, Mayer JA, Petereit J, Nair SS, Casey E, Hettich RL, Ceusters J, Ranjan P, Palla KJ, Yin H, Reyes-García C, Andrade JL, Freschi L, Beltran JD, Dever LV, Boxall SF, Waller J, Davies J, Bupphada P, Kadu N, Winter K, Sage RF, Aguilar CN, Schmutz J, Jenkins J, Holtum, J. A. M. (2015). A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist*, *207*(3), 491–504. https://doi.org/10.1111/nph.13393

Zhang, Y., Mo, Q., Xue, L., & Luo, J. (2021). Evaluation of deep learning approaches for modeling transcription factor sequence specificity. *Genomics*, *113*(6), 3774–3781. https://doi.org/10.1016/j.ygeno.2021.09.009

Zhu, C.-C., Wang, C.-X., Lu, C.-Y., Wang, J.-D., Zhou, Y., Xiong, M., Zhang, C.-Q., Liu, Q.-Q., & Li, Q.-F. (2021). Genome-Wide Identification and Expression Analysis of OsbZIP09 Target Genes in Rice Reveal Its Mechanism of Controlling Seed Germination. *International Journal of Molecular Sciences*, *22*(4), 1661. https://doi.org/10.3390/ijms22041661

Appendix Table 1. Full list of prey clone interactions for all yeast one-hybrid bait probes. Results are sorted by bait probe then by highest to lowest PBS score category. Redundant binding interactions corresponding to the same *Kalanchoe fedtschenkoi* transcript ID have been eliminated. The number of redundant occurences per transcript is indicated in a separate column.

Y1H bait probe	**Global PBS score**	**Number of redundant predictions in screen**	***K. fedtschenkoi* transcript ID**	***Arabidopsis* gene ID**	***Arabidopsis* gene name**	***Arabidopsis* description**
evening element (EE)	C	2	Kaladp0015s0142.1	AT4G00730.1	AHDP	Encodes a homeodomain protein of the HD-GLABRA2 group. Involved in the accumulation of anthocyanin and in root development. Loss of function mutants have increased cell wall polysaccharide content.
	C	2	Kaladp0018s0211.1	AT4G38900.1	BASIC LEUCINE-ZIPPER 29 (BZIP29)	Basic-leucine zipper (bZIP) transcription factor family protein
	D	1	Kaladp0059s0129.1	AT2G36720.1	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain-containing protein	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain-containing protein

GATA box	A	28	Kaladp0403s0001.1	AT5G07580.1	ERF106	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5.
	B	5	Kaladp0063s0022.1	AT1G49010.1	MYBS1	Duplicated homeodomain-like superfamily protein
	B	1	Kaladp0039s0361.1	AT1G49010.1	MYBS1	Duplicated homeodomain-like superfamily protein
	C	1	Kaladp0011s0947.1	AT2G05100.1	LIGHT-HARVESTING CHLOROPHYLL B-BINDING 2 (LHCB2)	Lhcb2.1 protein encoding a subunit of the light harvesting complex II. Member of a gene family with high degree of sequence similarity. Initially LHCB2.3 was considered as a separate gene but appears to be an allele of LHCB2.1.
	C	1	Kaladp0059s0024.1	AT2G34430.1	LIGHT-HARVESTING CHLOROPHYLL-PROTEIN COMPLEX II SUBUNIT B1 (DEG11)	Photosystem II type I chlorophyll a/b-binding protein The mRNA is cell-to-cell mobile.

	C	3	Kaladp0006s0005.1	AT1G19000.1	Homeodomain-like superfamily protein	Homeodomain-like superfamily protein
	C	3	Kaladp0102s0112.1	AT2G25760.2	AEL1	Casein kinase involved in phosphorylation and ubiquination of RYR/PYLs, resulting in negative regulation of ABA response. Plays a role in repressing the transition from vegetative to reproductive phase
	D	1	Kaladp0023s0108.1	AT1G10140.1	Uncharacterized conserved protein UCP031279	Uncharacterized conserved protein UCP031279
	D	1	Kaladp1321s0003.1	AT1G43760.1	DNAse I-like superfamily protein	DNAse I-like superfamily protein
	D	2	Kaladp0047s0252.1	AT1G78900.1	VACUOLAR ATP SYNTHASE SUBUNIT A (VHA-A)	Encodes catalytic subunit A of the vacuolar ATP synthase. Mutants are devoid of vacuolar ATPase activity as subunit A is encoded only by this gene and show strong defects in male gametophyte development and in Golgi stack morphology. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0008s0486.1	AT5G49650.1	XYLULOSE KINASE 2 (XK-2)	Encodes a cytosolic protein capable of phosphorylating xylulose and deoxy-xylulose. It

						most likely plays a role in producing precursors for isoprenoid biosynthesis.
	D	1	Kaladp0024s0017.1	AT2G25760.2	ARABIDOPSIS EL1-LIKE 1 (AEL1)	Casein kinase involved in phosphorylation and ubiquination of RYR/PYLs, resulting in negative regulation of ABA response. Plays a role in repressing the transition from vegetative to reproductive phase.
	D	1	Kaladp0053s0327.1	AT1G10840.1	TRANSLATION INITIATION FACTOR 3 SUBUNIT H1 (TIF3H1)	Encodes eukaryotic initiation factor 3H1 subunit (TIF3H1).
	D	3	Kaladp0053s0429.1	AT5G58900.1	DIVARICATA1 (DIV1)	R-R-type MYB protein
	D	2	Kaladp1295s0017.1	AT5G47390.1	KUODA1 (CHINESE FOR ENLARGE OR EXPAND)	Encodes a circadian-regulated transcription factor which specifically controls cell expansion during leaf development by controlling ROS homeostasis. The mRNA is cell-to-cell mobile.
	D	2	Kaladp0959s0007.1	AT1G01620.1	PLASMA MEMBRANE INTRINSIC PROTEIN 1;3 (PIP1;3)	a member of the plasma membrane intrinsic protein subfamily PIP1. localizes to the plasma membrane and exhibits water transport activity in

						Xenopus oocyte. expressed ubiquitously and protein level decreases slightly during leaf development. Involved redundantly with PIP1;1/2/4/5 in hydraulics and carbon fixation, regulates the expression of related genes that affect plant growth and development.
	D	1	Kaladp0011s1131.1	AT2G08665.1	Natural antisense transcript overlaps with AT2G34420	Natural antisense transcript overlaps with AT2G34420
	D	1	Kaladp0087s0025.1	AT1G10760.1	GWD	Encodes an α-glucan, water dikinase required for starch degradation. Involved in cold-induced freezing tolerance. Mutations that eliminate the GWD protein or affect the dikinase domain of the enzyme dramatically reduce both the amount of phosphate in the amylopectin and the rate of starch degradation. Mature leaves of these mutants accumulate amounts of starch

						up to seven times greater than those in wild-type leaves. NMR analysis of the mutants, suggests that the gene is specifically involved in the phosphorylation of the glucosyl residues of starch at the C6 position.
	D	1	Kaladp0098s0126.1	AT3G05190.1	D-aminoacid aminotransferase-like PLP-dependent enzymes superfamily protein	D-aminoacid aminotransferase-like PLP-dependent enzymes superfamily protein
	D	1	Kaladp0095s0055.1	AT1G53310.1	PEP(PHOSPHOENOLPYRUVATE) CARBOXYLASE 1 (ATPEPC1)	Encodes one of four Arabidopsis phosphoenolpyruvate carboxylase proteins.Plays an important role in carbon and nitrogen metabolism.
	D	1	Kaladp0048s0346.1	AT1G60970.1	SNARE-like superfamily protein	SNARE-like superfamily protein
	D	1	Kaladp0006s0004.1	AT3G08720.1	ARABIDOPSIS THALIANA PROTEIN KINASE 19 (ATPK19)	Encodes a ribosomal-protein S6 kinase. Gene expression is induced by cold and salt (NaCl). Activation of AtS6k is regulated by 1-naphthylacetic acid and kinetin, at least in part, via a lipid kinase-dependent

						pathway. Phosphorylates specifically mammalian and plant S6 at 25 degrees C but not at 37 degrees C. Involved in translational up-regulation of ribosomal proteins.
	D	1	Kaladp0042s0238.1	AT3G17410.1	PUTATIVE RECEPTOR-LIKE CYTOPLASMIC KINASE 1 (CARK1)	Positively regulates ABA-mediated physiological responses via phosphorylation on RCAR3/ RCAR11.
	D	1	Kaladp0822s0005.1	AT2G05300.1	pseudogene of the F-box protein family	pseudogene of the F-box protein family
	D	1	Kaladp0403s0001.1	AT5G07580.1	ERF106	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5.
	D	1	Kaladp0032s0450.1	AT4G38770.1	ARABIDOPSIS THALIANA PROLINE-RICH PROTEIN 4 (ATPRP4)	Encodes one of four proline-rich proteins in Arabidopsis which are predicted to localize to the cell wall. Transcripts are most abundant in aerial organs of the plant.

	D	1	Kaladp0095s0716.1	AT5G07580.1	ERF106	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5.
	D	1	Kaladp0071s0417.1	AT2G02170.1	Remorin family protein	Remorin family protein
	D	1	Kaladp0040s0122.1	AT1G01630.1	Sec14p-like phosphatidylinositol transfer family protein	Sec14p-like phosphatidylinositol transfer family protein
	D	1	Kaladp0011s0747.1	AT3G03770.1	Leucine-rich repeat protein kinase family protein	Leucine-rich repeat protein kinase family protein
	D	1	Kaladp0049s0038.1	AT3G19240.1	Vacuolar import/degradation, Vid27-related protein	Vacuolar import/degradation, Vid27-related protein
	D	1	Kaladp0047s0023.1	AT4G38950.1	ATP binding microtubule motor family protein	ATP binding microtubule motor family protein
	D	1	Kaladp0008s0232.1	AT1G64350.1	SEH1H	seh1-like protein
	D	1	Kaladp0019s0123.1	AT1G18335.1	Acyl-CoA N-acyltransferases (NAT) superfamily protein	Acyl-CoA N-acyltransferases (NAT) superfamily protein
	D	1	Kaladp0050s0276.1	AT3G25680.1	SLH domain protein	SLH domain protein
	D	1	Kaladp0047s0023.1	AT4G38950.1	ATP binding microtubule motor family protein	ATP binding microtubule motor family protein

	D	1	Kaladp0048s0538.1	AT1G53730.2	STRUBBELIG-RECEPTOR FAMILY 6 (SRF6)	STRUBBELIG-receptor family 6
	D	1	Kaladp0053s0475.1	AT3G18520.2	HISTONE DEACETYLASE 15 (ATHDA15)	Encodes a protein with similarity to histone deacetylases. The histone deacetylase domain of HDA15 (HDA15HD) assembles as tetrameric forms with each monomer composed of 12 alpha-helices and 9 beta-sheets (DOI: 10.1104/pp.20.00604).Plants expressing RNAi directed against this gene show a moderate resistance to agrobacterium-mediated root transformation. Class II RPD3-like family HDAC member which controls negative responses to salinity stress.
	D	1	Kaladp0008s0478.1	AT5G22640.1	EMBRYO DEFECTIVE 1211 (EMB1211)	EMB1211 is a MORN (multiple membrane occupation and recognition nexus) motif containing protein involved in embryo development and chloroplast biogenesis. The

						mRNA is cell-to-cell mobile.
	D	1	Kaladp0674s0030.1	AT5G63160.1	BTB AND TAZ DOMAIN PROTEIN 1 (BT1)	BTB and TAZ domain protein. Short-lived nuclear-cytoplasmic protein targeted for degradation by the 26S proteosome pathway. Acts redundantly with BT2 and BT3 during female gametophyte development.
	D	1	Kaladp0018s0034.1	AT5G08060.1	furry	furry
	D	1	Kaladp0008s0191.1	AT4G10080.1	transmembrane protein	transmembrane protein
	D	1	Kaladp0042s0147.1	AT5G41400.1	RING/U-box superfamily protein	RING/U-box superfamily protein
	D	1	Kaladp0045s0054.1	AT4G10340.1	LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5 (LHCB5)	photosystem II encoding the light-harvesting chlorophyll a/b binding protein CP26 of the antenna system of the photosynthetic apparatus The mRNA is cell-to-cell mobile.
	D	3	Kaladp0863s0010.1	AT3G16370.1	GDSL-motif esterase/acyltransferase/lipase	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or

						hydrolase reactions with lipid and non-lipid substrates. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0039s0748.1	AT5G25610.1	RESPONSIVE TO DESICCATION 22 (ATRD22)	responsive to dehydration 22 (RD22) mediated by ABA
	D	1	Kaladp0021s0003.1	AT5G09890.2	NUCLEAR DBF2-RELATED 8 (NDR8)	Protein kinase family protein
	D	1	Kaladp0032s0009.1	AT1G62750.1	SNOWY COTYLEDON 1 (ATSCO1)	Nuclear encoded protein consists of the five domains conserved in EF-G proteins, with two GTP-binding sites in the first domain, and an additional transit peptide at the N-terminus. Localized in chloroplasts. Point mutation results in a delay in the onset of germination. At early developmental stage embryos still contain undifferentiated proplastids. The greening of cotyledons is severely impaired in light-grown mutant sco1 seedlings, whereas the following true leaves develop normally as in wild-type plants.

