

Diurnal Cycling Transcription Factors of Pineapple Revealed by Genome-Wide Annotation and Global Transcriptomic Analysis

Anupma Sharma¹, Ching Man Wai^{2,3}, Ray Ming^{2,3}, and Qingyi Yu^{1,3,4,*}

¹Texas A&M Agrilife Research Center at Dallas, Texas A&M University System, Dallas

²Department of Plant Biology, University of Illinois at Urbana-Champaign

³Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Haixia Institute of Science and Technology, Fujian Agriculture and Forestry University, Fuzhou, Fujian Province, China

⁴Department of Plant Pathology and Microbiology, Texas A&M University

*Corresponding author: E-mail: qyu@ag.tamu.edu.

Accepted: August 22, 2017

Data deposition: The RNA-Seq data used in this study have been deposited and publicly available in NCBI under BioProject PRJNA305042.

Abstract

Circadian clock provides fitness advantage by coordinating internal metabolic and physiological processes to external cyclic environments. Core clock components exhibit daily rhythmic changes in gene expression, and the majority of them are transcription factors (TFs) and transcription coregulators (TCs). We annotated 1,398 TFs from 67 TF families and 80 TCs from 20 TC families in pineapple, and analyzed their tissue-specific and diurnal expression patterns. Approximately 42% of TFs and 45% of TCs displayed diel rhythmic expression, including 177 TF/TCs cycling only in the nonphotosynthetic leaf tissue, 247 cycling only in the photosynthetic leaf tissue, and 201 cycling in both. We identified 68 TF/TCs whose cycling expression was tightly coupled between the photosynthetic and nonphotosynthetic leaf tissues. These TF/TCs likely coordinate key biological processes in pineapple as we demonstrated that this group is enriched in homologous genes that form the core circadian clock in *Arabidopsis* and includes a STOP1 homolog. Two lines of evidence support the important role of the STOP1 homolog in regulating CAM photosynthesis in pineapple. First, STOP1 responds to acidic pH and regulates a malate channel in multiple plant species. Second, the cycling expression pattern of the pineapple STOP1 and the diurnal pattern of malate accumulation in pineapple leaf are correlated. We further examined duplicate-gene retention and loss in major known circadian genes and refined their evolutionary relationships between pineapple and other plants. Significant variations in duplicate-gene retention and loss were observed for most clock genes in both monocots and dicots.

Key words: *Ananas comosus*, CAM photosynthesis, phylogenomics, diurnal, circadian.

Introduction

Transcription factors (TFs) and transcription coregulators (TCs) play important roles in regulating plant growth and development, physiological and metabolic processes, cell cycle, and responses to biotic and abiotic stresses (Chen et al. 2002; Nakashima et al. 2009; Wilkins et al. 2009). TFs and TCs regulate gene expression directly or via a cascade of transcriptional regulation (Scott 2000). TFs control the expression of target genes through specific binding of TF to the *cis*-regulatory DNA elements (Arnone and Davidson 1997; Wray et al. 2003) while

TCs act by interacting with TFs and/or RNA polymerase. Most TFs can regulate numerous downstream target genes (Walhout 2006; Ishihama et al. 2016). As a result, the transcription of ~27,000 protein-coding genes in *Arabidopsis* genome is regulated by a relatively small number of TFs, around 1,500 TFs in total and proximately 6% of the estimated total number of genes in the genome (Riechmann et al. 2000). Genes are often regulated by more than one TF in a combinatorial manner to ensure precise spatial and temporal expression for appropriate functional outcomes (Narlikar and Ovcharenko 2009).

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Most TFs contain several functional domains, such as DNA-binding domains, protein–protein interaction domains, and domains that serve as intracellular trafficking signals (Frietze and Farnham 2011). DNA-binding domains are essential components that mediate the specificity of TF–DNA interaction (Franco-Zorrilla et al. 2014) and have been widely used for TF classification. Computational predictions of TF repertoires by searching for genes containing DNA-binding domains have been used in several plant species, including *Arabidopsis* (Riechmann et al. 2000), rice (Gao et al. 2006), maize, and foxtail millet (Lin et al. 2014).

Determining when and where genes are expressed and how their expressions are regulated are of critical importance to understanding the molecular mechanisms underlying plant growth and development. Tissue-specific patterns of gene expression play fundamental roles in tissue development, and determining distinctive features of cell types and functions. Therefore, identification of tissue-specific gene regulatory networks can yield insights into the molecular basis of a tissue's development and function. It is a widespread phenomenon that genes functioning in common processes are highly coordinately expressed (Niehrs and Pollet 1999). Ascertaining syexpression groups would also represent an important step towards delineating the transcriptional networks of functionally interacting genes.

The rhythmic environmental fluctuations caused by the planet's 24 h rotation have driven the evolution of the circadian clock in almost all living organisms on Earth. The circadian clock is one of the most important biological regulators controlling a wide range of physiological, developmental, and metabolic processes (Paranjpe and Sharma 2005). It can maintain diurnal rhythms in constant conditions and in the absence of external time-giving cues. Global profiling of transcriptomes in rice and poplar revealed 2- to 4-fold fewer rhythmic transcripts in the circadian (free-running) conditions relative to their respective diurnal conditions (Filichkin et al. 2011).

In plants, the genetics and molecular biology of circadian rhythms have been best characterized in *Arabidopsis thaliana*. The core elements that make up the circadian clock have been identified in *A. thaliana*. Circadian clocks are composed of three basic components, input pathways, rhythm-generating oscillators, and output pathways (Barak et al. 2000). Input pathways receive environmental cues and transmit them to circadian oscillators. Oscillators generate the circadian rhythm and synchronize the phase of rhythm with the outside environment. Output pathways regulate various physiological, metabolic, and developmental processes. Expression of core clock components is regulated by a complex network of interlocked transcriptional/translational feedback loops to ensure robust sustained circadian rhythmicity (Troein et al. 2009; Haydon et al. 2011; Nohales and Kay 2016).

Circadian-regulated transcription is widespread in plants. Studies using different approaches revealed that

approximately one-third of the transcriptome is regulated by the circadian clock in *A. thaliana* (Michael and McClung 2003; Covington et al. 2008; Nakamichi et al. 2009; Dong et al. 2011). The circadian clock not only regulates transcription of pathways associated with metabolism, growth, and development (Smith et al. 2004; Bläsing et al. 2005; Covington et al. 2008), but also modulates the response to the abiotic and biotic stresses (Fowler et al. 2005; Wilkins et al. 2010; Wang et al. 2011; Goodspeed et al. 2012). TFs and TCs are key regulators of gene expression. Therefore, it is not surprising that most circadian clock components are TFs (Wang et al. 1997; Schaffer et al. 1998; Para et al. 2007; Pruneda-Paz et al. 2009; Rawat et al. 2009; Nakamichi et al. 2010; Dai et al. 2011; Gendron et al. 2012) or TCs (Xie et al. 2014).

Pineapple (*Ananas comosus* (L.) Merr.) is an important tropical fruit crop utilizing Crassulacean acid metabolism (CAM), an efficient photosynthetic CO₂ fixation pathway that evolved in some plants as an adaptation to arid habitats (Cushman 2001). CAM involves a mechanism by which assimilation of CO₂ is temporally separated from the incorporation of CO₂ into carbohydrates. Therefore, CAM plants have higher water-use efficiency than C3 and C4 plants by fine control of nocturnal opening and diurnal closure of stomata. All the enzymes involved in CAM are also found in C3 plants. CAM photosynthesis evolved from C3 through reorganization of metabolic processes (Crayn et al. 2004; West-Eberhard et al. 2011). Understanding the circadian regulation of CAM metabolic activities is the key to fully elucidating CAM and successful application of CAM into crop improvement.

Pineapple genome is fully sequenced (Ming et al. 2015). The availability of high quality genomic and transcriptomic resources for pineapple (Fang et al. 2016; Paull et al. 2016; Singh et al. 2016; Wai et al. 2016a, 2016b; Zheng et al. 2016; Zhang et al. 2016) make it an ideal system to study the repertoire of TFs and their temporal and tissue-specific expression to gather insights into the molecular mechanisms governing the rhythm especially of the CO₂ metabolism in CAM plants. In this study, we identified and classified pineapple TFs and TCs based on their conserved signature domains, evaluated their tissue-specific and diurnal expression patterns, and predicted candidates related to the circadian rhythm. Our results provide a solid foundation for further systematic characterization of pineapple TFs and TCs in their biological context.

Materials and Methods

Plant Materials

Pineapple varieties F153 and MD-2 were grown and maintained at the Kunia Station of the Hawaii Agriculture Research Center and the field of Dole Plantation at Wahiawa on Oahu Island (Hawaii), respectively. Pineapple leaf, root, and flower tissues collected from cultivar F153 and fruits of five different

developing stages harvested from cultivar MD-2 were used for tissue-specific gene expression analysis. The green leaf tip and white leaf base tissues were harvested from cultivar MD-2 at thirteen time points over a 24-h period and used for temporal gene expression profiling. The harvested tissues were snap-frozen by dropping directly into liquid nitrogen and stored in a freezer at -80°C until RNA extraction.

RNA Extraction and RNA-Seq Library Construction

Total RNA was extracted from the fine powder of the ground tissues using Qiagen RNeasy Plant Mini Kit (Qiagen, <http://www.qiagen.com/>; last accessed August 30, 2017) following the manufacturer's protocol. DNA contamination was then eliminated using Invitrogen Ambion DNA-free DNA Removal Kit (Life Technologies, <http://www.lifetechnologies.com/>; last accessed August 30, 2017). RNA-Seq libraries were constructed using Illumina TruSeq Stranded RNA Sample Preparation kit (Illumina, <http://www.illumina.com/>; last accessed August 30, 2017) and sequenced on an Illumina HiSeq2500 using 100-nt pair-end sequencing mode.

Sequencing Read Processing and Gene Expression Analysis

The raw RNA-Seq reads generated for both tissue-specific and time course experiments were deposited into NCBI under BioProject PRJNA305042. Raw reads were trimmed with TRIMMOMATIC v0.30 to remove Illumina adapter sequence, any base below quality phred score 3 and any read less than 36 bp in length (Bolger et al. 2014). The trimmed paired-end reads of each sample were aligned to repeat-masked pineapple assembly using TopHat (v2.1.1) with default settings (Trapnell et al. 2012). The uniquely mapped reads were then used to calculate the number of reads falling into each gene and normalized to fragments per kilobase of exon per million fragments mapped (FPKM) using Cufflinks (v2.2.1) followed by Cuffnorm (v2.2.1) with default settings and pineapple gene model annotation provided.

Identification of TFs and TCs in Pineapple Genome

An HMM database was compiled for the signature domains (DNA binding, auxiliary and forbidden domains) of TFs and TCs enlisted by Lin et al. (2014). HMM models were either downloaded from Pfam 27.0 or self-built for domains that are not included in the Pfam 27.0 database using the DNA binding domain alignments downloaded from PlantTFDB 3.0 (Jin et al. 2014). HMMER 3.0 hmmscan was used to search the pineapple gene models version 3 (pineapple.v3.20141007.-pep.fasta) (Ming et al. 2015) against this database. TFs were classified based on the domain and sequence cut-off thresholds and family classification rules used by Lin et al. (2014). The empirical cut-off thresholds for self-built HMMs were the lowest scores obtained by hmmscan against the respective TF/TC families in the PlantTFDB 3.0. The supplementary table

S1, Supplementary Material online, contains the details of HMM and the cut-off thresholds used in the study. The homologs of pineapple TF/TCs in the public TF/TC databases were identified using blastp of the BLAST+ 2.2.30 package.