	D	1	Kaladp0071s0090.1	AT2G03150.1	EMBRYO DEFECTIVE 1579 (EMB1579)	Encodes a nuclear-localized calcium-binding protein RSA1 (SHORT ROOT IN SALT MEDIUM 1), which is required for salt tolerance.
	D	7	Kaladp0016s0384.1	AT4G26210.1	Mitochondrial ATP synthase subunit G protein	Mitochondrial ATP synthase subunit G protein
	D	1	Kaladp0040s0529.1	AT2G43640.1	Signal recognition particle, SRP9/SRP14 subunit	Signal recognition particle, SRP9/SRP14 subunit
	D	1	Kaladp0070s0153.1	AT5G06000.1	ARABIDOPSIS THALIANA EUKARYOTIC TRANSLATION INITIATION FACTOR 3G2 (ATEIF3G2)	One of the 2 genes that code for the G subunit of eukaryotic initiation factor 3 (EIF3).
	D	1	Kaladp0053s0058.1	AT2G34510.1	ATHB-1	Protein of unknown function, DUF642. Found in cellulose enriched cell wall fractions.
	D	1	Kaladp0032s0450.1	AT4G38770.1	ARABIDOPSIS THALIANA PROLINE-RICH PROTEIN 4 (ATPRP4)	Encodes one of four proline-rich proteins in Arabidopsis which are predicted to localize to the cell wall. Transcripts are most abundant in aerial organs of the plant.
	D	1	Kaladp0001s0056.1	AT1G75700.1	HVA22-LIKE PROTEIN G	HVA22-like protein G
	D	1	Kaladp0008s0049.1	AT2G30390.2	FERROCHELATASE 2 (ATFC-II)	Encodes one of two ferrochelatase genes in Arabidopsis. Ferrochelatase is

						the terminal enzyme of heme biosynthesis. FC-II is speculated to operate in photosynthetic cytochromes.
	D	1	Kaladp0019s0067.1	AT5G52040.2	ARGININE/SERINE-RICH SPLICING FACTOR 41 (AT-RS41)	Encodes an arginine/serine-rich splicing factor. Transcript is alternatively spliced and is differentially expressed in different tissues (flowers, roots, stems, and leaves) examined. Barta et al (2010) have proposed a nomenclature for Serine/Arginine-Rich Protein Splicing Factors (SR proteins): Plant Cell. 2010, 22:2926. RS41 binds to HYL1 and co-localizes to the nuclear dicing body. Along with RS41, it appears to be involved in pri-miRNA processing and miRNA biogenesis.
	D	1	Kaladp0022s0232.1	AT3G27380.1	SUCCINATE DEHYDROGENASE 2-1 (SDH2-1)	One of three isoforms of the iron-sulfur component of the succinate dehydrogenase complex, a component of the

						mitochondrial respiratory chain complex II. The product of the nuclear encoded gene is imported into the mitochondrion. Expressed during germination and post-germinative growth.
	D	1	Kaladp0036s0066.1	AT5G27120.1	SAR DNA-binding protein	SAR DNA-binding protein, putative, strong similarity to SAR DNA-binding protein-1 (Pisum sativum) GI:3132696; contains Pfam profile PF01798: Putative snoRNA binding domain; has similarity to MAR binding NOP58 protein The mRNA is cell-to-cell mobile.
	D	1	Kaladp0045s0206.1	AT2G40080.1	EARLY FLOWERING 4 (ELF4)	Encodes a novel nuclear 111 amino-acid phytochrome-regulated component of a negative feedback loop involving the circadian clock central oscillator components CCA1 and LHY. ELF4 is necessary for light-induced expression of both CCA1 and LHY, and conversely, CCA1 and LHY act

						negatively on light-induced ELF4 expression. ELF4 promotes clock accuracy and is required for sustained rhythms in the absence of daily light/dark cycles. It is involved in the phyB-mediated constant red light induced seedling de-etiolation process and may function to coregulate the expression of a subset of phyB-regulated genes.
	D	1	Kaladp0055s0107.1	AT4G38350.2	ATNPC1-2	Patched family protein
	D	1	Kaladp0076s0290.1	AT4G15910.1	DROUGHT-INDUCED 21 (ATDI21)	encodes a gene whose transcript level in root and leaves increases to progressive drought stress. The transcript level is also affected by changes of endogenous or exogenous abscisic acid level. It appears to be a member of plant-specific gene family that includes late embryo-abundant and zinc- IAA-induced proteins in other plants. The

						mRNA is cell-to-cell mobile.
	D	1	Kaladp0095s0559.1	AT1G18730.1	NDH DEPENDENT FLOW 6 (NDF6)	likely a subunit of the chloroplast NAD(P)H dehydrogenase complex, involved in PSI cyclic electron transport. Located on the thylakoid membrane. Mutant has impaired NAD(P)H dehydrogenase activity. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0080s0159.1	AT5G57910.1	ribosomal RNA small subunit methyltransferase G	ribosomal RNA small subunit methyltransferase G
	D	1	Kaladp0056s0057.1	AT2G26280.1	CTC-INTERACTING DOMAIN 7 (CID7)	smr (Small MutS Related) domain-containing protein
	D	1	Kaladp0040s0712.1	AT3G08820.1	Pentatricopeptide repeat (PPR) superfamily protein	Pentatricopeptide repeat (PPR) superfamily protein
	D	3	Kaladp0008s0071.1	AT4G11380.2	Adaptin family protein	Adaptin family protein
	D	1	Kaladp0515s0177.1	AT3G51600.1	LIPID TRANSFER PROTEIN 5 (LTP5)	Predicted to encode a PR (pathogenesis-related) protein. Belongs to the lipid transfer protein (PR-14) family with the following members: At2g38540/LTP1, At2g38530/LTP2, At5g59320/LTP3,

						At5g59310/LTP4, At3g51600/LTP5, At3g08770/LTP6, At2g15050/LTP7, At2g18370/LTP8, At2g15325/LTP9, At5g01870/LTP10, At4g33355/LTP11, At3g51590/LTP12, At5g44265/LTP13, At5g62065/LTP14, At4g08530/LTP15.
	D	1	Kaladp0070s0058.1	AT2G26330.1	ERECTA	Homologous to receptor protein kinases. Involved in specification of organs originating from the shoot apical meristem. Contains a cytoplasmic protein kinase catalytic domain, a transmembrane region, and an extracellular leucine-rich repeat. ER has been identified as a quantitative trait locus for transpiration efficiency by influencing epidermal and mesophyll development, stomatal density and porosity of leaves. It has been implicated in resistance to the bacterium Ralstonia solanacearum and

						to the necrotrophic fungus Plectosphaerella cucumerina. Together with ERL1 and ERL2, ER governs the initial decision of protodermal cells to either divide proliferatively to produce pavement cells or divide asymmetrically to generate stomatal complexes. ER binds to the peptides STOMAGEN and EPF2 which compete for the same binding site. The ER-EFP2 complex activates MAPK signaling that inhibits stomatal development. ER-STOMAGEN does not activate MAPK signaling. Plants harboring loss of function alleles of er are more susceptible to heat stress than wild type. In Arabidopsis and other organisms, overexpression of ER confers thermotolerance via as yet undefined mechanisms.

	D	1	Kaladp0043s0275.1	AT3G16350.1	NITROGEN RESPONSE DEFICIENCY 1 (NID1)	MYB-like transcription factor involved in nitrate signaling trough regulation of CHL1.
	D	1	Kaladp0011s0198.1	AT3G14470.1	NB-ARC domain-containing disease resistance protein	NB-ARC domain-containing disease resistance protein
	D	1	Kaladp0005s0043.1	AT1G07705.2	NEGATIVE ON TATA LESS2A (NOT2A)	NOT2 / NOT3 / NOT5 family
	D	1	Kaladp0043s0281.1	AT3G16240.1	AQP1	Delta tonoplast intrinsic protein, functions as a water channel and ammonium (NH3) transporter. Highly expressed in flower, shoot, and stem. Expression shows diurnal regulation and is induced by ammonium (NH3). Protein localized to vacuolar membrane. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0080s0162.1	AT4G30950.1	FATTY ACID DESATURASE 6 (FAD6)	Chloroplastic enzyme responsible for the synthesis of 16:2 and 18:2 fatty acids from galactolipids, sulpholipids and phosphatidylglycerol. Uses ferredoxin as electron donor. Gene mutation resulted in reduced level of unsaturated fatty

						acids leading to susceptibility to photoinhibition.
	D	1	Kaladp0033s0100.1	AT1G80130.1	Tetratricopeptide repeat (TPR)-like superfamily protein	Tetratricopeptide repeat (TPR)-like superfamily protein
	D	1	Kaladp0048s0813.1	AT4G12460.1	OSBP(OXYSTEROL BINDING PROTEIN)-RELATED PROTEIN 2B (ORP2B)	OSBP(oxysterol binding protein)-related protein 2B
	D	1	Kaladp0068s0369.1	AT1G20020.1	FERREDOXIN-NADP(+)-OXIDOREDUCTASE 2 (ATLFNR2)	Encodes a leaf-type ferredoxin:NADP(H) oxidoreductase. It is present in both chloroplast stroma and thylakoid membranes but is more abundant in the stroma The mRNA is cell-to-cell mobile.
	D	1	Kaladp0024s0136.1	AT1G20020.1	FERREDOXIN-NADP(+)-OXIDOREDUCTASE 2 (ATLFNR2)	Encodes a leaf-type ferredoxin:NADP(H) oxidoreductase. It is present in both chloroplast stroma and thylakoid membranes but is more abundant in the stroma The mRNA is cell-to-cell mobile.
	D	2	Kaladp0102s0020.1	AT1G04420.1	NAD(P)-linked oxidoreductase superfamily protein	NAD(P)-linked oxidoreductase superfamily protein
	D	5	Kaladp0085s0147.1	AT4G29680.1	Alkaline-phosphatase-like family protein	Alkaline-phosphatase-like family protein

	D	1	Kaladp0087s0188.1	AT3G57680.1	Peptidase S41 family protein	Peptidase S41 family protein
	D	1	Kaladp0064s0112.1	AT3G13670.1	ARABIDOPSIS EL1-LIKE 2 (AEL2)	MUT9-like protein kinase. Contributes to phosphorylation of photoexcited CRY2. Interaction with CRY2 occurs via the non catalytic PPKC domain.MLK4 phosphorylates the conserved H2A serine 95 residue. Synthetic mutants that cannot phosphorylate H2AS95 fail to complement the late flowering phenotype suggesting that MLK4 promotes long day flowering via phosphorylation.MLK4 is required for H2A295 phosphorylation of GI.
	D	1	Kaladp0070s0045.1	AT1G32860.1	Glycosyl hydrolase superfamily protein	Glycosyl hydrolase superfamily protein
	D	1	Kaladp0092s0024.1	AT2G17840.1	EARLY-RESPONSIVE TO DEHYDRATION 7 (ERD7)	Identified as drought-inducible gene by differential hybridization. Upregulated by high light, drought, cold and salt stress determined by microarray analysis.