Gene Expression Analysis of Pineapple TFs and TCs in Four Pineapple Tissues

The normalized FPKM values were calculated using the Cufflinks/Cuffnorm pipeline (<http://cufflinks.cbcb.umd.edu/>; last accessed August 30, 2017). The average of the normalized FPKM values across the five fruit libraries was used as the normalized FPKM value for the fruit tissue. Genes with no detectable expression (FPKM value "0" in all four tissues), or low expression (FPKM less than 10 in all tissues) were filtered. A pseudocount of 1 was added to the normalized FPKM values to avoid taking log of zeros in the downstream analysis. Cluster Affinity Search Technique (CAST) module in MeV (MultiExperiment Viewer) was used to cluster the log2-transformed expression values (FPKM + 1) using Pearson correlation distance metric and a threshold parameter of 0.9. Each cluster was further clustered hierarchically using Pearson correlation and complete linkage, and visualized as heatmaps in MeV.

Temporal Expression Profiling of TF and TC Genes

The RNA-Seq data for diurnal expression profiling is taken from Ming et al. (2015). This data was obtained from the green leaf tip and white leaf base tissues collected over a 24-h period, between October 24, 2013 (10 AM HST) and October 25, 2013 (9 AM HST) at times 6, 8, and 10 PM, midnight, 2, 4, 6, 8, and 10 AM, noon, 1, 3, and 4 PM, from field pineapple plants growing on Oahu Island (Hawaii). The time of sunset on October 24, 2013 was 6:01 PM HST and the civil twilight ended at 6:24 PM HST. The time of sunrise on October 25, 2013 was at 6:32 AM HST on October 25, 2013 and the civil twilight started at 6:09 AM HST. The time series expression data was analyzed using Haystack Version 2.0 (<http://haystack.mocklerlab.org/>; last accessed August 30, 2017) (Mockler et al. 2007) to identify best-fit model and phase of the expression using a correlation cut off 0.7, fold change cutoff 2, *P* value cutoff 0.05, and background cutoff 1. We derived the models for our specific time points using the hourly shift values of the predefined models of cycling transcripts given in the study by Endo et al. (2014). The amplitude of cycling TF and TC transcripts was estimated by subtracting the mean expression value from the maximum expression value and genes with amplitude less than ten were assumed to be noise and were filtered. Time lagged correlation (Pearson's *r* using CORREL function in Excel) between the stationary leaf time series and the lagged white time series were estimated by treated the white time series as circular and shifting it incrementally.

Identification of Orthologous Relationships of Core Clock Circadian Genes

The initial synteny-based orthology detection was done using Proteinortho-PoFF from the Proteinortho_v5.13 package (Lechner et al. 2014) and SynMap2 (<https://genomeevolution.org/CoGe/SynMap.pl>; last accessed August 30, 2017). To refine orthologous relationships, we conducted maximum likelihood phylogenetic analysis using PhyML with Smart Model Selection (Guindon et al. 2010).

Phylogenetic Analysis of Core Circadian Clock Genes

The protein sequences of *Arabidopsis* TFs known to function primarily within the circadian clock based on Hsu and Harmer (Hsu and Harmer 2014) were retrieved from the *Arabidopsis* TAIR10 database (<http://www.arabidopsis.org>; last accessed August 30, 2017) and used as queries to blastp (BLAST+ 2.2.30 package) search against the protein sequences of plant and algal genomes downloaded from Phytozome v11.0 (<https://phytozome.jgi.doe.gov/pz/portal.html>; last accessed August 30, 2017). The multiple sequence alignment of homologous proteins was generated using MUSCLE v3.8.31 (Edgar 2004). To improve the alignments, we manually checked the initial alignments and removed poorly aligned regions at the ends of multiple sequence alignments generated by the full-length protein sequences using BioEdit v7.2.0 (Hall 2011). The resulting alignments were then analyzed by PhyML with Smart Model Selection (<http://www.atgc-montpellier.fr/phyml-sms/>; last accessed August 30, 2017) (Guindon et al. 2010). Branch support values were calculated using the SH-like approximate likelihood ratio test (aLRT). The resulting trees were visualized using HyperTree v1.2.2 (Bingham and Sudarsanam 2000).

Results

Genome-Wide Prediction and Classification of TFs and TCs in Pineapple

Lin et al. (2014) compiled a comprehensive list of 67 TF families and 29 TC families based on the survey of signature domains in all available TF databases and other resources, and identified TFs in maize and foxtail millet genomes based on selected hmm bit score thresholds and a set of classification rules. We used the same thresholds and classification rules to identify and classify TFs in the pineapple genome, except that we separately grouped TFs with more than two copies of B3 domain in REM family as suggested by Romanel et al. (2009) and the zf-B_box containing proteins that do not have a CTT domain in BBX1 family as suggested by Huang et al. (2012) and Gangappa and Botto (Gangappa and Botto 2014) (supplementary tables S1 and S2, Supplementary Material online). We identified 1,398 TFs from 67 TF families and 80 TCs from 20 TC families in the

pineapple genome (fig. 1 and supplementary tables S3 and S4, Supplementary Material online). The most abundant TC families of pineapple were AUX/IAA and OFP, accounting for 50% of the total TCs, and the most abundant TF families of pineapple were bHLH, MYB, ERF, NAC, C2H2, MYB_related, FAR1, WRKY, and bZIP that together make up ~50% of the total pineapple TFs. No TF and TC were identified from the TF families HMG1Y and NZZ/SPL and TC families Med11, Med15, Med15_fungi, Med18, Med19, Med3, Med5, Med8, and Med9 in pineapple genome.

Although most pineapple TFs and TCs contain the canonical DNA binding and auxiliary domains enlisted by Lin et al. (2014), there are few notable exceptions. First, eight TFs of the sixteen GRF family members lacked a QLQ auxiliary domain. Although it has been reported that QLQ auxiliary domain was a conserved feature of most GRF family members (Kim et al. 2003) and it might play roles in protein–protein interaction (van der Knaap et al. 2000), some GRF family members that have been identified in plants lacked a QLQ domain. It still remains unclear whether these proteins lacking a QLQ domain are functional (Omidbakhshfard et al. 2015). Second, three TFs containing the GATA and CCT DNA binding domains, and one containing the GATA and FAR1 DNA binding domains were identified. We classified these four proteins into the GATA TF family as it was reported that some GATA family members contained the CCT and FAR1 domains (Reyes et al. 2004). Third, we classified five SRF-TF domain proteins to the MIKC group despite lacking a K-box because the MADS box domain from these proteins was grouped in the MIKC group based on MADS box phylogenetic analysis. It is plausible that missing K-box in these MIKC group members may have resulted from a deletion or truncation in the protein. Fourth, three B3: ARF TFs contained a presumably diverged (or truncated) B3 domain that had a lower bit score threshold than the default.

Homologs of Pineapple TFs and TCs in Other Plant Species

To further validate the pineapple TFs and TCs identified in this study, we searched for their best homologs using blastp in the public plant TF and TC databases—PlnTFDB (Pérez-Rodríguez et al. 2010), PlantTFDB 3.0 (Jin et al. 2014), Grassius (Yilmaz et al. 2009), ProFITS (<http://bioinfo.cau.edu.cn/ProFITS/index.php>; last accessed August 30, 2017), and iTAK (<http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi>; last accessed August 30, 2017). Approximately 98% of the pineapple TFs (1,371 out of 1,398) and 81% of the pineapple TCs (65 out of 80) had a good match (query coverage per subject >45%, e-value < 1e-5) in the plant TF and TC databases (supplementary tables S3 and S5, Supplementary Material online). In most cases, the pineapple TFs (96%, 1,340 out of 1,398) and TCs (65%, 52 out of 80) belonged to the same family or superfamily as those of their best homologs in the public TF and TC databases (table 1, supplementary tables S3 and S5, Supplementary Material online). Since TF and TC family

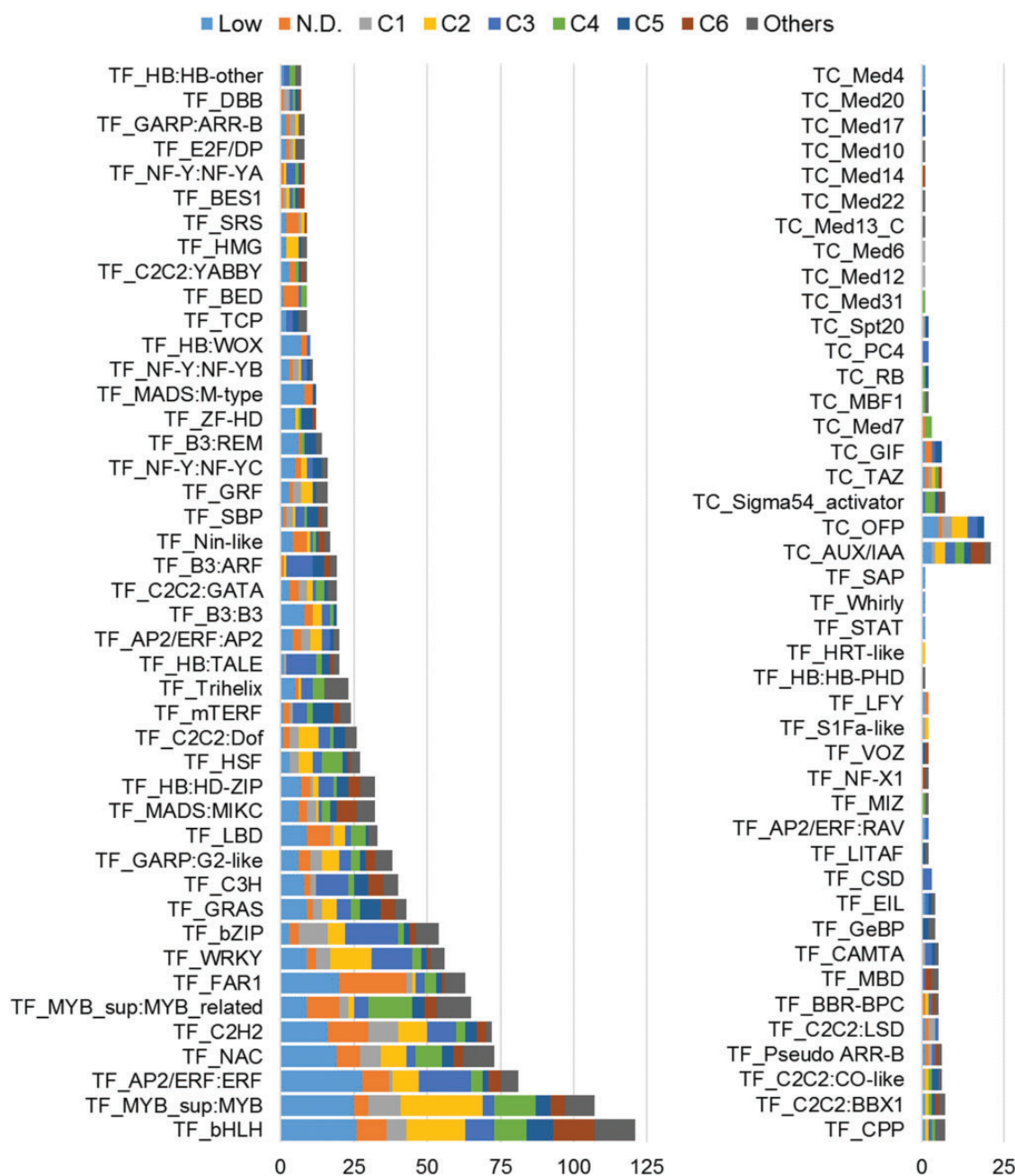


Fig. 1.—The abundance and tissue expression patterns of pineapple TFs and TCs. The bar graph shows the total number of members (*X*-axis) identified for each TF/TC family (listed along *Y*-axis) in pineapple. The TFs and TCs in each family are further grouped and color coded based on their expression in the four tissues—flower, leaf, root, and fruit. N.D., not detected (FPKM 0 in all tissues); Low (FPKM <10 in all tissues), groups C1 to C6 are the largest clusters obtained by clustering the normalized FPKM values from four tissues (see fig. 2 for the major clusters).

names adopted by different TF and TC databases may differ, synonymous TF and TC family names were taken into consideration when applicable, for example, TF families B3 (or ABI3VP1), NF-YC (or CCAAT), Nin-like (or RWP-RK), LBD (or LOB), GIF (or SSXT), and TC families Med31 (or SOH1), and PC4 (or coactivator p15).