	D	1	Kaladp0022s0218.1	AT2G32720.1	ARABIDOPSIS CYTOCHROME B5 ISOFORM B (ATCB5-B)	Participates with ELO2 in VLCFA synthesis
	D	1	Kaladp0036s0172.1	AT3G13810.2	INDETERMINATE(ID)-DOMAIN 11 (ATIDD11)	indeterminate(ID)-domain 11
	N/A	3	Kaladp0102s0112.1	AT2G25760.2	ARABIDOPSIS EL1-LIKE 1 (AEL1)	Casein kinase involved in phosphorylation and ubiquination of RYR/PYLs, resulting in negative regulation of ABA response. Plays a role in repressing the transition from vegetative to reproductive phase.
	N/A	1	Kaladp0016s0377.1	AT5G35170.1	adenylate kinase family protein	adenylate kinase family protein
	N/A	1	Kaladp0045s0143.1	AT1G04200.1	dyggve-melchior-clausen syndrome protein	dyggve-melchior-clausen syndrome protein
	N/A	1	Kaladp0047s0189.1	AT5G49720.1	DEFECTIVE CYTOKINESIS (DEC)	Encodes a membrane-bound endo-1,4-beta-D-glucanase, involved in cellulose biosynthesis. Loss-of-function mutants have severe cellulose-deficient phenotypes. During cell elongation, KOR1 is associated with Golgi apparatus and early endosome. Inhibition of

						cellulose biosynthesis promoted a redistribution of KOR1 in subcellular locations. These observations suggest that deposition of cellulose involves the intracellular cycling of KOR1.
	N/A	1	Kaladp0674s0166.1	AT3G47470.1	LIGHT-HARVESTING CHLOROPHYLL-PROTEIN COMPLEX I SUBUNIT A4 (LHCA4)	Encodes a chlorophyll a/b-binding protein that is more similar to the PSI Cab proteins than the PSII cab proteins. The predicted protein is about 20 amino acids shorter than most known Cab proteins.
	N/A	1	Kaladp0492s0002.1	AT3G54890.1	PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 1 (LHCA1)	Encodes a component of the light harvesting complex associated with photosystem I.
	N/A	1	Kaladp0060s0371.1	AT5G54770.1	THI1	Encodes a thiamine biosynthetic gene that has a dual function in thiamine biosynthesis and mitochondrial DNA damage tolerance. It appears to be involved in producing the thiazole portion of thiamine (vitamin

						B1). A crystal structure of the protein reveals that it forms a 2-ring homo-octamer. The mRNA is cell-to-cell mobile.
	N/A	2	Kaladp1277s0002.1	AT4G21310.1	DEAL2	DUF1218 family member
	N/A	1	Kaladp0081s0301.1	AT5G55190.1	ATRAN3	A member of RAN GTPase gene family. Encodes a small soluble GTP-binding protein. Likely to be involved in nuclear translocation of proteins. May also be involved in cell cycle progression. Role in seed and endosperm development.
	N/A	1	Kaladp0037s0392.1	AT3G56940.1	ACSF	Encodes a putative ZIP protein with varying mRNA accumulation in leaves, stems and roots. Has a consensus carboxylate-bridged di-iron binding site. The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0018s0220.1	AT4G38970.1	FRUCTOSE-BISPHOSPHATE ALDOLASE 2 (ATFBA2)	Protein is tyrosine-phosphorylated and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds.

	N/A	1	Kaladp0048s0369.1	AT5G48300.1	ADP GLUCOSE PYROPHOSPHORYLASE 1 (ADG1)	Encodes the small subunit of ADP-glucose pyrophosphorylase. The small subunit is the catalytic isoform responsible for ADP-glucose pyrophosphorylase activity. The presence of the small subunit is required for large subunit stability. Two isoforms of the small subunit (ApS1 and ApS2) have been described. ApS1 is the major small subunit isoform present in all plant tissues tested. The mRNA is cell-to-cell mobile.
	N/A	3	Kaladp0085s0062.1	AT1G06960.1	U2B"-LIKE (U2BL)	Similar to U2B'. Functions in cold stress induced alternative splicing.
	N/A	1	Kaladp0016s0249.1	AT4G38970.1	FRUCTOSE-BISPHOSPHATE ALDOLASE 2 (ATFBA2)	Protein is tyrosine-phosphorylated and its phosphorylation state is modulated in response to ABA in Arabidopsis thaliana seeds.
	N/A	1	Kaladp0058s0476.1	AT1G43710.1	ARABIDOPSIS THALIANA SERINE DECARBOXYLASE 1 (ATSDC1)	Encodes a serine decarboxylase that is involved in ethanolamine metabolism and is

						crucial for plant growth.
	N/A	1	Kaladp0516s0003.1	AT1G50010.1	TUBULIN ALPHA-2 CHAIN (TUA2)	Encodes alpha-2,4 tubulin. TUA2 and TUA4 encode identical proteins. The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0101s0056.1	AT5G59690.1	Histone superfamily protein	Histone superfamily protein
	N/A	1	Kaladp0095s0103.1	AT5G64860.1	DISPROPORTIONATING ENZYME (ATDPE1)	Encodes a maltotriose-metabolizing enzyme with chloroplastic α-1,4-glucanotransferase activity. Mutant has altered starch degradation.
	N/A	2	Kaladp0008s0534.1	AT2G39810.1	EARLY IN SHORT DAYS 6 (ESD6)	A novel protein with a RING finger motif near the amino terminus. Negative regulator of cold responses. Functions as an E3 ligase required for the ubiquitination of ICE1. HOS1 physically interacts with ICE1 and mediates the ubiquitination of ICE1 both in vitro and in vivo. Overexpression represses the expression of CBFs and their downstream genes and confers increased sensitivity to

						freezing stress. The mRNA is cell-to-cell mobile.
	N/A	17	Kaladp0075s0059.1	AT2G39730.1	RUBISCO ACTIVASE (RCA)	Rubisco activase, a nuclear-encoded chloroplast protein that consists of two isoforms arising from alternative splicing in most plants. Required for the light activation of rubisco. Involved in jasmonate-induced leaf senescence.
	N/A	1	Kaladp0091s0051.1	AT5G55190.1	RAN GTPASE 3 (ATRAN3)	A member of RAN GTPase gene family. Encodes a small soluble GTP-binding protein. Likely to be involved in nuclear translocation of proteins. May also be involved in cell cycle progression. Role in seed and endosperm development.
	N/A	1	Kaladp0064s0112.1	AT3G13670.1	ARABIDOPSIS EL1-LIKE 2 (AEL2)	MUT9-like protein kinase. Contributes to phosphorylation of photoexcited CRY2. Interaction with CRY2 occurs via the non catalytic PPKC domain.MLK4 phosphorylates the conserved H2A serine 95 residue. Synthetic mutants that cannot

						phosphorylate H2AS95 fail to complement the late flowering phenotype suggesting that MLK4 promotes long day flowering via phosphorylation.MLK4 is required for H2A295 phosphorylation of GI.
	N/A	1	Kaladp0674s0124.1	AT5G62670.1	H(+)-ATPASE 11 (AHA11)	H[+]-ATPase 11
	N/A	1	Kaladp0095s0055.1	AT1G53310.1	PEP(PHOSPHOENOLPYRUVATE) CARBOXYLASE 1 (ATPEPC1)	Encodes one of four Arabidopsis phosphoenolpyruvate carboxylase proteins.Plays an important role in carbon and nitrogen metabolism.
	N/A	1	Kaladp0058s0194.1	AT4G14290.1	alpha/beta-Hydrolases superfamily protein	alpha/beta-Hydrolases superfamily protein
	N/A	1	Kaladp0036s0297.1	AT1G06950.1	ARABIDOPSIS THALIANA TRANSLOCON AT THE INNER ENVELOPE MEMBRANE OF CHLOROPLASTS 110 (ATTIC110)	Encodes a protein thought to be a part of the translocon at the chloroplast inner envelope. Involved in protein import into the chloroplast and chloroplast biogenesis. C-terminal half of Tic110 functions as scaffolds for protein-protein interactions.

	N/A	1	Kaladp0032s0450.1	AT4G38770.1	ARABIDOPSIS THALIANA PROLINE-RICH PROTEIN 4 (ATPRP4)	Encodes one of four proline-rich proteins in Arabidopsis which are predicted to localize to the cell wall. Transcripts are most abundant in aerial organs of the plant.
	N/A	1	Kaladp0009s0009.1	AT1G33360.1	Encodes ClpX3, a subunit of the Clp protease complex.	Encodes ClpX3, a subunit of the Clp protease complex.
	N/A	1	Kaladp0067s0207.1	AT1G79550.1	PHOSPHOGLYCERATE KINASE 3 (PGK3)	Encodes cytosolic phosphoglycerate kinase (PGK3). Expression studies in PGK mutants showed that PGK1 and PGK3 were down-regulated in pgk3.2 and pgk1.1, respectively. These results indicate that the down-regulation of photosynthetic activity could be a plant strategy when glycolysis is impaired to achieve metabolic adjustment and optimize growth.
	N/A	1	Kaladp0045s0220.1	AT5G58900.1	DIVARICATA1 (DIV1)	R-R-type MYB protein
	N/A	1	Kaladp0099s0108.1	AT5G48030.1	GAMETOPHYTIC FACTOR 2 (GFA2)	encodes a mitochondrially targeted DNAJ protein involved in female gametophyte development.

	N/A	1	Kaladp0863s0010.1	AT3G16370.1	GDSL-motif esterase/acyltransferase/lipase	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactions with lipid and non-lipid substrates. The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0008s0319.1	AT5G25610.1	RESPONSIVE TO DESICCATION 22 (ATRD22)	responsive to dehydration 22 (RD22) mediated by ABA
	N/A	1	Kaladp0001s0263.1	AT5G65630.1	GLOBAL TRANSCRIPTION FACTOR GROUP E7 (GTE7)	This gene is predicted to encode a bromodomain-containing protein. Plant lines expressing RNAi constructs targeted against GTE7 show some resistance to agrobacterium-mediated root transformation.
	N/A	1	Kaladp0035s0087.1	AT5G41010.1	NRPB12	Non-catalytic subunit common to nuclear DNA-dependent RNA polymerases II, IV and V; homologous to budding yeast RPB12.
	N/A	1	Kaladp0024s0238.1	AT1G05190.1	EMBRYO DEFECTIVE 2394 (EMB2394)	Encodes the plastid 50S ribosomal protein L6.

	N/A	1	Kaladp0292s0007.1	AT4G22670.1	HSP70-INTERACTING PROTEIN 1 (ATHIP1)	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins (Prasad 2010, Pubmed ID: 20856808) with potential to interact with Hsp90/Hsp70 as co-chaperones. The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0040s0495.1	AT3G15360.1	ARABIDOPSIS THIOREDOXIN M-TYPE 4 (ATHM4)	encodes a prokaryotic thioredoxin The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0087s0084.1	AT3G52500.1	Eukaryotic aspartyl protease family protein	Eukaryotic aspartyl protease family protein
	N/A	1	Kaladp0022s0218.1	AT2G32720.1	ARABIDOPSIS CYTOCHROME B5 ISOFORM B (ATCB5-B)	Participates with ELO2 in VLCFA synthesis.
	N/A	1	Kaladp0959s0007.1	AT1G01620.1	a member of the plasma membrane intrinsic protein subfamily PIP1	a member of the plasma membrane intrinsic protein subfamily PIP1. localizes to the plasma membrane and exhibits water transport activity in Xenopus oocyte. expressed ubiquitously and protein level decreases slightly during leaf development. Involved redundantly with PIP1;1/2/4/5 in hydraulics and carbon fixation,

						regulates the expression of related genes that affect plant growth and development.
G-mesophyll	B	3	Kaladp0036s0319.1	AT4G32010.1	HSI2-LIKE 1	Transcriptional repressor involved in the recruitment of PRC2 for genome-wide polycomb silencing.
	B	4	Kaladp0053s0637.1	AT2G20180.2	PHY-INTERACTING FACTOR 1	Encodes a novel Myc-related bHLH transcription factor that has transcriptional activation activity in the dark. It is a key negative regulator of phytochrome-mediated seed germination and acts by inhibiting chlorophyll biosynthesis, light-mediated suppression of hypocotyl elongation and far-red light-mediated suppression of seed germination, and promoting negative gravitropism in hypocotyls. Light reduces this activity in a phy-dependent manner. The protein preferentially interacts with the Pfr forms of

						Phytochrome A (PhyA) and Phytochrome B (PhyB), is physically associated with APRR1/TOC1 and is degraded in red (R) and far-red (FR) light through the ubiquitin (ub)-26S proteasome pathway to optimize photomorphogenic development in Arabidopsis. It also negatively regulates GA3 oxidase expression.
	B	6	Kaladp0010s0058.1	AT5G44210.1	ATERF-9	encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-9). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole
	D	2	Kaladp0036s0113.1	AT4G11420.1	EUKARYOTIC TRANSLATION INITIATION FACTOR 3A (TIF3A1)	Encodes a subunit of eukaryotic initiation factor 3 (eIF3), a multisubunit complex that is required for binding of mRNA to 40 S ribosomal

						subunits, stabilization of ternary complex binding to 40 S subunits, and dissociation of 40 and 60 S subunits.
	D	1	Kaladp0011s0523.1	AT3G54050.1	HIGH CYCLIC ELECTRON FLOW 1	Encodes a chloroplastic fructose 1,6-bisphosphate phosphatase. also known as HCEF1 (High Cyclic Electron Flow 1). hcef1 mutants have constitutively elevated electron flow (CEFI) and plants with antisense suppression of this enzyme have higher levels of net leaf photosynthesis and increased sucrose biosynthesis. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0040s0625.1	AT1G06070.1	BASIC LEUCINE-ZIPPER 69 (BZIP69)	Basic-leucine zipper (bZIP) transcription factor family protein
	D	1	Kaladp0032s0085.1	AT2G36400.1	GROWTH-REGULATING FACTOR 3	Growth regulating factor encoding transcription activator. One of the nine members of a GRF gene family, containing nuclear targeting domain. Mutants result in smaller leaves indicating

						the role of the gene in leaf development. Expressed in root, shoot and flower.
	D	1	Kaladp0042s0395.1	ATCG01180.1	RNA23S	chloroplast-encoded 23S ribosomal RNA
	D	1	Kaladp0010s0058.1	AT5G44210.1	ERF DOMAIN PROTEIN- 9 (ERF9)	encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-9). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole.
	D	1	Kaladp0808s0040.1	AT4G16830.1	ATRGGA	Encodes a perinuclear and cytoplasmically localized mRNA binding protein. AtRGGA is likely involved in stress responsivness. It is induced by salt and osmotic stress and loss of function mutations are more sensitive to stress. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0037s0187.1	AT2G40410.2	ATCAN2	Encodes a Ca(2+)-dependent nuclease that can degrade both DNA and RNA.

	D	1	Kaladp0075s0014.1	AT2G33590.1	CCR(CINNAMOYL COA:NADP OXIDOREDUCTASE)-LIKE 1	Encodes a protein with homology to members of the dihydroflavonol-4-reductase (DFR) superfamily. The expression pattern of AtCRL1 indicates that CRL1 has a role in embryogenesis and seed germination. AtCRL1 is induced by ABA, drought and heat, and is highly expressed in seeds. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0011s1278.1	AT4G37740.1	GROWTH-REGULATING FACTOR 2	Growth regulating factor encoding transcription activator. One of the nine members of a GRF gene family, containing nuclear targeting domain. Mutants result in smaller leaves indicating the role of the gene in leaf development. Expressed in root, shoot and flower
	D	2	Kaladp0095s0662.1	AT4G35800.1	RNA POLYMERASE II LARGE SUBUNIT	Encodes the unique largest subunit of nuclear DNA-dependent RNA polymerase II; the ortholog of budding yeast RPB1 and a homolog of the E. coli RNA

						polymerase beta prime subunit.
	D	1	Kaladp0094s0022.1	AT5G61030.1	GLYCINE-RICH RNA-BINDING PROTEIN 3	Encodes a glycine-rich RNA binding protein that is involved in C-> U RNA editing in mitochondria. Gene expression is induced by cold. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0042s0027.1	AT1G75170.1	Sec14p-like phosphatidylinositol transfer family protein	Sec14p-like phosphatidylinositol transfer family protein
	D	1	Kaladp0045s0478.1	AT1G05210.1	Transmembrane protein 97	Transmembrane protein 97
	D	1	Kaladp0048s0336.1	AT1G75380.1	BIFUNCTIONAL NUCLEASE IN BASAL DEFENSE RESPONSE 1	Encodes a nuclease involved in ABA-mediated callose deposition. It has been shown to interact with JAZ proteins, binds to a jasmonic acid-responsive element (JARE) and repress AtJMT expression.
	D	1	Kaladp0964s0028.1	AT1G79090.1	ARABIDOPSIS HOMOLOG OF YEAST PAT1	Encodes a component of the mRNA decapping machinery that is found in complexes with and is phosphorylated by MPK4. The protein can be found in P-bodies after treatment with PAMPs. Pat1 mutants exhibit dwarfism and de-

						repressed immunity dependent on the immune receptor SUMM2. The pat1 constitutive defense phenotype is suppressed by summ2 such that pat1 summ2 mutants display a wild-type phenotype in response to biotrophic and necrotrophic pathogens.
	D	1	Kaladp0039s0246.1	AT1G28330.5	DORMANCY-ASSOCIATED PROTEIN 1	dormancy-associated protein (DRM1)
	D	1	Kaladp0113s0002.1	AT1G25350.2	OVULE ABORTION 9	glutamine-tRNA ligase, putative / glutaminyl-tRNA synthetase, putative / GlnRS
CACT-mesophyll	A	5	Kaladp0023s0075.1	AT2G22800.1	HAT9	Encodes homeobox protein HAT9
	A	4	Kaladp0844s0013.1	AT4G16780.1	ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 2 (ATHB-2)	Encodes a homeodomain-leucine zipper protein that is rapidly and strongly induced by changes in the ratio of red to far-red light. It is also involved in cell expansion and cell proliferation and in the response to auxin. The mRNA

						is cell-to-cell mobile.
	A	10	Kaladp0079s0030.1	AT2G44910.1	ARABIDOPSIS THALIANA HOMEOBOX-LEUCINE ZIPPER PROTEIN 4 (ATHB-4)	Encodes a homeodomain protein whose expression displays a dependence on phyB for both red and far-red light response. Also involved in the shade avoidance syndrome.
	B	4	Kaladp0058s0639.1	AT3G53600.1	ZAT18	Nuclear C2H2 zinc finger protein.Expression is induced by cold, osmotic, salt, and drought stress. Over expression confers some drought tolerance whereas mutants display some drought sensitivity.
	C	2	Kaladp0071s0123.1	AT5G15150.1	ATHB-3	homeobox-containing gene with an unusual feature: a leucine zipper motif adjacent to the carboxyl-terminal of the homeodomain structure. This gene is expressed primarily in the cortex of the root and the stem.
	C	2	Kaladp0102s0021.1	AT5G28490.1	LIGHT-DEPENDENT SHORT HYPOCOTYLS 1 (LSH1)	Encodes a nuclear protein that mediates light regulation of seedling development in a

						phytochrome-dependent manner.
	C	2	Kaladp0323s0002.1	AT5G65310.1	HOMEOBOX PROTEIN 5 (ATHB-5)	Encodes a class I HDZip (homeodomain-leucine zipper) protein that is a positive regulator of ABA-responsiveness, mediating the inhibitory effect of ABA on growth during seedling establishment.
	D	2	Kaladp0808s0045.1	AT2G44910.1	ARABIDOPSIS THALIANA HOMEOBOX-LEUCINE ZIPPER PROTEIN 4 (ATHB-4)	Encodes a homeodomain protein whose expression displays a dependence on phyB for both red and far-red light response. Also involved in the shade avoidance syndrome.
	D	1	Kaladp0067s0146.1	AT4G10770.1	ARABIDOPSIS THALIANA OLIGOPEPTIDE TRANSPORTER 7 (ATOPT7)	oligopeptide transporter
	D	1	Kaladp0070s0188.1	AT3G17780.1	B-cell receptor-associated-like protein	B-cell receptor-associated-like protein
	D	1	Kaladp0024s0501.1	AT5G65310.1	HOMEOBOX PROTEIN 5 (ATHB-5)	Encodes a class I HDZip (homeodomain-leucine zipper) protein that is a positive regulator of ABA-responsiveness, mediating the inhibitory effect of ABA on growth

						during seedling establishment.
	D	1	Kaladp0016s0352.1	AT1G15820.1	LIGHT HARVESTING COMPLEX PHOTOSYSTEM II SUBUNIT 6	Lhcb6 protein (Lhcb6), light harvesting complex of photosystem II.
	D	1	Kaladp0058s0639.1	AT3G53600.1	ZAT18	Nuclear C2H2 zinc finger protein.Expression is induced by cold, osmotic, salt, and drought stress. Over expression confers some drought tolerance whereas mutants display some drought sensitivity.
	D	1	Kaladp0022s0180.1	AT3G04780.1	Encodes a protein with little sequence identity with any other protein of known structure or function. Part of this protein shows a 42% sequence identity with the C-terminal domain of the 32-kD human thioredoxin-like protein.	Encodes a protein with little sequence identity with any other protein of known structure or function. Part of this protein shows a 42% sequence identity with the C-terminal domain of the 32-kD human thioredoxin-like protein.
	D	1	Kaladp0051s0016.1	AT4G22360.1	SWIB complex BAF60b domain-containing protein	SWIB complex BAF60b domain-containing protein
	D	1	Kaladp0076s0328.1	AT3G19250.1	transmembrane protein, putative (DUF677)	transmembrane protein, putative (DUF677)
	D	1	Kaladp0024s0051.1	AT4G17740.1	Peptidase S41 family protein	Peptidase S41 family protein
	D	1	Kaladp0033s0026.1	AT2G36830.1	GAMMA TONOPLAST INTRINSIC PROTEIN 1	Encodes a tonoplast intrinsic protein, which functions as water

						channel. It has also been shown to be able to facilitate the transport of urea and hydrogen peroxide. Highly expressed in vascular tissues of the root, stem, cauline leaves and flowers but not in the apical meristems. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0047s0062.1	AT1G27730.1	SALT TOLERANCE ZINC FINGER, STZ, ZAT10	Related to Cys2/His2-type zinc-finger proteins found in higher plants. Compensated for a subset of calcineurin deficiency in yeast. Salt tolerance produced by ZAT10 appeared to be partially dependent on ENA1/PMR2, a P-type ATPase required for Li+ and Na+ efflux in yeast. The protein is localized to the nucleus, acts as a transcriptional repressor and is responsive to chitin oligomers. Also involved in response to photooxidative stress.

	D	1	Kaladp0029s0125.1	AT4G31890.1	ARM repeat superfamily protein	ARM repeat superfamily protein
	D	4	Kaladp0042s0108.1	AT5G04340.1	ATZAT6	Encodes a C2H2 zinc finger transcription factor that coordinately activates phytochelatin-synthesis related gene expression and directly targets GSH1 by binding to its promoter to positively regulate Cd accumulation and tolerance.
	D	1	Kaladp0045s0160.1	AT2G28200.1	C2H2-type zinc finger family protein	C2H2-type zinc finger family protein
	D	3	Kaladp0040s0265.1	AT1G27730.1	SALT TOLERANCE ZINC FINGER, STZ, ZAT10	Related to Cys2/His2-type zinc-finger proteins found in higher plants. Compensated for a subset of calcineurin deficiency in yeast. Salt tolerance produced by ZAT10 appeared to be partially dependent on ENA1/PMR2, a P-type ATPase required for Li+ and Na+ efflux in yeast. The protein is localized to the nucleus, acts as a transcriptional repressor and is responsive to chitin oligomers. Also

						involved in response to photooxidative stress.
	D	1	Kaladp0062s0180.1	AT2G22540.1	AGAMOUS-LIKE 22	Encodes a nuclear protein that acts as a floral repressor and that functions within the thermosensory pathway. SVP represses FT expression via direct binding to the vCArG III motif in the FT promoter.
	D	1	Kaladp0042s0024.1	AT1G75170.1	Sec14p-like phosphatidylinositol transfer family protein	Sec14p-like phosphatidylinositol transfer family protein
	D	1	Kaladp0103s0006.1	AT5G18940.1	Mo25 family protein	Mo25 family protein
	N/A	1	Kaladp0039s0498.1	AT5G38410.3	RUBISCO SMALL SUBUNIT 3B	Encodes a member of the Rubisco small subunit (RBCS) multigene family: RBCS1A (At1g67090), RBCS1B (At5g38430), RBCS2B (At5g38420), and RBCS3B (At5g38410). Functions to yield sufficient Rubisco content for leaf photosynthetic capacity.
	N/A	1	Kaladp0035s0044.1	AT3G55610.1	DELTA 1-PYRROLINE-5-CARBOXYLATE SYNTHASE 2	encodes delta 1-pyrroline-5-carboxylate synthetase B. Gene