Eighty-six pineapple TFs and TCs had either no homolog or had a homolog with a different classification in public TF and

TC databases. We therefore compared our annotations of these 86 TFs and TCs to those predicted independently by iTAK and PlantTFcat webservers (supplementary table S3, Supplementary Material online). Our classification matched iTAK and/or PlantTFcat predictions for 55 TF and TC proteins but differed for five proteins due to the presence of alternate DNA binding or auxiliary domain in these proteins that resulted in ambiguous classification. iTAK and PlantTFcat

Table 1
Pineapple TFs and TCs and Their Homologs in Public Databases

	TF	TC
Total	1398	80
With any homolog in TF databases	1371	65
With same family homolog in TF databases	1340	52
With different family homolog in TF databases	31	13
No Homolog in TF databases	27	15

NOTE.—Detailed information on annotation of TFs and TCs can be found in supplementary tables S3 and S5, Supplementary Material online.

did not predict TF and TC for the 26 remaining proteins, 17 of which belonged to 10TC families (Med10, Med12, Med13_C, Med14, Med17, Med20, Med22, Med4, Sigma54_activator, and Spt20) that are not incorporated in these databases. Therefore, the vast majority of our classification of pineapple TFs and TCs is consistent with their homologs from other plant genomes.

Expression Pattern of Pineapple TFs and TCs among Four Different Tissues

Tissue-specific expression of TFs is important to establish cell identity and function. To study the expression pattern of pineapple TFs and TCs in different tissues, we obtained normalized FPKM values of pineapple genes based on RNAseq libraries prepared from four pineapple tissues: flower, leaf, and root from cultivar F153, and fruit from cultivar MD-2. The expression patterns of TFs and TCs among the four tissues were interrogated by clustering. A total of 468 genes, including 164 with no detectable expression (normalized FPKM “0” in all tissues) and 304 with insufficient read depth (maximum FPKM less than 10 in any tissue), were filtered before clustering to reduce noise in the final clusters. We chose CAST (Cluster Affinity Search Technique) algorithm to explore the underlying structure of pineapple TF/TC expression in the four tissues because this method does not require a predefined number of clusters, and can handle outliers efficiently. We clustered 1,010 pineapple TF/TC genes into 19 clusters using a strict threshold parameter of 0.9. Each of the 19 clusters was hierarchically clustered and shown as heatmaps (figs. 2 and 3). The gene expression graphs and the mean expression pattern (centroids) for all 19 clusters are given in figures 2 and 3 as well. Most TF/TC genes (82%) were grouped into one of the six largest clusters—C1 (11%), C2 (17%), C3 (19%), C4 (13%), C5 (12%), C6 (9%). These six clusters can be subdivided into three pairs with inversely correlated centroids, that is, C1 and C5 (Pearson’s $r = -0.94$), C2 and C6 (Pearson’s $r = -0.95$), C3 and C4 (Pearson’s $r = -0.93$). The opposing expression patterns are 1) low in fruit (C1) or high in fruit (C5) relative to the other three tissues; 2) high in root (C2) or low in root (C6) relative to the other three tissues; and 3) low in flower and leaf (C3) or high in flower and leaf (C4) relative to other two tissues. Therefore, in addition to a large number

of TF/TCs with fruit-specific and root-specific regulation, a noticeably large number TF/TCs exhibited coregulation pairwise in leaf and flower, and in root and fruit tissues.

We then used Fisher’s exact test to assess if any TF or TC family was preferentially enriched in a specific cluster (supplementary fig. S1, Supplementary Material online). We found that 26 TF/TC families were enriched in one or two of the six largest clusters. These included three (C2H2, bZIP, and C2C2: LSD) enriched in C1 (higher mean expression in flower, leaf, and root), four (WRKY, MYB_sup: MYB, C2C2: Dof, and HMG) in C2 (higher expression in roots), eight (HB: TALE, bZIP, B3: ARF, CSD, WRKY, C3H, AP2/ERF: ERF, PC4) in C3 (higher expression in root and fruit), five (TFs MYB_sup: MYB_related, HSF, Sigma54_activator, and Med7) in C4 (higher expression in flower and leaf), five (mTERF, SBP, ZF-HD, GeBP, and B3: REM) in C5 (higher expression in fruit), and four (TFs bHLH, MADS: MIKC, MBD and AUX/IAA) in C6 (higher expression in flower, leaf, and fruit). In addition, 11 TF families were enriched in the group with no detectable expression (FPKM 0) or little expression (FPKM < 10). These included FAR1 (enriched in both groups), C2H2, LBD, Nin-like, BED, and SRS (enriched in the FPKM 0 group), and AP2/ERF: ERF, B3: REM, B3: B3, MADS: M-type, HB: WOX (enriched in the FPKM < 10 group).

Diurnal Expression Pattern of Pineapple TFs and TCs

Most clock components and clock-regulated genes exhibit diurnal periodicity in gene expression and this rhythmicity generates the circadian rhythms in plant physiology. We used time-course gene expression data of pineapple photosynthetic green tip and nonphotosynthetic white base leaf tissue over a 24-h period to identify TFs and TCs whose expression patterns fit a predefined model of cycling genes using Haystack (Mockler et al. 2007). We tailored the models to fit our collection time points based on the models defined by Endo et al. (2014). Detailed information about the derived models is given in supplementary figure S2 and table S6, Supplementary Material online. We empirically defined cycling TFs and TCs as those with a strong correlation ($r > 0.7$) to a predefined model of cycling genes, a fold change > 2 , P value > 0.05 and an amplitude > 10 . Based on this rule, ~42% of TFs (589 out of 1,392) and 45% (36 out of 80) of TCs were found to be cycling in either one or both of green tip and white base leaf tissues. The 625 cycling TFs/TCs included 177 (28%) cycling in the white leaf base only, 247 (40%) cycling in the green tissue only, and 201 (32%) cycling in both tissues (table 2 and supplementary table S3, Supplementary Material online). Diurnal expression profiles of cycling TF/TCs with a diel peak expression in both the white base and green tip, as well as in the white base and green tip only are shown in figures 4 and 5.

The distribution of cycling TFs and TCs in each TF and TC family is given in supplementary figure S3, Supplementary

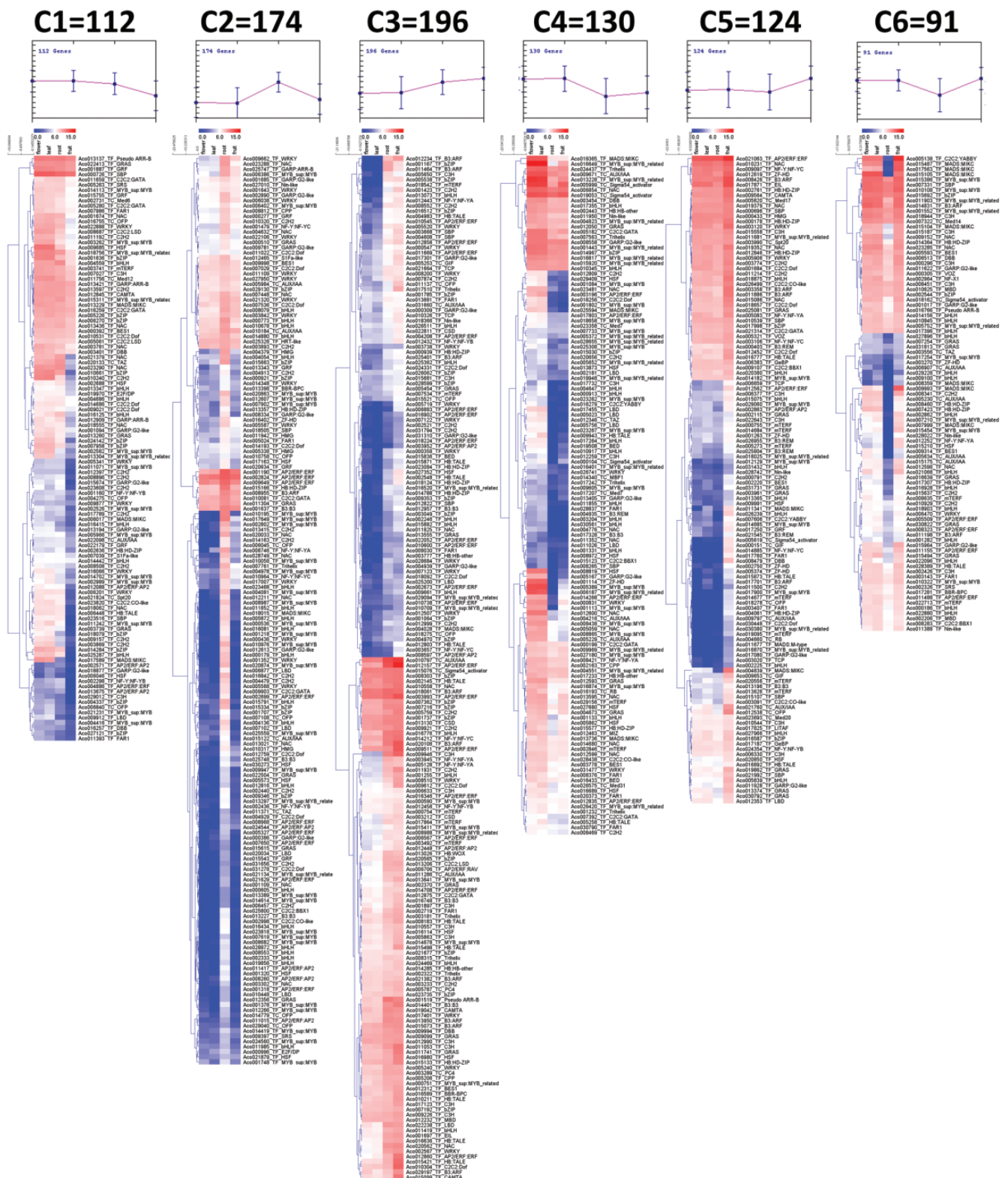


FIG. 2.—Heatmaps of the six largest clusters obtained by clustering the expression data of pineapple TFs and TCs in four tissues. The number of TF/TC genes in each cluster is listed at the top of each cluster. A small graph on the top of each graph shows the mean expression pattern (centroid) of each cluster.