						expression is induced by dehydration, high salt and ABA. Knock-out mutations in P5CS2 are embryo-lethal. P5CS2 appears to be present in different cells and/or different subcellular locations from P5CS1 in a tissue-dependent manner. Mutants are defective in pollen development.
	N/A	1	Kaladp0040s0265.1	AT1G27730.1	SALT TOLERANCE ZINC FINGER, STZ, ZAT10	Related to Cys2/His2-type zinc-finger proteins found in higher plants. Compensated for a subset of calcineurin deficiency in yeast. Salt tolerance produced by ZAT10 appeared to be partially dependent on ENA1/PMR2, a P-type ATPase required for Li+ and Na+ efflux in yeast. The protein is localized to the nucleus, acts as a transcriptional repressor and is responsive to chitin oligomers. Also involved in

						response to photooxidative stress.
enhancer-like elements	D	2	Kaladp0036s0297.1	AT1G06950.1	ARABIDOPSIS THALIANA TRANSLOCON AT THE INNER ENVELOPE MEMBRANE OF CHLOROPLASTS 110 (ATTIC110)	Encodes a protein thought to be a part of the translocon at the chloroplast inner envelope. Involved in protein import into the chloroplast and chloroplast biogenesis. C-terminal half of Tic110 functions as scaffolds for protein-protein interactions.
	D	1	Kaladp0033s0028.1	AT5G59780.3	ATMYB59	Encodes a putative transcription factor (MYB59). In roots it is involved in K+/NO3- transport and expression of the NPF7.3 transporter.
	D	1	Kaladp0102s0112.1	AT2G25760.2	ARABIDOPSIS EL1-LIKE 1 (AEL1)	Casein kinase involved in phosphorylation and ubiquination of RYR/PYLs, resulting in negative regulation of ABA response. Plays a role in repressing the transition from vegetative to reproductive phase.
	D	1	Kaladp0081s0239.1	AT2G27980.1	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger

					domain-containing protein	domain-containing protein
	D	1	Kaladp0063s0015.1	AT2G35110.1	GNARLED	Component of the WAVE protein complex which act as activators of ARP2/3 complex involved in actin nucleation. Required for trichome morphogenesis. Mutant displays distorted trichomes, a phenotype that can be phenocopied by treatment of WT plants with actin-interacting drugs. Its ER localization is independent of upstream SPK1 activation signaling and the physical association with the WAVE/SCAR Regulatory complex binding partner SRA1. ER-localized NAP1 has little colocalization with actin and represent the inactive pool of NAP1 in these cell types.
	D	1	Kaladp0060s0147.1	AT1G01590.1	FERRIC REDUCTION OXIDASE 1 (FRO1)	Encodes a ferric-chelate reductase that is expressed at extremely low levels in Fe deficiency-induced seedlings.

	D	1	Kaladp1303s0001.1	AT1G01590.1	FERRIC REDUCTION OXIDASE 1 (FRO1)	Encodes a ferric-chelate reductase that is expressed at extremely low levels in Fe deficiency-induced seedlings.
	N/A	1	Kaladp0102s0112.1	AT2G25760.2	ARABIDOPSIS EL1-LIKE 1 (AEL1)	Casein kinase involved in phosphorylation and ubiquination of RYR/PYLs, resulting in negative regulation of ABA response. Plays a role in repressing the transition from vegetative to reproductive phase.
***KfPEPC1* promoter (-1 to -200)**	D	1	Kaladp0021s0075.1	AT2G08665.1	antisense_long_noncoding_rna	Natural antisense transcript overlaps with AT2G34420
	D	1	Kaladp0016s0352.1	AT1G15820.1	LIGHT HARVESTING COMPLEX PHOTOSYSTEM II SUBUNIT 6	Lhcb6 protein (Lhcb6), light harvesting complex of photosystem II.
	D	1	Kaladp0024s0493.1	AT4G40040.1	HISTONE 3.3	Histone superfamily protein
	D	1	Kaladp0076s0229.1	AT4G15530.5	PYRUVATE ORTHOPHOSPHATE DIKINASE (PPDK)	Encodes a dual-targeted protein believed to act as a pyruvate, orthophosphate dikinase. These enzymes are normally

						associated with C4 photosynthesis which does not occur in Arabidopsis. However, PPDK may play a role in remobilizing nitrogen during leaf senescence in Arabidopsis. The product of the long transcript (.1 gene model) was shown to be targeted to the chloroplast, whereas the shorter transcript (no targeting sequence) accumulates in the cytosol. The two proteins were also found to be expressed in slightly different tissues.
	D	1	Kaladp0082s0004.1	AT2G33260.1	Tryptophan/tyrosine permease	Tryptophan/tyrosine permease
	D	1	Kaladp0061s0105.1	AT2G30570.1	PHOTOSYSTEM II REACTION CENTER W	Encodes PsbW, a protein similar to photosystem II reaction center subunit W. Loss of PsbW destabilizes the supramolecular organization of PSII.
	D	1	Kaladp1241s0004.1	AT5G19860.1	transmembrane protein, putative (Protein of unknown function, DUF538)	transmembrane protein, putative (Protein of unknown function, DUF538)

	D	1	Kaladp0068s0349.1	AT4G18530.1	SGC, STICKY GENERATIVE CELL	lysine ketoglutarate reductase trans-splicing-like protein, putative (DUF707)
	D	1	Kaladp0076s0366.1	AT5G48580.1	FK506- AND RAPAMYCIN-BINDING PROTEIN 15 KD-2 (FKBP15-2)	Endoplasmic reticulum (ER) localized immunophilin protein which possesses PPIase activity. Positively regulates plant immunity in response to Phytophthora infection. Host target of PcAvr3a12 during early P. capsici infection. Involved in ER stress sensing and is required for ER stress-mediated plant immunity.
	D	1	Kaladp0024s0977.1	AT4G39040.1	CRM FAMILY MEMBER SUBFAMILY 4	RNA-binding CRS1 / YhbY (CRM) domain protein
	D	1	Kaladp0052s0030.1	AT1G69750.1	CYTOCHROME C OXIDASE 19-2	CYTOCHROME C OXIDASE 19-2
	D	4	Kaladp0101s0026.1	AT2G45860.1	hypothetical protein	hypothetical protein
	D	1	Kaladp0068s0288.1	AT3G54720.1	ALTERED MERISTEM PROGRAM 1	Encodes glutamate carboxypeptidase. Various alleles show-increased cotyledon number and rate of leaf initiation, show transformation of

						leaves to cotyledons, altered flowering time and photomorphogenes is and an increased level of cytokinin biosynthesis. Involved in ethylene enhanced hypocotyl elongation in the light. Strong genetic interaction between TGH and AMP1.
	D	1	Kaladp0034s0228.1	AT4G14740.1	FORKED-LIKE3	FORKED-LIKE family member, part of Group 1 (FKD1, FL1-FL3; Group 2 consists of FL4 and FL8 and Group 3 consists of FL5- FL7). May coordinate leaf size with vein density, where Group 1 members and Group 3 members have opposing functions.
	D	1	Kaladp0016s0330.1	AT4G21860.1	METHIONINE SULFOXIDE REDUCTASE B 2	METHIONINE SULFOXIDE REDUCTASE B 2
	D	1	Kaladp0008s0330.1	AT3G20770.1	ETHYLENE-INSENSITIVE3	Encodes EIN3 (ethylene-insensitive3), a nuclear transcription factor that initiates downstream transcriptional cascades for ethylene responses. EIN3 interacts with MYC2, MYC3 and

						MYC4 to inhibit jasmonate-induced expression of wound-responsive genes and herbivory-inducible genes, and plant defense against generalist herbivores.
	D	1	Kaladp0068s0050.1	AT3G25240.1	sulfate/thiosulfate import ATP-binding protein, putative (DUF506)	sulfate/thiosulfate import ATP-binding protein, putative (DUF506)
	D	1	Kaladp0037s0018.1	AT4G30960.1	CBL-INTERACTING PROTEIN KINASE 6	Encodes CBL-interacting protein kinase 6 (CIPK6). Required for development and salt tolerance. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0840s0019.1	AT4G18390.1	TCP2	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2
	D	1	Kaladp0036s0095.1	AT1G56220.3	Dormancy/auxin associated family protein	Dormancy/auxin associated family protein
	D	1	Kaladp0011s0229.1	AT5G57710.1	SUPPRESSOR OF MAX2 1	SMAX1 (SUPPRESSOR OF MAX2 1) is a member of an eight-gene family in Arabidopsis that has weak similarity to AtHSP101, a ClpB chaperonin required for thermotolerance. SMAX1 is an important component of KAR/SL signaling

						during seed germination and seedling growth, but is not necessary for all MAX2-dependent responses. The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0863s0031.1	AT1G17720.1	ATB BETA	type 2A protein serine/threonine phosphatase 55 kDa B
	N/A	1	Kaladp0024s0275.1	AT5G22060.1	ARABIDOPSIS THALIANA DNAJ HOMOLOGUE 2	Co-chaperonin similar to E. coli DnaJ
	N/A	1	Kaladp0056s0113.1	AT4G21105.1	cytochrome-c oxidase/ electron carrier	cytochrome-c oxidase/ electron carrier
	N/A	1	Kaladp0037s0356.1	AT1G80070.1	ABNORMAL SUSPENSOR 2	Encodes a factor that influences pre-mRNA splicing and is required for embryonic development. Mutations result in an abnormal suspensor and embryo lethality. The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0011s1141.1	AT2G30790.1	PHOTOSYSTEM II SUBUNIT P-2	Encodes a 23 kD extrinsic protein that is part of photosystem II and participates in the regulation of oxygen evolution.
	N/A	1	Kaladp0099s0114.1	AT5G02500.1	ARABIDOPSIS THALIANA HEAT SHOCK COGNATE PROTEIN 70-1	Encodes a member of heat shock protein 70 family. Hsc70-1 negatively regulates the expression of Hsp101 through

						HsfA1d, HsfA1e and HsfA2. During non-HS condition, Hsc70-1 attenuates the activity of HsfAs and finally affects the expression of HsfA2 and Hsp101 genes. hsc70-1 mutant showed thermotolerance phenotype due to higher expression of Hsp101 and other HS inducible genes.
	N/A	2	Kaladp0064s0010.1	ATCG01210.1	RIBOSOMAL RNA 16S	chloroplast-encoded 16S ribosomal RNA
***KfPEPC1* promoter (-171 to -370)**	B	1	Kaladp0095s0716.1	AT5G07580.1	ERF106	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5.
	B	9	Kaladp0403s0001.1	AT5G07580.1	ERF106	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this

						subfamily including ATERF-1, ATERF-2, AND ATERF-5.
	B	2	Kaladp0042s0395.1	ATCG01180.1	RIBOSOMAL RNA23S	chloroplast-encoded 23S ribosomal RNA
	B	3	Kaladp0008s0505.1	AT5G08720.1	PIN2 PROMOTER BINDING PROTEIN 1 (PPP1)	Encodes PIN2 PROMOTER BINDING PROTEIN 1 (PPP1), an evolutionary conserved plant-specific DNA binding protein that acts on transcription of PIN genes. Also named as HCF145. Mutations in HCF145 have reduced level of the tricistronic psaA-psaB-rps (small-subunit ribosomal protein)14 mRNA which encodes for the major subunits of the photosystem I (PSI). HCF145 binds to the 5'UTR of PSAA via a novel TMR domain. It functions to stabilize the PSAA transcript.
	C	2	Kaladp0011s0663.1	AT5G59690.1	Histone superfamily protein	Histone superfamily protein
	C	2	Kaladp0082s0119.1	AT2G44840.1	ETHYLENE-RESPONSIVE ELEMENT	encodes a member of the ERF (ethylene response