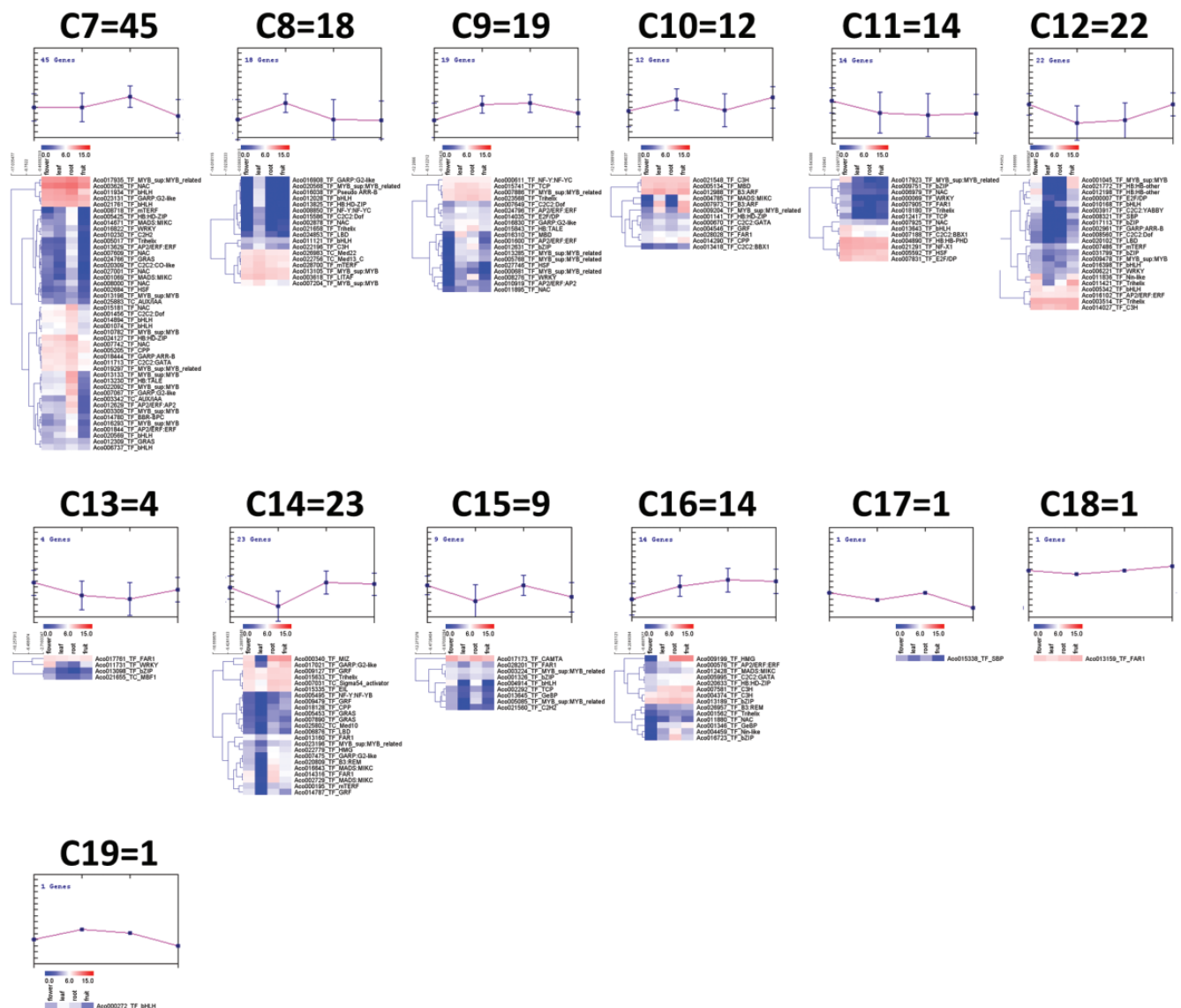


FIG. 3.—Heatmaps of the small clusters obtained by clustering the expression data of pineapple TFs and TCs in four tissues. The number of TF/TC genes in each cluster is listed at the top of each cluster. A small graph on the top of each graph shows the mean expression pattern (centroid) of each cluster.

Material online. Among the TF families with ten or more members, the families with a high fraction (50–68%) of cycling genes included C2C2: GATA, B3: ARF, bZIP, mTERF, SBP, WRKY, HB: HD-ZIP, and HB: TALE. Although TF families bHLH, NAC, AP2/ERF: ERF, MYB_{sup}: MYB, MYB_{sup}: MYB_{related}, and C2H2 contained a high number of cycling TFs (ranging from 26 to 56), the fraction of their cycling TFs was low, ranging from 29% to 46%.

Phase Correlation of Cycling TFs and TCs with Diurnal Expression Peaks in Both Nonphotosynthetic and Photosynthetic Leaf Tissues

In general, all the cells in plants contain an autonomous circadian clock (McClung 2006). Unlike animal that has a

“master clock” in the brain coordinating all the body clocks, plant circadian clock had been thought to be uncoupled but recent studies have deciphered previously unknown interactions (Endo et al. 2014; Takahashi et al. 2015). The 201 TF/TCs that exhibited diel peak expression in both nonphotosynthetic and photosynthetic leaf tissues offered us an opportunity to investigate if there is a phase correlation of the TF/TCs cycling in the two tissues.

To ascertain whether the cycling TF/TCs with diurnal expression peaks in both green leaf tip and white leaf base have synchronous expression in the two tissues, we estimated the correlation (Pearson’s *r*) at 2-h interval lags between the green leaf tip and white leaf base for the time series expression data of the 201 cycling TF/TCs that showed diurnal expression peaks in both tissues. A total of 105 TF/TCs had strong

Table 2

Diurnal Expression Pattern of Pineapple TFs and TCs

	TC	TF
Cycling in both green tip and white base leaf tissue	15	186
Cycling in white base leaf tissue only	11	166
Cycling in green tip leaf tissue only	10	237
Non-Cycling	44	803
Total number	80	1392

NOTE.—Please see [supplementary table S3, Supplementary Material](#) online, for details of the TF/TCs summarized above.

correlation (>0.7) between the green leaf tip and white leaf base at various lags. The highest correlation value for most of these genes was obtained at lag 0 (31 genes) or lag 2 (29 genes), suggesting that these TF/TCs have synchronized expression between the two tissues (fig. 6). Since no peak was observed at lag 22 (-2 h relative to ZT0), there was a slight phase delay of up to ~ 2 h in the white leaf base (29 genes) relative to the green tissue for nearly half of the TF/TCs that showed synchronous expression between the two tissues (fig. 6). In addition, eight genes, two with the highest correlation at lag 4 and six with the highest correlation at lag 22, were also classified as synchronously expressed because their correlation values at lag 0 or lag 2 were high ($r > 0.7$). Therefore, we grouped these 201 cycling TF/TCs into three groups based on their correlation of expression pattern between the green leaf tip and white leaf base. Group I contained 68 genes with strong correlation (>0.7) at lag 0 and lag 2 (fig. 4, top panel), group II contained 37 genes with strong correlation (>0.7) at lags 4–22 (fig. 4, middle panel), and group III contains 96 genes with weak correlation (fig. 4, bottom panel).

Diurnal cycling genes have peak transcript or protein abundance at specific time of the day. A plot of the number of TF/TCs against either the predicted phase (model-based determination of peak expression time) or the time of maximum expression level as determined by FPKM showed that peak expression was distributed throughout the 24-h period. More cycling TF/TCs had expression patterns that peaked during the day than those with peak expression at night (supplementary fig. S4, Supplementary Material online). A large number of TF/TCs showed peak expression between 7 AM and 3 PM and very few had peaks between 5 and 11 PM. This estimate may be biased as there was an extra sample collected during the day (8 AM–6 PM) compared with the ones collected at night (8 PM–6 AM).

The expression pattern of most cycling TF/TCs in pineapple match the “spike” model, which was most successful in identifying cycling TF/TC genes in pineapple, accounting for 45% of the cycling genes, followed by the models *asyMt2* (16%), *mt* (11%), and *asyMt1* (7%) (supplementary fig. S5, Supplementary Material online). The “cosine” and “spike” models were found to be the most successful among the six models, “asymmetric,” “rigid,” “spike,” “cosine,” “sine,” and “box-like,” in identifying true oscillating transcripts in

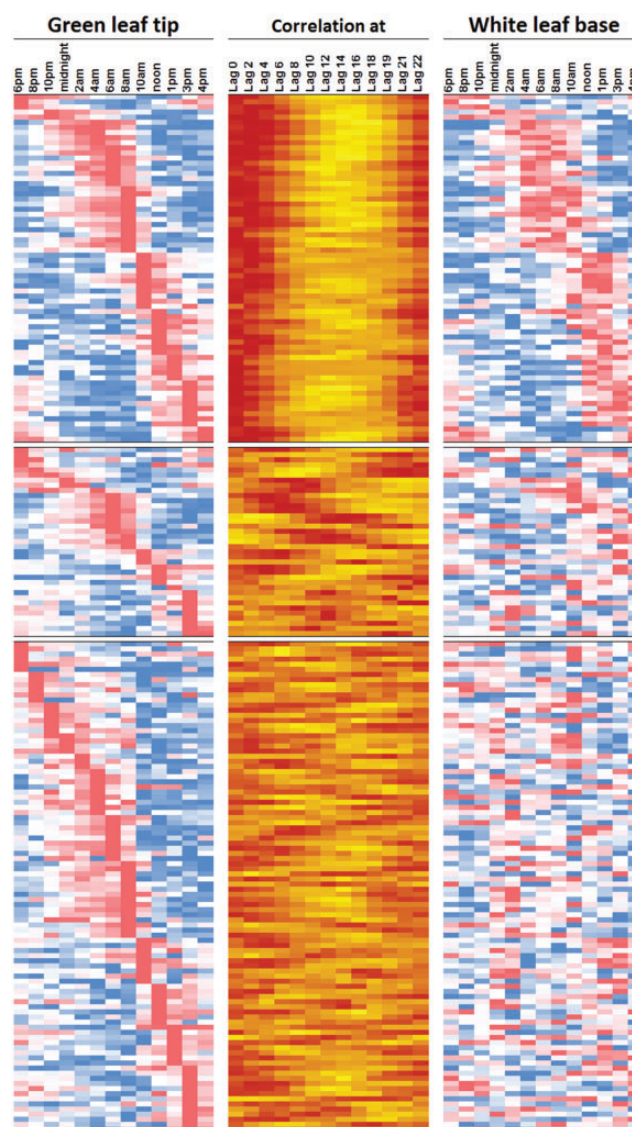


FIG. 4.—Diurnal expression profiles of cycling TFs and TCs with a diel peak expression in both photosynthetic and nonphotosynthetic leaf tissues. The expression levels of each gene in a given tissue are color coded in red-white-blue color scale, where red represents the highest expression, blue represents the lowest expression, and white represents an intermediate expression. The middle vertical panel shows the correlation coefficient at 2-h lags between the time course expression values in the green and white leaf tissues. The top horizontal panel includes genes exhibiting strong correlation (>0.7) at lag 0 and lag 2, middle horizontal panel includes genes exhibiting strong correlation (>0.7) at lags 4–22, and the bottom horizontal panel includes genes exhibiting weak or no correlation in the expression between the two tissues.

an artificial time series (Walter et al. 2014). Our results match with the results of Walter et al (2014) for the spike model but not for the cos model. This difference could be caused by an artificial time series used in their study, additional models used in our study, or the expression pattern of cycling TF/TCs.