					BINDING FACTOR 13 (ERF13)	factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0037s0023.1	AT5G57940.1	CYCLIC NUCLEOTIDE GATED CHANNEL 5	Encodes a cyclic GMP activated Ca2+-permeable cation channel in the plasma membrane of guard cells. Required for constitutive growth of root hairs as Ca2+-permeable channels.
	D	1	Kaladp0844s0013.1	AT4G16780.1	ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 2 (ATHB-2)	Encodes a homeodomain-leucine zipper protein that is rapidly and strongly induced by changes in the ratio of red to far-red light. It is also involved in cell expansion and cell proliferation and in the response to auxin. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0010s0058.1	AT5G44210.1	ATERF-9	encodes a member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2

						transcription factor family (ATERF-9). The protein contains one AP2 domain. There are 15 members in this subfamily including ATERF-3, ATERF-4, ATERF-7, and leafy petiole.
	D	1	Kaladp0053s0123.1	AT5G13710.1	STEROL METHYLTRANSFERASE 1	SMT1 controls the level of cholesterol in plants
	D	1	Kaladp0058s0342.1	AT3G62290.1	ADP-RIBOSYLATION FACTOR A1E	A member of ARF GTPase family. A thaliana has 21 members of this family, known to be essential for vesicle coating and uncoating and functions in GTP-binding. The gene is shown to play a role in cell division, cell expansion and cellulose production using antisense construct. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0053s0626.1	AT4G19390.1	Uncharacterized protein family (UPF0114)	Uncharacterized protein family (UPF0114)
	D	1	Kaladp0081s0189.1	AT1G14230.1	GDA1/CD39 nucleoside phosphatase family protein	GDA1/CD39 nucleoside phosphatase family protein
	D	1	Kaladp1116s0006.1	AT1G52190.1	NITRATE TRANSPORTER 1.11	Encodes a low affinity nitrate transporter that is expressed in the plasma membrane

						and found in the phloem of the major veins of leaves. It is responsible for nitrate redistribution to young leaves.
	D	1	Kaladp0015s0100.1	AT2G47730.1	ARABIDOPSIS THALIANA GLUTATHIONE S-TRANSFERASE PHI 8	Encodes glutathione transferase belonging to the phi class of GSTs. Naming convention according to Wagner et al. (2002).
	D	1	Kaladp0081s0343.1	AT5G60410.2	SAP AND MIZ1 DOMAIN-CONTAINING LIGASE1	Encodes a plant small ubiquitin-like modifier (SUMO) E3 ligase that is a focal controller of Pi starvation-dependent responses. Also required for SA and PAD4-mediated R gene signalling, which in turn confers innate immunity in Arabidopsis. Also involved in the regulation of plant growth, drought responses and freezing tolerance. This latter effect is most likely due to SIZ1 dependent ABI5 sumoylation. Regulates leaf cell division and

						expansion through salicylic acid accumulation. signaling
	D	1	Kaladp0012s0059.1	AT3G63140.1	CHLOROPLAST STEM-LOOP BINDING PROTEIN OF 41 KDA	Encodes a protein with ribonuclease activity that is involved in plastid rRNA maturation
	D	1	Kaladp0062s0207.1	AT1G60990.1	Encodes a chloroplast-localized COG0354 protein that requires folate for its function in Fe/S cluster biogenesis.	Encodes a chloroplast-localized COG0354 protein that requires folate for its function in Fe/S cluster biogenesis.
	D	1	Kaladp0060s0381.1	AT4G21960.1	PRXR1	Encodes AT4g21960 (AT4g21960/T8O5_170). The mRNA is cell-to-cell mobile.
	D	1	Kaladp0574s0016.1	AT5G63160.1	BTB AND TAZ DOMAIN PROTEIN 1	BTB and TAZ domain protein. Short-lived nuclear-cytoplasmic protein targeted for degradation by the 26S proteosome pathway. Acts redundantly with BT2 and BT3 during female gametophyte development.
	D	1	Kaladp0047s0075.1	AT4G33760.1	OKI1	Aminoacyl tRNA synthetase functions in SAM maintenance.
	D	1	Kaladp0081s0218.1	AT3G56160.1	Sodium Bile acid symporter family	Sodium Bile acid symporter family

	D	1	Kaladp0050s0155.1	AT4G19650.1	MTERF26	Mitochondrial transcription termination factor family protein
	D	1	Kaladp0067s0283.1	AT2G47730.1	ARABIDOPSIS THALIANA GLUTATHIONE S-TRANSFERASE PHI 8	Encodes glutathione transferase belonging to the phi class of GSTs. Naming convention according to Wagner et al. (2002).
	D	1	Kaladp0100s0017.1	AT3G61040.1	CYP76C7	encodes a protein with cytochrome P450 domain
	D	1	Kaladp0101s0026.1	AT2G45860.1	hypothetical protein	hypothetical protein
	N/A	1	Kaladp0096s0124.1	AT4G21120.1	CATIONIC AMINO ACID TRANSPORTER 1	Encodes a member of the cationic amino acid transporter (CAT) subfamily of amino acid polyamine choline transporters. Mediates efficient uptake of Lys, Arg and Glu in a yeast system. The mRNA is cell-to-cell mobile.
	N/A	2	Kaladp0053s0637.1	AT2G20180.2	PHY-INTERACTING FACTOR 1 (PIF1)	Encodes a novel Myc-related bHLH transcription factor that has transcriptional activation activity in the dark. It is a key negative regulator of phytochrome-mediated seed

						germination and acts by inhibiting chlorophyll biosynthesis, light-mediated suppression of hypocotyl elongation and far-red light-mediated suppression of seed germination, and promoting negative gravitropism in hypocotyls. Light reduces this activity in a phy-dependent manner. The protein preferentially interacts with the Pfr forms of Phytochrome A (PhyA) and Phytochrome B (PhyB), is physically associated with APRR1/TOC1 and is degraded in red (R) and far-red (FR) light through the ubiquitin (ub)-26S proteasome pathway to optimize photomorphogenic development in Arabidopsis. It also negatively regulates GA3 oxidase expression.
	N/A	1	Kaladp0001s0016.1	AT4G35090.1	CATALASE 2	Encodes a peroxisomal catalase, highly expressed in bolts

						and leaves. mRNA expression patterns show circadian regulation with mRNA levels being high in the subjective early morning. Loss of function mutations have increased H2O2 levels and increased H2O2 sensitivity. Mutants accumulate more toxic ions yet show decreased sensitivity to Li+. This decreased sensitivity is most likely due to an insensitivity to ethylene. Note that in Queval et al. (2007) Plant Journal, 52(4):640, SALK_057998 is named as cat2-1, SALK_076998 is named as cat2-2; in Bueso et al. (2007) Plant Journal, 52(6):1052, SALK_076998 is named as cat2-1. TAIR has adopted the nomenclature consistent with that in Bueso et al. (2007) after consultation with the authors: SALK_076998 (cat2-1),

						SALK_057998 (cat2-2).
	N/A	2	Kaladp0492s0002.1	AT3G54890.1	PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 1 (LHCA1)	Encodes a component of the light harvesting complex associated with photosystem I.
	N/A	1	Kaladp0039s0498.1	AT5G38410.3	RUBISCO SMALL SUBUNIT 3B	Encodes a member of the Rubisco small subunit (RBCS) multigene family: RBCS1A (At1g67090), RBCS1B (At5g38430), RBCS2B (At5g38420), and RBCS3B (At5g38410). Functions to yield sufficient Rubisco content for leaf photosynthetic capacity.
	N/A	1	Kaladp0674s0166.1	AT3G47470.1	LIGHT-HARVESTING CHLOROPHYLL-PROTEIN COMPLEX I SUBUNIT A4 (LHCA4)	Encodes a chlorophyll a/b-binding protein that is more similar to the PSI Cab proteins than the PSII cab proteins. The predicted protein is about 20 amino acids shorter than most known Cab proteins.
	N/A	1	Kaladp0019s0045.1	AT1G01520.1	ALTERED SEED GERMINATION 4	RVE3 is one of eleven homologous MYB-like transcription factors in Arabidopsis and a member of the RVE8 clade. Plays

						a minor role in clock regulation.
	N/A	1	Kaladp0042s0395.1	ATCG01180.1	RIBOSOMAL RNA23S	chloroplast-encoded 23S ribosomal RNA
	N/A	1	Kaladp0067s0207.1	AT1G79550.1	PHOSPHOGLYCERATE KINASE 3	Encodes cytosolic phosphoglycerate kinase (PGK3). Expression studies in PGK mutants showed that PGK1 and PGK3 were down-regulated in pgk3.2 and pgk1.1, respectively. These results indicate that the down-regulation of photosynthetic activity could be a plant strategy when glycolysis is impaired to achieve metabolic adjustment and optimize growth.
	N/A	2	Kaladp0059s0024.1	AT2G34430.1	LIGHT-HARVESTING CHLOROPHYLL-PROTEIN COMPLEX II SUBUNIT B1	Photosystem II type I chlorophyll a/b-binding protein The mRNA is cell-to-cell mobile
	N/A	1	Kaladp0059s0105.1	AT2G36530.1	ENOLASE 2	Involved in light-dependent cold tolerance and encodes an enolase. Protein is tyrosine-phosphorylated and its phosphorylation state is modulated in response to ABA in Arabidopsis

						thaliana seeds. Affects seed size and weight by adjusting cytokinin content and forming ENO2-bZIP75 complex.
	N/A	1	Kaladp0131s0039.1	AT3G05100.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
	N/A	1	Kaladp0011s0049.1	AT2G02450.2	ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 34 (ANAC034)	NAC domain containing protein 35
	N/A	1	Kaladp0095s0427.1	AT1G50010.1	TUBULIN ALPHA-2 CHAIN	Encodes alpha-2,4 tubulin. TUA2 and TUA4 encode identical proteins. The mRNA is cell-to-cell mobile.
	N/A	1	Kaladp0076s0229.1	AT4G15530.5	PYRUVATE ORTHOPHOSPHATE DIKINASE (PPDK)	Encodes a dual-targeted protein believed to act as a pyruvate, orthophosphate dikinase. These enzymes are normally associated with C4 photosynthesis which does not occur in Arabidopsis. However, PPDK may play a role in remobilizing nitrogen during leaf senescence in Arabidopsis. The product of the long transcript (.1 gene

						model) was shown to be targeted to the chloroplast, whereas the shorter transcript (no targeting sequence) accumulates in the cytosol. The two proteins were also found to be expressed in slightly different tissues.
	N/A	1	Kaladp0076s0373.1	AT3G07030.1	Alba DNA/RNA-binding protein (ALBA6)	Alba DNA/RNA-binding protein
	N/A	1	Kaladp0056s0019.1	AT5G07540.1	GLYCINE-RICH PROTEIN 16 (GRP16)	encodes a glycine-rich protein that is expressed only in flowers during a specific developmental stage (flower stages 11 and 12).
***KfPEPC1* promoter (-341 to -540)**	A	5	Kaladp0011s1344.1	AT3G15030.1	TCP FAMILY TRANSCRIPTION FACTOR 4 (TCP4)	Arabidopsis thaliana TCP family transcription factor. Regulated by miR319. Involved in heterchronic regulation of leaf differentiation.
	B	5	Kaladp0022s0030.1	AT4G00730.1	ARABIDOPSIS THALIANA HOMEODOMAIN PROTEIN	Encodes a homeodomain protein of the HD-GLABRA2 group. Involved in the accumulation of anthocyanin and in root development. Loss of function mutants have

						increased cell wall polysaccharide content.
	C	2	Kaladp0840s0019.1	AT4G18390.1	CYCLOIDEA AND PCF TRANSCRIPTION FACTOR 2 (TCP2)	TEOSINTE BRANCHED 1, cycloidea and PCF transcription factor 2
	D	1	Kaladp0048s0463.1	AT3G20670.1	HISTONE H2A	Encodes HTA13, a histone H2A protein.
	D	1	Kaladp0043s0282.1	AT1G79840.2	GLABRA 2 (GL2)	Glabra 2, a homeodomain protein affects epidermal cell identity including trichomes, root hairs, and seed coat. It also down-regulates seed oil content. Expressed in atrichoblasts and required to suppress root hair development. Also expressed abundantly during early seed development. Directly regulated by WER.
	D	1	Kaladp0039s0155.1	AT1G09200.1	HISTONE 3.1	Histone superfamily protein
	D	10	Kaladp0040s0359.1	AT4G21750.1	MERISTEM LAYER 1 (ATML1)	Encodes a homeobox protein similar to GL2. It is expressed in both the apical and basal daughter cells of the zygote as well as their progeny. Expression is detected starting

						the two-celled stage of embryo development and is later restricted to the outermost, epidermal cell layer from its inception. Its promoter is highly modular with each region contributing to specific aspects of the gene's spatial and temporal expression. Double mutant analysis with PDF2, another L1-specific gene, suggests that their functions are partially redundant and the absence of both of the genes result in abnormal shoot development.
	D	2	Kaladp0076s0024.1	AT2G36610.1	ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 22 (ATHB22)	Encodes a homeodomain leucine zipper class I (HD-Zip I) protein.
	D	1	Kaladp0093s0030.1	AT4G00730.1	ANTHOCYANINLESS 2 (ANL2)	Encodes a homeodomain protein of the HD-GLABRA2 group. Involved in the accumulation of anthocyanin and in root development. Loss of function mutants have increased cell wall polysaccharide content.