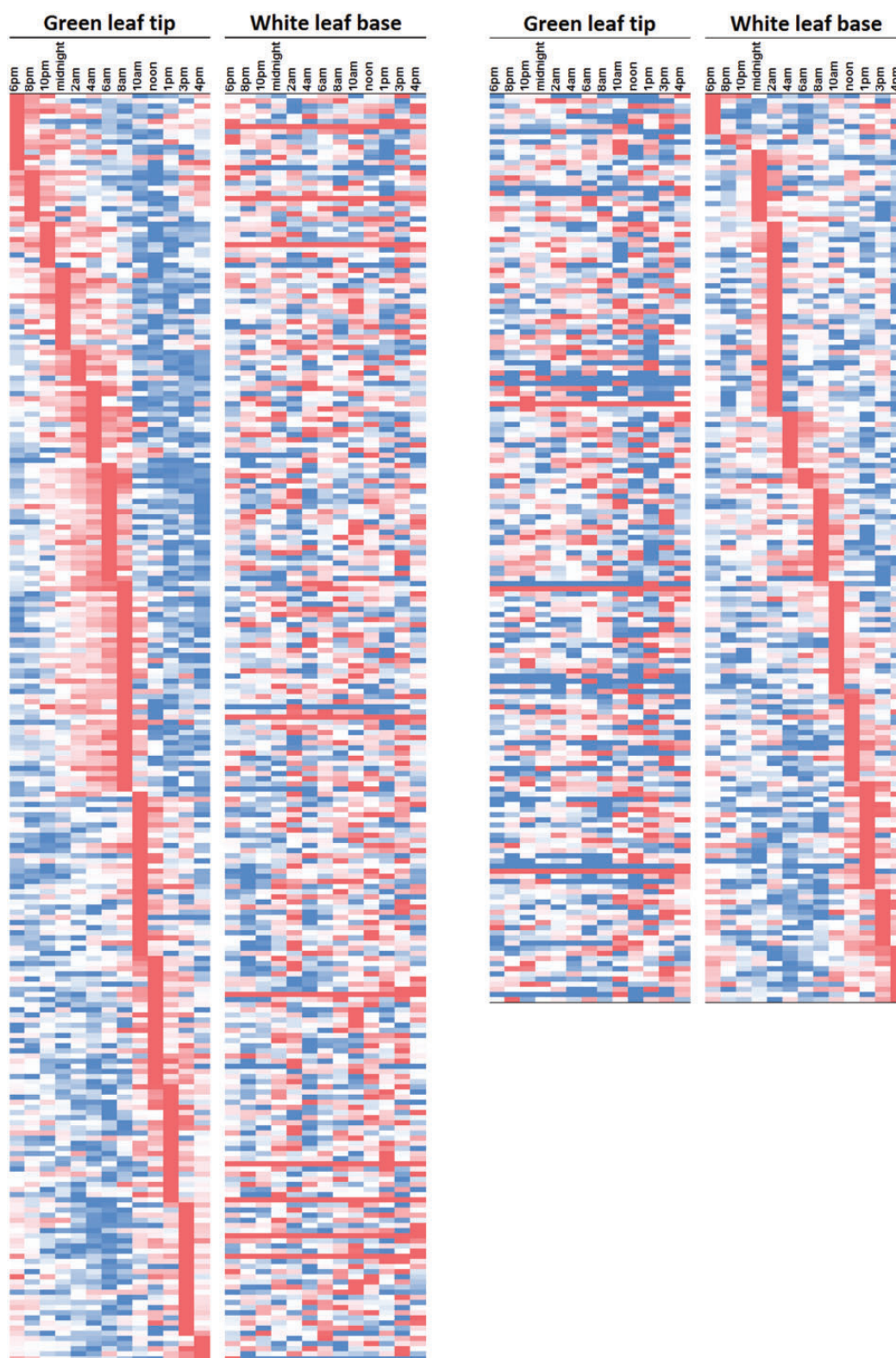


Fig. 5.—Diurnal expression of cycling TFs and TCs that have diel peak expression in white leaf base or green leaf tip only. (A) The expression of TF/TC genes having a diel peak only in the green leaf tip is shown with respect to their expression in the white leaf base (left panel). (B) The expression of TF/TC genes having a diel peak only in the white leaf base is shown with respect to their expression in the green leaf tip (right panel). The normalized expression values are color coded in red-white-blue color scale, where red represents the highest expression, blue represents the lowest expression, and white represents an intermediate expression.

We identified pineapple orthologs of known circadian clock related TF/TC genes in *Arabidopsis* (Cheng and Wang 2005; Nakamichi 2011; Hsu and Harmer 2014; Greenham and McClung 2015) and checked their distribution among the three groups of cycling TF/TCs described above. Among the 17 identified orthologs, 15 showed cycling expression in pineapple leaf tissues. Eleven out of the 15 orthologs (73% of the cycling circadian orthologs) belonged to the group I – genes cycling synchronously in both green leaf tip and white leaf base (16%, 11/68), whereas four of

them (27% of the cycling circadian orthologs) belonged to the group of the genes cycling in green leaf tip only (1.6%, 4/247) (table 3). All the 11 circadian-related gene homologs in group I had high amplitude and appeared among the top 22 high amplitude genes. The enrichment of homologs of known core circadian oscillators and circadian regulated genes in the set of high amplitude synchronously expressed genes implies that this gene set can be a promising candidate pool for identification of circadian associated genes in pineapple.

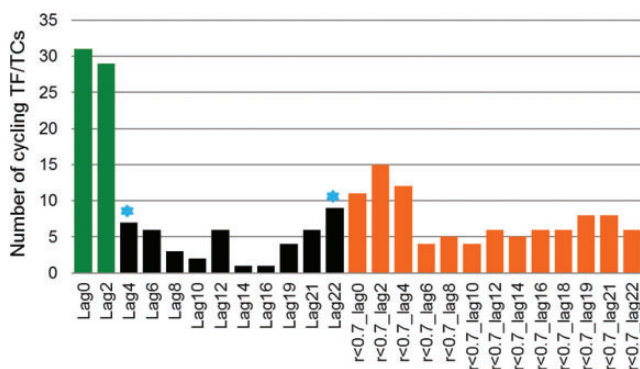


Fig. 6.—Lagged correlation of time course expression of cycling TFs and TCs between the green leaf tip and white leaf base. The number of TFs and TCs (Y-axis) showing high correlation (>0.7, green and black columns) or low correlation (<0.7, orange columns) of their time course expression between the green leaf tip and white leaf base at time lags 0–22 (X-axis).

Besides the circadian genes listed in table 3, the group 1 included five heat shock factors, Aco005999.1 (homolog of AT-CLPB1/AT-CLPB2), Aco018162.1 (homolog of AT-CLPB3/AT-CLPB4), Aco005592.1 (homolog of AT-HSFA4A), Aco008819.1 (homolog of AT-HSFA6B), and Aco027680.1 (AT-HSFB2A/AT-HSFB2B). These five heat shock factor genes had peak expression between 10 and 12 AM (fig. 7). Heat shock factors are activated in response to temperature stress and confer thermotolerance. In plants, the expression of heat shock factors, such as AT-HSFB2b, is required for accurate circadian rhythms under temperature and/or salt stress (Kolmos et al. 2014). In mammals, Hsf1 plays a key role in resetting and compensation of the circadian clock to temperature (Reinke et al. 2008; Buhr et al. 2010). Therefore, it is conceivable that cycling heat shock factors in pineapple might play a role in modulating circadian clock in response to temperature changes in addition to providing thermotolerance. Aco002824.1 and Aco016346.1, pineapple homologs of the DREB subfamily A-6 genes, and Aco022517.1 and

Table 3
Pineapple Homologs of *Arabidopsis* Circadian Clock-Related Genes

Pineapple Homologs	<i>Arabidopsis</i> Circadian Clock-Associated Genes
Aco001684 (Group I)	AT5G39660 - CYCLING DOF FACTOR 2 (CDF2)
Aco018657 (Group I)	
<u>Aco009612</u> (Group I)	AT3G47500 - CYCLING DOF FACTOR 3 (CDF3)
	AT5G62430 - CYCLING DOF FACTOR 1 (CDF1)
Aco013228 (Group I)	AT2G46830 CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)
<u>Aco016649</u> (Group I)	AT1G01060 LATE ELONGATED HYPOCOTYL (LHY)
Aco016830 (Group I)	AT5G59570 BROTHER OF LUX ARRHYTHMO (BOA)
	AT3G46640 PHYTOCLOCK 1 (PCL1) or LUX
Aco026499 (Group I)	AT3G07650 CONSTANS-LIKE 9 (COL9)
Aco003091 (Group I)	AT5G48250 best match to CONSTANS-like 9
<u>Aco013137</u> (Group I)	AT5G24470 PSEUDO-RESPONSE REGULATOR 5 (PRR5)
Aco016766 (Green Leaf)	
Aco016928 (Green Leaf)	AT5G58010 (AtLRL3)
	AT4G30980 (bHLH69 or AtLRL2)
	AT2G24260 (AtLRL1)
	AT5G61380 TIMING OF CAB EXPRESSION 1 (TOC1)
Aco016038 (Not Cycling)	
Aco021178 (Not Cycling)	
<u>Aco013643</u> (Green Leaf)	AT1G09530 PHYTOCHROME INTERACTING FACTOR 3 (PIF3)
<u>Aco011903</u> (Group I)	AT5G17300 Putative RVE1 like TF (RVE1)
Aco009204 (Group I)	AT5G37260 REVEILLE 2 (RVE2)
Aco011012 (Green Leaf)	AT5G02810.1 PSEUDO-RESPONSE REGULATOR 7 (PRR7)

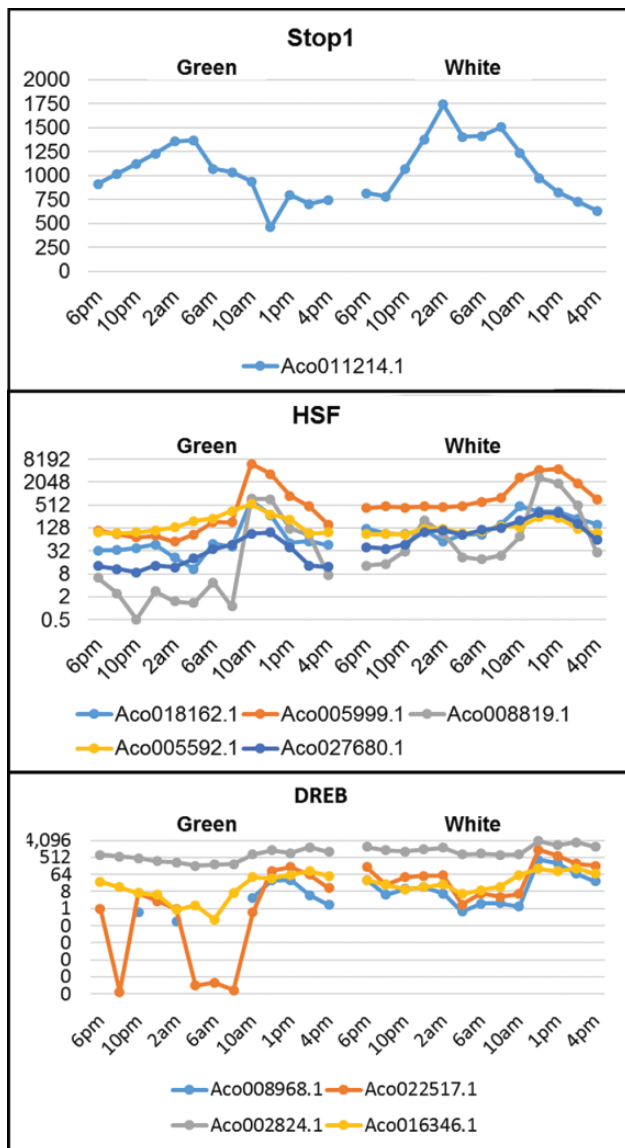


FIG. 7.—Diurnal expression patterns of the pineapple homologous genes of a STOP1 (Sensitive TO Proton rhizotoxicity), five heat shock factors, and four DREB subfamilies A-6 and A-1 genes in green leaf tip and white leaf base. The X-axis represents time points starting from ZT0 (6am) to ZT22 (4am) and Y-axis represents gene expression level in FPKM value.

Aco008968.1, pineapple homolog of DREB subfamily A-1, were also included in the group I. RAP 2.4, an *Arabidopsis* DREB subfamily A-6 gene, regulates multiple developmental processes and drought stress tolerance by mediating the cross-talk between the light and ethylene signaling pathways (Lin et al. 2008). Therefore, Aco002824.1 and Aco016346.1 may play roles in coordinate internal biological processes with changes in light and abiotic stress signals. Group I also contains Aco011214.1, a homolog of AT-STOP1 that responds to acidic pH and activates a malate efflux transporter (Iuchi et al. 2007). Aco011214.1, the pineapple

STOP1 homolog, had the lowest expression at noon and the highest expression at 2 to 4 AM (fig. 7), which coincided with diurnal oscillation of malate concentration in pineapple leaf (Kenyon et al. 1985; Rainha et al. 2016), suggesting a potentially important role of Aco011214.1 in regulating CAM photosynthesis.

Tissue-Specific Expression of Paralogs of Three Cycling TFs in Pineapple

We identified two pineapple homologs each of the central circadian oscillator At-CCA1/LHY (Mizoguchi et al. 2002), circadian clock regulated gene At-COL9 (Cheng and Wang 2005) and At-CDF2 (Imaizumi et al. 2005; Fornara et al. 2009; Greenham and McClung 2015) in group I (table 3). Since group I genes have the same phase between the photosynthetic and nonphotosynthetic leaf tissues, we further examined these three pairs of genes in detail to understand their origin, evolution, and tissue-specific expression patterns.

The two LHY homologs of pineapple, the 698aa Ac_LHY-1 (Aco016649) and the 696aa Ac_LHY-2 (Aco013228), showed 39% sequence identity over their full-length alignment (supplementary fig. S6A, Supplementary Material online). Detailed phylogenetic analysis with CCA1/LHY homologs from other dicot and monocot genomes is described in the next section “Phylogenetic relationships of core circadian clock genes between pineapple and other plant species.” Interestingly, although both Ac_LHY-1 and Ac_LHY-2 had the same phase (expression peaking at dawn) between the two leaf tissues, differential expression patterns were observed between the two LHY homologs and between the two leaf tissues. Ac_LHY-2 exhibited a high amplitude rhythm in both green leaf tip and white leaf base, whereas Ac_LHY-1 showed a low-level of expression in the white leaf base and a much higher level of expression in the green leaf tip (fig. 8).

The two COL9 homologs of pineapple, the 402aa Ac_COL9-1 (Aco003091.1) and 411aa Ac_COL9-2 (Aco026499.1), are 54% identical over their full-length alignment (supplementary fig. S6B, Supplementary Material online). Phylogenetic analysis revealed that both Ac_COL9-1 and Ac_COL9-2 were present in the common ancestor of commelinids and grasses (Poaceae) but Ac_COL9-2 homolog had been lost in the grass family after grasses diverged from commelinids (supplementary fig. S7, Supplementary Material online). Ac_COL9-2 (Aco026499.1) showed a much higher level of expression than Ac_COL9-1 (Aco003091.1) in both green leaf tip and the white leaf base, and both Ac_COL9-1 and Ac_COL9-2 exhibited increased expression in the green leaf tip compared with the white leaf base (fig. 8).

The two At-CDF2 homologs of pineapple, 478aa Ac_CDF-1 (Aco001684.1) and 497aa Ac_CDF-2 (Aco018657.1), share 58% identity over their full-length alignment (supplementary fig. S6C, Supplementary Material online). Phylogenetic analysis with CDF homologs from other monocots and dicots

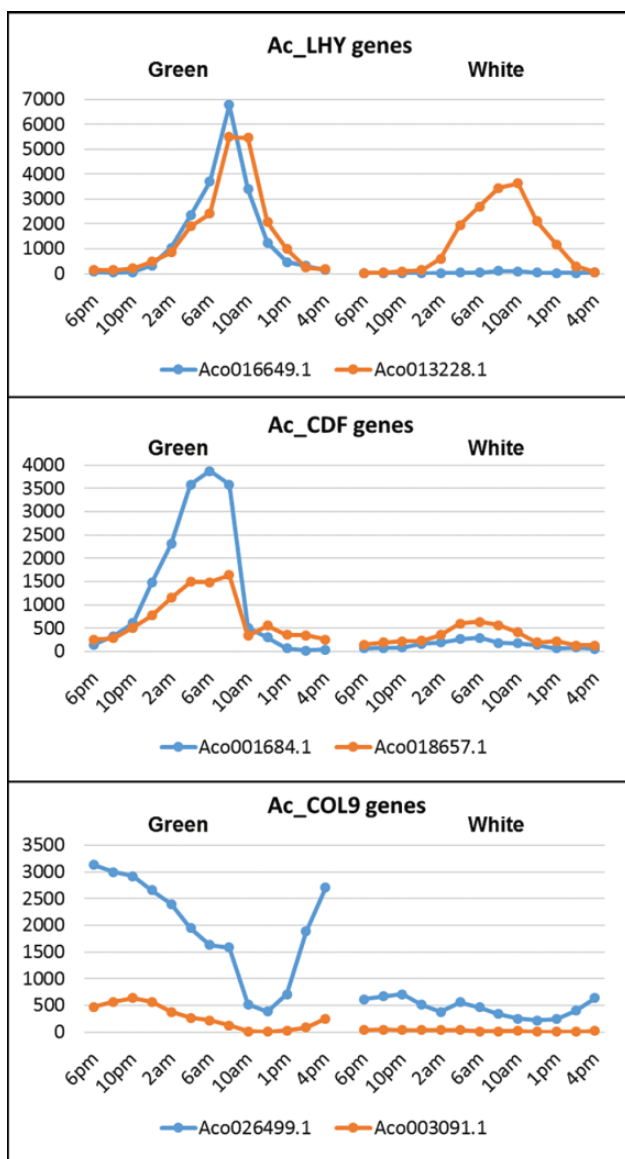


FIG. 8.—Differential gene expression patterns of the three pairs of cycling genes in green leaf tip and white leaf base. The diurnal expression pattern of pineapple *Ac_LHY* (top), *Ac_CDF* (middle), and *Ac_COL9* (bottom) genes are shown where X-axis represents time points starting from ZT0 (6am) to ZT22 (4am) and Y-axis represents normalized gene expression in FPKM units.

showed that *Ac_CDF-1* and *Ac_CDF-2* shared similar distance with other CDF homologs from the grass family and they might have originated from a gene duplication event after the pineapple lineage split from the lineage leading to the grass family (supplementary fig. S7, Supplementary Material online). Both *Ac_CDF-1* and *Ac_CDF-2* exhibited increased expression in the green leaf tip compared with the white leaf base. *Ac_CDF-1* (Aco001684.1) showed a significantly increased level of expression in green leaf tip compared with the white leaf base (fig. 8).

Phylogenetic Relationships of Core Circadian Clock Genes between Pineapple and Other Plant Species

Orthologs tend to have similar function and therefore their identification is important for gene annotation and gene function prediction. The reciprocal best blast hit based methods are commonly employed for genome-wide prediction of orthologs. However, these methods tend to be less accurate than the phylogeny-based approaches (Fulton et al. 2006). Moreover, the reciprocal best hit method generates a high rate of false negatives in the duplication-rich genomes as it detects 1 to 1 orthologs, and misses as much as 60% of the orthologous relations that originated by gene duplication events (Dalquen and Dessimoz 2013). Therefore, we used the phylogenetic approach in addition to the methods based on Reciprocal Best BLAST hit and synteny information including Proteinortho-Poff (Lechner et al. 2014) and SynMap (<https://genomeevolution.org/coge/SynMap.pl>; last accessed August 30, 2017) to identify orthologous relationships of major TFs known to function primarily within the circadian clock between pineapple and other plant genomes. For a better resolution of evolutionary relationships, we included homologous proteins from 64 sequenced plant genomes that were available at Phytozome v11 (supplementary table S7, Supplementary Material online) to construct phylogenetic trees using maximum likelihood method (Guindon et al. 2010). As expected, the phylogenetic approach was better at resolving the evolutionary relationships. Our detailed ortholog analysis of core circadian genes are described and summarized below.

The PRR Family

In *Arabidopsis*, PRR1, PRR5, PRR9, PRR7, and PRR3 are directly involved in circadian clock and contain a Pseudo-Receiver and CCT domains (Makino et al. 2000; Strayer et al. 2000). Maximum likelihood phylogenetic trees constructed from the conserved N-terminal domain of PRR1/TOC1, PRR9, PRR5, PRR3, and PRR7 homologs revealed three distinct clades, the PRR1/TOC1 clade, the PRR3/PRR7 clade, and the PRR5/PRR9 clade (supplementary fig. S8, Supplementary Material online). This grouping is similar to those obtained by Takata et al. (2010). To discern the relationship of pineapple PRR homologs to those of *Arabidopsis* and rice, we constructed separate phylogenetic trees for each of the three clades.