	D	1	Kaladp0060s0288.1	AT2G37180.1	PLASMA MEMBRANE INTRINSIC PROTEIN 2 (PIP2C)	a member of the plasma membrane intrinsic protein PIP2. functions as aquaporin and is involved in desiccation.
	D	1	Kaladp0048s0578.1	AT1G53310.1	PEP(PHOSPHOENOLPYRUVATE) CARBOXYLASE 1 (ATPPC1)	Encodes one of four Arabidopsis phosphoenolpyruvate carboxylase proteins.Plays an important role in carbon and nitrogen metabolism.
	D	1	Kaladp0060s0345.1	ATCG00650.1	RIBOSOMAL PROTEIN S18	chloroplast-encoded ribosomal protein S18
	D	1	Kaladp0053s0264.1	AT3G07160.1	GLUCAN SYNTHASE-LIKE 10 (GSL10)	Encodes GSL10, a member of the Glucan Synthase-Like (GSL) family believed to be involved in the synthesis of the cell wall component callose. GSL10 is required for male gametophyte development and plant growth. Has a role in entry of microspores into mitosis. GSL10 mutation leads to perturbation of microspore division symmetry, irregular callose deposition and failure of generative cell engulfment by the

						vegetative cell cytoplasm. Also refer to GSL8 (At2g36850).
	D	1	Kaladp0630s0026.1	AT3G14980.1	INCREASED DNA METHYLATION 1 (IDM1)	IDM1 is a histone H3 acetyltransferase that is capable of recognizing methylated DNA through its MBD domain and recognizing unmethylated histone H3K4 through its PHD domain. It negatively regulates DNA demethylation, preventing DNA hypermethylation of highly homologous multicopy genes and other repetitive sequences.
	D	1	Kaladp0041s0063.1	AT5G08200.1	peptidoglycan-binding LysM domain-containing protein	peptidoglycan-binding LysM domain-containing protein
	D	1	Kaladp0056s0019.1	AT5G07540.1	GLYCINE-RICH PROTEIN 16 (GRP16)	encodes a glycine-rich protein that is expressed only in flowers during a specific developmental stage (flower stages 11 and 12).
	D	1	Kaladp0071s0337.1	AT3G03460.1	BRAHMA-INTERACTING PROTEIN 1 (BRIP1)	One of two paralogous GLTSCR domain containing proteins and a core component of

						SWI/SNF complexes. Interacts with BRM and may be responsible for ensuring proper complex assembly and association with chromatin. Function is dependent upon the GLTSCR domain.
	N/A	1	Kaladp0029s0137.1	AT5G27640.2	ARABIDOPSIS THALIANA TRANSLATION INITIATION FACTOR 3B1 (ATEIF3B-1)	encodes a member of eukaryotic translation initiation factor 3B family.
	N/A	1	Kaladp0101s0273.1	AT5G15150.1	ATHB-3	homeobox-containing gene with an unusual feature: a leucine zipper motif adjacent to the carboxyl-terminal of the homeodomain structure. This gene is expressed primarily in the cortex of the root and the stem.
	N/A	1	Kaladp1246s0003.1	AT2G02760.1	UBIQUITIN-CONJUGATING ENZYME 2 (UBC2)	ubiquitin conjugating enzyme UBC2. Homolog of the yeast RAD6 gene.
	N/A	1	Kaladp0039s0092.1	AT4G15530.5	PYRUVATE ORTHOPHOSPHATE DIKINASE (PPDK)	Encodes a dual-targeted protein believed to act as a pyruvate, orthophosphate dikinase. These enzymes are normally

						associated with C4 photosynthesis which does not occur in Arabidopsis. However, PPDK may play a role in remobilizing nitrogen during leaf senescence in Arabidopsis. The product of the long transcript (.1 gene model) was shown to be targeted to the chloroplast, whereas the shorter transcript (no targeting sequence) accumulates in the cytosol. The two proteins were also found to be expressed in slightly different tissues.
	N/A	1	Kaladp0050s0155.1	AT4G19650.1	MTERF26	Mitochondrial transcription termination factor family protein
***KfPEPC1* promoter (-511 to -710)**	C	2	Kaladp0492s0002.1	AT3G54890.1	PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 1 (LHCA1)	Encodes a component of the light harvesting complex associated with photosystem I
	C	2	Kaladp0011s1344.1	AT3G15030.1	TCP FAMILY TRANSCRIPTION FACTOR 4 (TCP4)	Arabidopsis thaliana TCP family transcription factor. Regulated

						by miR319. Involved in heterchronic regulation of leaf differentiation.
	C	2	Kaladp0039s0668.1	AT3G41768.1	rRNA	ribosomal RNA (noncoding RNA)
	D	1	Kaladp0011s0537.1	AT2G37400.1	Tetratricopeptide repeat (TPR)-like superfamily protein	Tetratricopeptide repeat (TPR)-like superfamily protein
	D	1	Kaladp0067s0135.1	AT3G51240.1	FLAVANONE 3-HYDROXYLASE (F3H)	Encodes flavanone 3-hydroxylase that is coordinately expressed with chalcone synthase and chalcone isomerases and is involved in flavonoid biosynthesis. Not responsive to auxin or ethylene stimulus (qRT-PCR).
	D	1	Kaladp0374s0001.1	AT4G26270.1	PHOSPHOFRUCTOKINASE 3 (PFK3)	PHOSPHOFRUCTOKINASE 3 (PFK3)
	D	1	Kaladp0045s0054.1	AT4G10340.1	LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5 (LHCB5)	photosystem II encoding the light-harvesting chlorophyll a/b binding protein CP26 of the antenna system of the photosynthetic apparatus The mRNA is cell-to-cell mobile.
	D	1	Kaladp0131s0010.1	AT3G01930.2	Major facilitator superfamily protein	Major facilitator superfamily protein

	D	1	Kaladp0096s0056.1	AT1G80325.1	Natural antisense transcript overlaps with AT1G80320	Natural antisense transcript overlaps with AT1G80320
	D	1	Kaladp0403s0001.1	AT5G07580.1	ERF106	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5.
	D	1	Kaladp0550s0009.1	AT1G08380.1	PHOTOSYSTEM I SUBUNIT O (PSAO)	Encodes subunit O of photosystem I.
	D	1	Kaladp0630s0020.1	AT3G15030.1	TCP FAMILY TRANSCRIPTION FACTOR 4 (TCP4)	Arabidopsis thaliana TCP family transcription factor. Regulated by miR319. Involved in heterchronic regulation of leaf differentiation.
	D	1	Kaladp0008s0657.1	AT1G27180.1	disease resistance protein (TIR-NBS-LRR class)	disease resistance protein (TIR-NBS-LRR class)
	D	1	Kaladp0040s0679.1	AT1G04160.1	ARABIDOPSIS THALIANA MYOSIN XI B	Encodes a member of the type XI myosin protein family involved in root hair elongation.
	D	1	Kaladp0808s0063.1	AT5G47110.1	LIGHT-HARVESTING-LIKE 3:2	Encodes a light-harvesting-like protein that is involved in chlorophyll and tocopherol

						biosynthesis anchoring geranylgeranyl reductase in the thylakoid membrane.
	D	1	Kaladp0016s0220.1	AT4G38460.1	GERANYL(GERANYL)DIPHOSPHATE SYNTHASE 12	Encodes a type II small subunit of the heteromeric geranyl(geranyl) diphosphate synthase that is localized to the chloroplast, expressed in petals and sepals and is involved in monoterpene biosynthesis. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0058s0293.1	AT3G26810.1	AUXIN SIGNALING F-BOX 2 (AFB2)	Auxin F box protein, the dominant auxin receptor in roots.
	D	1	Kaladp0010s0064.1	AT1G28240.1	MUCILAGE-RELATED 70 (MUCI70)	strawberry notch protein (DUF616)
	D	1	Kaladp0748s0002.1	AT5G49530.1	SIN-like family protein	SIN-like family protein
	D	1	Kaladp0020s0147.1	AT4G36945.1	PLC-like phosphodiesterases superfamily protein	PLC-like phosphodiesterases superfamily protein
	D	1	Kaladp0042s0043.1	AT1G09740.1	Adenine nucleotide alpha hydrolases-like superfamily protein	Adenine nucleotide alpha hydrolases-like superfamily protein
	D	2	Kaladp0082s0119.1	AT2G44840.1	ETHYLENE-RESPONSIVE ELEMENT BINDING FACTOR 13 (ERF13)	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor

						family. The protein contains one AP2 domain. There are 18 members in this subfamily including ATERF-1, ATERF-2, AND ATERF-5. The mRNA is cell-to-cell mobile.
	D	1	Kaladp0035s0032.1	AT3G22240.1	CYSTEINE-RICH TRANSMEMBRANE MODULE 9 (ATHCYSTM9)	cysteine-rich/transmembrane domain PCC1-like protein
	D	1	Kaladp0067s0312.1	AT4G20430.1	Subtilase family protein	Subtilase family protein
	D	1	Kaladp0020s0142.1	AT3G19130.1	RNA-BINDING PROTEIN 47B (ATRBP47B)	RBP47B, is a component of the stress granule proteome and interacts with 2',3'-cAMP.
	D	1	Kaladp0058s0427.1	AT3G02760.1	Class II aaRS and biotin synthetases superfamily protein	Class II aaRS and biotin synthetases superfamily protein
	D	1	Kaladp0011s1349.1	AT4G05475.1	RNI-like superfamily protein	RNI-like superfamily protein
	D	1	Kaladp0059s0142.1	AT1G76810.1	EIF5B1	eukaryotic translation initiation factor 2 (eIF-2) family protein
	D	1	Kaladp0011s1308.1	AT5G54190.1	PROTOCHLOROPHYLLIDE OXIDOREDUCTASE A (PORA)	light-dependent NADPH:protochlorophyllide oxidoreductase A
	N/A	1	Kaladp0060s0465.1	AT3G17390.1	METHIONINE ADENOSYLTRANSFERASE 4 (MAT4)	S-adenosylmethionine synthetase

	N/A	1	Kaladp0040s0441.1	AT1G74470.1	Encodes for a multifunctional protein with geranylgeranyl reductase activity shown to catalyze the reduction of prenylated geranylgeranyl-chlorophyll a to phytyl-chlorophyll a (chlorophyll a) and free geranylgeranyl pyrophosphate to phytyl pyrophosphate. The mRNA is cell-to-cell mobile	Encodes for a multifunctional protein with geranylgeranyl reductase activity shown to catalyze the reduction of prenylated geranylgeranyl-chlorophyll a to phytyl-chlorophyll a (chlorophyll a) and free geranylgeranyl pyrophosphate to phytyl pyrophosphate. The mRNA is cell-to-cell mobile
	N/A	1	Kaladp1222s0026.1	AT1G79690.1	NUDIX HYDROLASE HOMOLOG 3 (ATNUDT3)	Encodes a dual activity enzyme which catalyses the hydrolysis of a peptide bond and of a phosphate bond, acting both as a dipeptidyl peptidase III and an atypical Nudix hydrolase.
	N/A	2	Kaladp0740s0002.1	AT5G01530.1	LIGHT HARVESTING COMPLEX PHOTOSYSTEM II (LHCB4.1)	light harvesting complex photosystem II
	N/A	1	Kaladp0067s0152.1	AT3G57390.1	AGAMOUS-LIKE 18 (AGL18)	encodes a MADS-box containing protein likely to be a transcription factor that is expressed in endosperm and developing gametophytes. The protein sequence is

						most similar to that of AGL15, which is expressed in developing embryos.
	N/A	1	Kaladp0840s0018.1	AT5G46170.1	F-box family protein	F-box family protein
	N/A	1	Kaladp0053s0064.1	AT2G08665.1	Natural antisense transcript overlaps with AT2G34420	Natural antisense transcript overlaps with AT2G34420
	N/A	1	Kaladp0050s0284.1	AT1G68830.1	STT7 HOMOLOG STN7 (STN7)	STN7 protein kinase; required for state transitions, phosphorylation of the major antenna complex (LHCII) between PSII and PSI, and light adaptation. STN7 is involved in state transitions.
	N/A	1	Kaladp0033s0048.1	AT2G39730.1	RUBISCO ACTIVASE (RCA)	Rubisco activase, a nuclear-encoded chloroplast protein that consists of two isoforms arising from alternative splicing in most plants. Required for the light activation of rubisco. Involved in jasmonate-induced leaf senescence.
	N/A	1	Kaladp0011s1131.1	AT2G08665.1	Natural antisense transcript overlaps with AT2G34420	Natural antisense transcript overlaps with AT2G34420
	N/A	1	Kaladp0515s0040.1	AT1G61520.1	PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 3 (LHCA3)	PSI type III chlorophyll a/b-binding protein (Lhca3*1) The mRNA is cell-to-cell mobile.