PRR1/TOC1 Clade

Only single copy of PRR1 gene from *Arabidopsis* (AT5G61380) and rice (LOC_Os02g40510.1) genomes was grouped in the PRR1/TOC1 clade in our phylogenetic analysis. Phylogenetic clustering of PRR1 homologs showed that the single copy of PRR1 gene in *Arabidopsis* was orthologous to all the PRR1 homologs in monocots (supplementary fig. S9, Supplementary Material online). We identified two copies of

PRR1 paralogous genes in pineapple genome, Ac_PRR1-1 (Aco021178) and Ac_PRR1-2 (Aco016038). These two PRR1 paralogs share 95% amino acid sequence identity (bit-score: 881). The rice PRR1 homolog, Os_PRR1, is approximately equidistant to the two pineapple PRR1 paralogs, sharing 57% amino acid sequence identity with Ac_PRR1-1 (bit-score: 540) and 56% identity with Ac_PRR1-2 (bit-score: 504). Our result suggested that the two pineapple PRR1 paralogs were likely derived from a recent duplication event after pineapple and grasses diverged from a common ancestor. Therefore, both of them were co-orthologous to the solo PRR1 copy in rice (LOC_Os02g40510.1) and *Arabidopsis* (AT5G61380) (fig. 9a and supplementary fig. S9, Supplementary Material online).

PRR3/PRR7 Clade

Phylogenetic analysis grouped the PRR3/PRR7 homologs of dicots into three clades, whereas all the PRR3/PRR7 homologs of monocots into a single clade, indicating that the PRR3/PRR7 homologs of dicots had experienced a triplication event, likely the gamma triplication event, after dicots diverged with monocots (fig. 9a and supplementary fig. S10, Supplementary Material online). Interestingly, differential retention after the triplication event was observed in PRR3/PRR7 homologs of dicots. All three copies of the PRR3/PRR7 homolog have been retained in *Theobroma cacao* and *Gossypium raimondii*. *G. raimondii* gained an additional copy through a lineage-specific gene duplication event. Only two of the three copies have been retained in *Arabidopsis*, *Carica papaya*, and *Amaranthus hypochondriacus*, and the three species showed different retention patterns. We identified two copies of PRR3/PRR7 homologs (Aco011012 and Aco001519) in pineapple genome. Since PRR3 and PRR7 were initially named in *Arabidopsis* and both of them are coorthologous to the monocot PRR3/PRR7 homologs, we named the two pineapple homologs “Ac_PRR3/7-1” (Aco011012) and “Ac_PRR3/7-2” (Aco001519). Only a single copy of PRR7/PRR3 homolog was identified in the basal angiosperm species *Amborella trichopoda*. Therefore, the two pineapple copies were likely originated from the sigma duplication event after the divergence of pineapple and grass ancestor from *Elaeis guineensis* and *Phoenix dactylis*. Furthermore, Ac_PRR3/7-1 (Aco011012) and Ac_PRR3/7-2 (Aco001519) share 55% amino acid sequence identity (bit-score: 657) and both have the best match to the same *E. guineensis* and *P. dactylis* orthologs. Three copies of PRR3/PRR7 homologs were identified in *Musa acuminata*. Our phylogenetic analysis showed these three copies of PRR3/PRR7 homologs were derived from *Musa* lineage-specific genome duplication events. All the PRR3/PRR7 homologs of grasses were clustered into two groups (supplementary fig. S10, Supplementary Material online), suggesting PRR3/PRR7 homologs of grasses had undergone a duplication event, likely the rho whole genome duplication event in the common

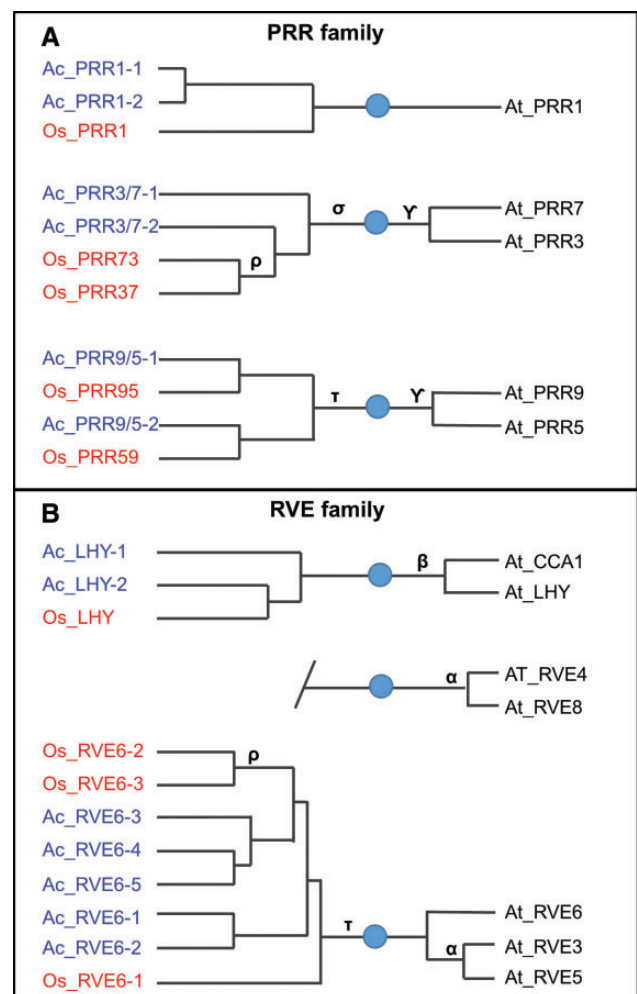


Fig. 9.—Evolutionary relationships of PRR and RVE genes. Schematic shows phylogenetic relationships between the PRR family members (A) and RVE family (B) of *Arabidopsis*, rice, and pineapple. The blue circle denotes common ancestor of dicot and monocots. The whole genome duplication events gamma, sigma, tau, and rho are shown as applicable and the lengths of the lines connecting taxa do not indicate evolutionary distance. *Arabidopsis* RVE1/RVE2/RVE7 genes and their homologs are not included here as these do not form the core circadian clock.

ancestor of grasses, after grasses diverged with pineapple (fig. 9a). Two PRR3/PRR7 homologs, Os_PRR73 (LOC_Os03g17570) and Os_PRR37 (LOC_Os07g49460), were identified in rice. Os_PRR37 and Os_PRR73 share ~66% amino acid sequence identity over 90% of their lengths and both share a higher percent of similarity to Ac_PRR3/7-2 than to Ac_PRR3/7-1, indicating the Ac_PRR3/7-1 counterpart have been lost in grasses during fractionation.

PRR5/PRR9 Clade

Phylogenetic analysis grouped the PRR5/PRR9 homologs of dicots into three clusters and the ones of monocots into

two clusters (supplementary fig. S11, Supplementary Material online). Only a single copy of PRR5/PRR9 was identified in the basal angiosperm species *A. trichopoda*. All together suggested that the PRR5/PRR9 homologs had undergone duplication events after dicots and monocots separated from a common ancestor. PRR5/PRR9 homologs of dicots likely originated during the gamma triplication event, and the ones of monocots likely originated during the Tau whole genome duplication (fig. 9a). Differential retention after the triplication event was observed in PRR5/PRR9 homologs of dicots. Only two of the three copies have been retained in *Arabidopsis* and *C. papaya* genomes, and the two species showed different retention patterns. After the Tau whole genome duplication, the two copies of PRR5/PRR9 homologs have been retained in both pineapple and rice. The two copies in pineapple were named Ac_PRR9/5-1 (Aco013137.1) and Ac_PRR9/5-2 (Aco016766.1). Ac_PRR9/5-1 was grouped into the same clade with its rice ortholog Os_PRR95 (LOC_Os09g36220.1) and they share 47% amino acid sequence identity (bit-score: 485). Ac_PRR9/5-2 was grouped into the same clade with its rice ortholog Os_PRR59 (LOC_Os11g05930.1) and they share 50% amino acid sequence identity (bit-score: 567).

The RVE Family

In *Arabidopsis*, several RVE gene family members, including LHY, CCA1, RVE6, RVE5, RVE3, RVE4, and RVE8, have been implicated to play an important role in circadian clock (Schaffer et al. 1998; Hsu and Harmer 2014). This gene family contains a single Myb domain. We constructed phylogenetic trees using the conserved Myb domain of LHY, CCA1, RVE6, RVE5, RVE3, RVE4, and RVE8 homologs in plants. Our result showed that these proteins formed two distinct clades, one containing CCA1 and LHY and the other containing RVE6, RVE5, RVE3, RVE4, and RVE8 (supplementary fig. S12, Supplementary Material online). We constructed separate phylogenetic trees for the two clades to refine their evolutionary relationships.

CCA1/LHY Clade

Phylogenetic analysis grouped CCA1/LHY homologs into two major clades, one containing all the homologs from dicots and the other one containing all the homologs from monocots (supplementary fig. S13, Supplementary Material online). Our result showed that CCA1 (AT2G46830) and LHY (AT1G01060) are paralogs that originated in a Brassicaceae lineage-specific duplication event, which is consistent with the finding by Lou et al. (2012) based on synteny analysis. Lou et al. (2012) further determined that LHY was the ancestral copy while CCA1 was derived from the duplicated copy. We identified two distinct LHY paralogs in pineapple genome, Ac_LHY-1 (Aco016649) and Ac_LHY-2 (Aco013228), which share 39% amino acid sequence identity. Our phylogenetic

analysis showed that Ac_LHY-1 was likely the ancestral copy while Ac_LHY-2 was derived from a duplication event after monocots diverged with dicots (fig. 9b and supplementary fig. S13, Supplementary Material online). We further identified both orthologs of Ac_LHY-1 and Ac_LHY-2 in *E. guineensis* and *P. dactylis*, suggesting Ac_LHY-2 was likely derived from Ac_LHY-1 in the tau whole genome duplication event. One of the LHY homologs in *E. guineensis*, XP_010919033.1, is more closely related to Ac_LHY-1 (49% sequence identity and 525 bit-score) than to Ac_LHY-2 (47% sequence identity and 266 bit-score). Another LHY homolog from *E. guineensis*, XP_010941176.1, is more closely related to Ac_LHY-2 (51% sequence identity and 543 bit-score) than to Ac_LHY-1 (40% sequence identity and 377 bit-score). In contrast, rice genome contains a single copy of LHY (LOC_Os08g061100) that shares a higher degree of similarity with Ac_LHY-2 (54% sequence identity and 362 bit-score) than Ac_LHY-1 (38% sequence identity and 320 bit-score), suggesting the ancestral copy of LHY homolog has been lost in the rice genome (fig. 9b).

RVE8/RVE4/RVE6/RVE5/RVE3 Clade

The phylogenetic tree constructed using the RVE8, RVE4, RVE6, RVE3, and RVE5 homologs in plants revealed two distinct clades, the RVE8/RVE4 clade and RVE6/RVE5/RVE3 clade (supplementary fig. S14, Supplementary Material online). Lou et al. (2012) proposed that RVE8 and RVE6/RVE3 clades separated after the gamma triplication event and each clade gained additional expansion through α and β duplication events in Brassicaceae family. And RVE4 arose from RVE8 and RVE5 arose from RVE3 via α duplication event (Lou et al. 2012). However, our results suggested that the RVE8 group might have separated from RVE6/RVE3 group much earlier as we discovered one ortholog for each group in the basal angiosperm *A. trichopoda*. Lou et al. (2012) had missed one of these two *A. trichopoda* homologs in their study. RVE6/RVE5/RVE3 orthologs from monocots formed a monophyletic cluster, but no monocot ortholog was found for RVE8/RVE4 group (fig. 9b and supplementary fig. S14, Supplementary Material online). Although *Arabidopsis* RVE4 and RVE8 showed best blast hit to pineapple Aco029094.1, Aco013238.1, and Aco017509.1, all of them clustered with the RVE6/RVE5/RVE3 group. Therefore, the RVE8/RVE4 orthologs likely have been lost in monocots after monocots diverged from dicots.