	N/A	1	Kaladp0081s0036.1	AT5G39740.1	RIBOSOMAL PROTEIN L5 B 9RPL5B0	Encodes a ribosomal protein RPL5B that is involved in ribosome biogenesis and plays a role in organ size control by promoting cell proliferation and preventing compensation in normal leaf development.
	N/A	2	Kaladp0095s0055.1	AT1G53310.1	PEP(PHOSPHOENOLPYRUVATE) CARBOXYLASE 1 (ATPEPC1)	Encodes one of four Arabidopsis phosphoenolpyruvate carboxylase proteins.Plays an important role in carbon and nitrogen metabolism
	N/A	1	Kaladp0021s0051.1	AT1G30380.1	PHOTOSYSTEM I SUBUNIT K (PSAK)	Encodes subunit K of photosystem I reaction center. The mRNA is cell-to-cell mobile.
***KfPEPC1* promoter (-681 to -880)**	A	5	Kaladp0011s1344.1	AT3G15030.1	TCP FAMILY TRANSCRIPTION FACTOR 4 (TCP4)	Arabidopsis thaliana TCP family transcription factor. Regulated by miR319. Involved in heterchronic regulation of leaf differentiation.
	D	1	Kaladp0047s0151.1	AT1G10960.1	FERREDOXIN 1 (FD1)	Phosphatidylinositol 4-phosphate 5-kinase (PIP5K) enzyme family member.

	D	1	Kaladp0066s0002.1	AT3G17000.1	UBIQUITIN-CONJUGATING ENZYME 32 (UBC32)	Group XIV ubiquitin-conjugating enzyme that functions negative regulation of drought stress.
	D	1	Kaladp0025s0037.1	AT1G09200.1	HISTONE 3.1 (H3.1)	Histone superfamily protein
	D	2	Kaladp0074s0058.1	AT5G65360.1	HISTONE 3.1 (H3.1)	Histone superfamily protein
	D	1	Kaladp0079s0159.1	AT4G40040.1	HISTONE 3.3 (H3.3)	Histone superfamily protein
	D	1	Kaladp1246s0007.1	AT5G15150.1	HOMEOBOX 3 (ATHB-3)	homeobox-containing gene with an unusual feature: a leucine zipper motif adjacent to the carboxyl-terminal of the homeodomain structure. This gene is expressed primarily in the cortex of the root and the stem.
	D	1	Kaladp0037s0474.1	AT3G60160.1	ATP-BINDING CASSETTE C9 (ABCC9)	member of MRP subfamily
	D	1	Kaladp0073s0058.1	AT1G68720.1	ARABIDOPSIS THALIANA TRNA ADENOSINE DEAMINASE A (ATTADA)	Encodes the chloroplastic A-to-I tRNA editing enzyme.
	D	1	Kaladp0080s0110.1	AT5G22060.1	ARABIDOPSIS THALIANA DNAJ HOMOLOGUE 2 (ATJ2)	Co-chaperonin similar to E. coli DnaJ
	D	1	Kaladp0192s0037.1	AT3G15110.1	transmembrane protein	transmembrane protein

	D	1	Kaladp0067s0212.1	AT1G64720.1	MEMBRANE RELATED PROTEIN CP5 (CP5)	membrane related protein CP5
	D	1	Kaladp0024s0173.1	AT3G02470.1	S-ADENOSYLMETHIONINE DECARBOXYLASE (SAMDC)	Encodes a S-adenosylmethionine decarboxylase involved in polyamine biosynthesis.
	D	1	Kaladp0008s0054.1	AT4G12490.1	AZI3	Encodes a member of the AZI family of lipid transfer proteins. Contains a PRR domain that appears to be required for localization to the chloroplast.
	D	1	Kaladp0039s0024.1	AT2G19790.1	ADAPTOR PROTEIN COMPLEX 4S (AP4S)	Encodes a component of the AP4 complex and is involved in vacuolar sorting of storage proteins.
	D	1	Kaladp0085s0021.1	AT2G47015.1	MICRORNA408 (MIR408)	Encodes a microRNA that targets both a Laccase and Plantacyanin-like family member. MicroRNAs are regulatory RNAs with a mature length of ~21-nucleotides that are processed from hairpin precursors by Dicer-like enzymes. MicroRNAs can negatively regulate gene expression by attenuating translation or by

						directing mRNA cleavage.Mature sequence: AUGCACUGCCU CUUCCCUGGC
	D	1	Kaladp0101s0026.1	AT2G45860.1	hypothetical protein	hypothetical protein
	N/A	1	Kaladp0006s0117.1	ATCG01180.1	RIBOSOMAL RNA23S (RRN23S.2)	chloroplast-encoded 23S ribosomal RNA

Travis M. Garcia Curriculum Vitae

Department of Biochemistry
University of Nevada, Reno
1664 N. Virginia Street
Reno, Nevada, USA 89557
775-213-3734
tgarcia@nevada.unr.edu

Education

PhD Biochemistry, 2021
University of Nevada, Reno
Research advisor: Dr. John C. Cushman
Thesis title: Characterizing the transcriptional regulation of crassulacean acid metabolism in *Kalanchoe*

B.S. Biochemistry and Molecular Biology, *cum laude*, 2011
University of Nevada, Reno
Research advisor: Dr. David K. Shintani
Thesis title: Extraction and visualization of three rubber particle associated small rubber particle proteins from *Taraxacum kok-saghyz* using 2D gel electrophoresis and mass spectrometry

High School Advanced High School Diploma, 2007
Churchill County High School
Fallon, Nevada, USA 89406

Experience

Graduate research assistant 2013 to 2021
University of Nevada, Reno, USA, Dr. John C. Cushman Lab

Doctoral dissertation research project focused on investigating the transcriptional regulation of plant crassulacean acid metabolism (CAM), with a focus on the phospho*enol*pyruvate carboxylase (*Ppc*) gene, through transcription factor-DNA interaction studies in the model CAM species *Kalanchoe*.

Graduate teaching assistant 2016 to 2021
University of Nevada, Reno, Department of Biochemistry

Worked as a graduate teaching assistant for both undergraduate and graduate molecular biology laboratory courses. Responsibilities included instructing individual lab sections, grading, and helping students learn techniques and troubleshoot experiments.

Graduate research assistant 2012 to 2013
University of Nevada, Reno, USA, Dr. David K. Shintani Lab

Continued the project described below to overexpress key phytate pathway genes in *Lemna.*

Research affiliate 2011 to 2012
University of Nevada, Reno, USA, Dr. David K. Shintani Lab

Worked on a project to induce over-accumulation of phytate in the small aquatic plant duckweed (*Lemna*) for the purpose of enhanced phosphate removal from polluted water. Project involved overexpression of key phytate pathway genes.

Undergraduate senior thesis student 2010 to 2011
University of Nevada, Reno, USA, Dr. David K. Shintani Lab

Bachelor's senior thesis involved extraction and analysis of proteins involved in the production of natural latex rubber from the Russian Dandelion species *Taraxacum kok-saghyz* to better understand latex rubber biosynthesis.

Mentoring Activities

Mentored biotechnology B.S./M.S. degree student 2012 to 2013

While working as a graduate research assistant on the *Lemna* phytate project described above, I mentored and guided a B.S./M.S. degree-seeker who worked alongside me on this project. Her research presentation on our work was selected for an oral presentation at a University of Nevada, Reno biotechnology symposium.

Mentored biochemistry B.S. degree student 2011 to 2012

While working as a university research affiliate after receiving my B.S. degree and before starting graduate school, I mentored a biochemistry B.S. degree senior thesis student on the *Lemna* phytate project. His research was presented as a poster display at the University of Nevada, Reno's annual undergraduate biosciences poster day.

Publications

Amin, A. B., Rathnayake, K. N., Yim, W. C., Garcia, T. M., Wone, B., Cushman, J. C., & Wone, B. W. M. (2019). Crassulacean Acid Metabolism Abiotic Stress-Responsive Transcription Factors: A Potential Genetic Engineering Approach for Improving Crop Tolerance to Abiotic Stress. *Frontiers in Plant Science*, *10*, 129. https://doi.org/10.3389/fpls.2019.00129

Garcia, T. M., Heyduk, K., Kuzmick, E., & Mayer, J. A. (2014). Crassulacean acid metabolism biology. *New Phytologist*, *204*(4), 738–740. https://doi.org/10.1111/nph.13127

Yang, X., Cushman, J. C., Borland, A. M., Edwards, E. J., Wullschleger, S. D., Tuskan, G. A., Owen, N. A., Griffiths, H., Smith, J. A. C., De Paoli, H. C., Weston, D. J., Cottingham, R., Hartwell, J., Davis, S. C., Silvera, K., Ming, R., Schlauch, K., Abraham, P., Stewart, J. R., … Holtum, J. A. M. (2015). A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist*, *207*(3), 491–504. https://doi.org/10.1111/nph.13393

Yang, X., Hu, R., Yin, H., Jenkins, J., Shu, S., Tang, H., Liu, D., Weighill, D. A., Cheol Yim, W., Ha, J., Heyduk, K., Goodstein, D. M., Guo, H.-B., Moseley, R. C., Fitzek, E., Jawdy, S., Zhang, Z., Xie, M., Hartwell, J., … Tuskan, G. A. (2017). The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications*, *8*(1), 1899. https://doi.org/10.1038/s41467-017-01491-7

Zhang, J., Hu, R., Sreedasyam, A., Garcia, T. M., Lipzen, A., Wang, M., Yerramsetty, P., Liu, D., Ng, V., Schmutz, J., Cushman, J. C., Borland, A. M., Pasha, A., Provart, N. J., Chen, J.-G., Muchero, W., Tuskan, G. A., & Yang, X. (2020). Light-responsive expression atlas reveals the effects of light quality and intensity in Kalanchoë fedtschenkoi, a plant with crassulacean acid metabolism. *GigaScience*, *9*(3), giaa018. https://doi.org/10.1093/gigascience/giaa018

Talks

Travis Garcia, "Characterizing the Transcriptional Regulation of Crassulacean Acid Metabolism in *Kalanchoe*"; seminar talk and PhD defense at the University of Nevada, Reno, November 2021.

Travis Garcia, "Characterizing the Transcriptional Regulation of Crassulacean Acid Metabolism in *Kalanchoe*"; seminar talk at the University of Nevada, Reno, September 2014.

Poster Presentations

Travis Garcia, "Characterization of the Transcriptional Regulation of Crassulacean Acid Metabolism (CAM) in *Kalanchoe laxiflora* Mesophyll Tissue"; poster presentation at the Biology of CAM Plants Conference, Phoenix, Arizona, April 2018

Travis Garcia, "Characterization of the Transcriptional Regulation of Crassulacean Acid Metabolism (CAM) in *Kalanchoe laxiflora* Mesophyll Tissue"; poster presentation at the Plant and Animal Genome Conference (PAG), San Diego, California, January 2017

Travis Garcia, "Characterization of the Transcriptional Regulation of Crassulacean Acid Metabolism (CAM) in *Kalanchoe laxiflora* Mesophyll Tissue"; poster presentation at the Plant and Animal Genome Conference (PAG), San Diego, California, January 2016

Travis Garcia, "Characterization of the Transcriptional Regulation of Crassulacean Acid Metabolism (CAM) in *Kalanchoe laxiflora* Mesophyll Tissue"; poster presentation at the

34th New Phytologist Symposium: Systems biology and ecology of CAM plants, Tahoe City, California, July 2014

Travis Garcia, "Characterization of the Transcriptional Regulation of Crassulacean Acid Metabolism (CAM) in *Kalanchoe laxiflora* Mesophyll Tissue"; poster presentation at the Western Section Meeting of the American Society of Plant Biologists, Santa Clara, California, May 2014