Five RVE proteins from pineapple and three from rice were identified to be orthologs of the *Arabidopsis* RVE6/RVE5/RVE3. Phylogenetic analysis revealed that the RVE6/RVE5/RVE3 orthologs of monocots had undergone several rounds of duplication and fractionation events. Since RVE6 represents the ancestor of the RVE6/RVE5/RVE3 clade the best, we named the five RVE6/RVE5/RVE3 orthologs in pineapple Ac_RVE6-1 (Aco017509), Ac_RVE6-2 (Aco023196),

Ac_RVE6-3 (Aco028506), Ac_RVE6-4 (Aco029094), and Ac_RVE6-5 (Aco013238). The three RVE6 homologs in rice are Os_RVE6-1 (LOC_Os02g45670), Os_RVE6-2 (LOC_Os06g01670), and Os_RVE6-3 (LOC_Os06g45840). The phylogenetic tree revealed two groups of the monocot RVE6 homologs, one containing Os_RVE6-1 and its homologs from grasses and banana but none from pineapple and the other one containing Os_RVE6-2, Os_RVE6-3 and homologs from grasses, banana as well as pineapple (fig. 9b and supplementary fig. S14, Supplementary Material online). The two groups of the monocot RVE6 homologs were likely derived from the Tau duplication event and only one copy had been retained in pineapple genome. The retained copy gained additional expansion and resulted in current five RVE6 homologs in pineapple genome. In rice, Os_RVE6-2 and Os_RVE6-3 share 72% identity (bit-score: 400) and likely originated during the recent rho whole genome duplication event.

LUX/BOA

Phylogenetic analysis clearly separated LUX/BOA homologs of dicots from the ones of monocots (supplementary fig. S15, Supplementary Material online). Only a single copy of LUX/BOA ortholog was identified in pineapple (Aco016830), rice (LOC_Os01g74020) and the basal angiosperm species *A. trichopoda*. *Arabidopsis* LUX (AT3G46640) and BOA (AT5G59570) are paralogous genes that were likely evolved from a Brassicaceae lineage-specific duplication event.

CHE

Phylogenetic analysis also clearly separated CHE homologs of dicots from the ones of monocots (supplementary fig. S16, Supplementary Material online). Most plant species retained a single copy of CHE. The *Arabidopsis* CHE (AT5G08330) has a paralog (AT5G23280) that was likely evolved from a Brassicaceae lineage-specific duplication event. A single copy of CHE ortholog was identified in pineapple (Aco010326) and rice (LOC_Os02g58180). The rice CHE ortholog was missing in the MSU Rice Genome Annotation due to incorrect gene prediction at LOC_Os02g58180. We extracted the alternative overlapping CDS from the same location and used it in our analysis. When the true ortholog was missing from the annotated rice proteins, both Proteinortho-PoFF and SynMap incorrectly predicted two rice genes LOC_Os04g44440.1 and LOC_Os02g42380.1 as CHE orthologs. This case illustrated how automated prediction failed when true ortholog was missing.

Discussion

Phenotypic variation can be caused by not only the differences in gene coding sequences that change protein functions, but also the differences in regulatory networks that affect gene expression (Oleksiak et al. 2002). Studies have demonstrated

that variation in gene expression is the primary source leading to natural variation (Oleksiak et al. 2002) and trans-acting regulatory variation is responsible for most differences in gene expression (Brem et al. 2002). Therefore, identification and characterization of TFs and TCs are important to understand not only the regulatory networks controlling the biological processes, but also the evolutionary driving force leading to the biodiversification.

Most TFs belong to multigene families and the size of each TF gene family varies considerably among genomes. Lineage-specific expansions of TF families are common and have played important roles in both plant and animal diversification. Compared with animal genomes, TF families in plants have undergone a much higher degree of expansion, partly caused by genome-wide duplications (Shiu et al. 2005). In plants, TF exhibited a higher rate of retention than other genes after genome-wide duplications (Shiu et al. 2005). In our study, we identified 1,398 TFs and 80 TCs in pineapple genome. Considering the total number of genes annotated in the pineapple draft sequence at 27,024 (Ming et al. 2015), pineapple TFs account for ~5.2% of its estimated total number of genes. This ratio is similar to the one estimated in *Arabidopsis* at 5.9–7.5% (Riechmann et al. 2000; Riaño-Pachón et al. 2007), the one in maize at 4.2–6.9% (Lin et al. 2014), and the one estimated for millet at 5.0–5.6% (Lin et al. 2014). Unlike the grass genomes, the pineapple genome hasn't undergone the pan-cereal genome duplication event (ρ) (Ming et al. 2015). The relatively smaller number TFs and TCs identified in the pineapple genome than grass genomes might have resulted from the absence of the recent genome-wide duplication. However, the size of each TF and TC family didn't show proportional changes between the pineapple and grass genomes, suggesting differential expansions of TF and TC families across lineages. Compared with grass genomes, the pineapple genome contains smaller numbers of TFs and TCs for most TF and TC families with few exceptions, such as the TF family GRF and the TC family GIF. Interestingly, GIFs are the transcriptional coactivators of GRFs. GRFs and GIFs form complexes and play important roles in regulating leaf growth and senescence (Debernardi et al. 2014). The interacting and functionally related proteins tend to coevolve (Juan et al. 2008) because changes in copy number of these genes may disrupt the stoichiometric balance of their gene products with those of other genes (Birchler and Veitia 2010; Birchler et al. 2001; Freeling and Thomas 2006). Coexpansion of GRFs and GIFs in the pineapple genome might be explained by this *dosage-balance selection* theory. It may be interesting to investigate if this lineage-specific expansion contributed to the emergence of adaptive phenotypic innovations.

Lineage-specific expansions of gene families may result from genome-wide duplication, as well as segmental and tandem duplication events (Wilkins et al. 2009; Nakano et al. 2006). Duplication of TF and TC creates selective advantage

in having a viable genetic system capable of regulating growth, development, and responses to the environment. Due to the deleterious effects of dosage imbalance (Birchler and Veitia 2010), duplicated TFs and TCs were subjected to differential retention under selective pressure. In this study, we examined duplicate-gene retention, loss, and evolution in major circadian clock genes and refine their evolutionary relationships. Our results revealed significant variations in the fractions of retained duplicated genes although the plant species share same ancient whole-genome duplications. Lineage-specific expansion and differential retention had observed in almost all the clock genes in both monocots and dicots. Duplicated genes provide the raw material for evolving new functions. Therefore, lineage-specific expansion and differential retention of clock genes are likely linked to the evolutionary success and conferred fitness benefits during species radiation.

The retained duplicated genes can further evolve new functions via neofunctionalization, partition their ancestral roles via subfunctionalization, or accumulate deleterious mutations and decay as pseudogenes (Lynch and Force 2000; Lynch et al. 2001). In this study, we identified two copies of homologous gene for each of the three circadian genes, CCA1/LHY, COL9, and CDF. Phylogenetic analysis indicated the duplication events of CCA1/LHY and COL9 occurred in the common ancestor of monocots but lost in the grass family later, suggesting differential retentions of CCA1/LHY and COL9 took place across monocots. While, the duplication of CDF is likely a pineapple-specific expansion event. We further investigated tissue-specific expression of the three pairs of circadian genes. Although all of them showed the same phase between nonphotosynthetic and photosynthetic leaf tissues, each pair exhibited different expression patterns between the two tissues. The differential gene expression patterns of these duplicated genes suggested they might have partitioned their functions in different tissues through subfunctionalization.

In a multicellular organism, different types of cells harboring the same genomic constituents can have distinct structure and perform dramatically different functions. Tissue-identity is believed to be achieved mainly through tissue-specific gene expression and regulatory mechanisms, including epigenetic modification and transcriptional and posttranscriptional regulation (Gaudinier et al. 2015). Therefore, tissue-specific genes are believed to contribute to the structural and functional diversification of different tissue types. Elucidation of regulatory networks controlling the spatial and temporal gene expression can provide broader and deeper insights into molecular mechanisms underlying tissue-specific functions. Tissue-specific TFs and TCs preferentially connect to genes with tissue-specific functions and play pivotal roles in orchestrating the complex regulatory networks that lead to cell or tissue identity. In this study, the largest number of TFs and TCs that showed dynamic changes in transcript abundance was

observed in root. Root plays a vital role in whole-plant development by providing anchorage to the ground, uptaking water and nutrients from soil, and synthesizing amino acids and hormones. Roots are the primary organs that first sense the soil environment and respond to biotic and abiotic stresses, and communicate with aboveground plant parts via signaling pathways. Significantly higher numbers of tissue-specific or tissue-enriched TFs and TCs in root than other tissues may reflect the distinct regulatory networks evolved in root to adapt to the unique underground environment and enable the dynamic regulation in response to the constantly changing environment.

Genes with similar expression patterns and genes with similar functions are largely and tightly coregulated (Tavazoie et al. 1999). Based on this hypothesis, we investigated the coexpression patterns of TFs and TCs among different tissues. Our study showed that a large number of TFs and TCs exhibited coexpression patterns between leaf and flower. It is widely accepted that flowers are modified shoots and floral organs are modified leaves (Goethe 1790). A large number of TFs and TCs shared similar expression patterns between leaf and flower may suggest conserved regulatory networks have remained in regulating leaf and flower development due to their common origin. Interestingly, a large number of TFs and TCs that shared similar expression patterns were also observed between root and fruit. Root and fruit are major sink organs (Wardlaw 1990). Unraveling networks of coexpressed genes may help us to gain further insights into evolutionary conservation of regulatory networks between these two morphologically distinct tissues.

Circadian clock plays the fundamental role in the regulation of plant metabolic reactions, physiological processes, development, and response to environment. An *in vivo* enhancer trapping study revealed that 36% of the *Arabidopsis* transcriptome was regulated by the circadian clock (Michael and McClung 2003). A much larger portion of the transcriptome was identified to be under circadian control when the assessment was conducted on multiple photo- and thermo-cycles and combinations of these photo- and thermo-cycles (Michael et al. 2008). By analyzing 11 diurnal and circadian time courses, 89% of *Arabidopsis* transcripts were found to exhibit cycling expression patterns in at least one condition (Michael et al. 2008). In our study, ~42% of TFs and 45% of TCs displayed rhythmic expression patterns in pineapple genome. Since only leaf tissues were included in our diurnal expression analysis, we would expect the breadth of circadian regulation of transcription in pineapple genome could be even wider.

The circadian clock plays a major role in the temporal regulation of an organism's metabolism and physiology. An a large number of rhythmic TFs and TCs are required in order to achieve robust and tightly regulated circadian oscillations and ensure appropriate biological processes occur at the right time of the day (Michael and McClung 2003; Michael et al.

2008). Most circadian expressed genes showed peak expression around dawn or dusk to assist plants anticipating daily light transitions (Doherty and Kay 2010). Our results agree with this finding that a large number of TFs and TCs showed peak expression around dawn or dusk with 10 AM having the most and 8 and 10 PM having the least. We also found more cycling TF/TCs showing peak expression at daytime than the ones at night, which may reflect the wider range of environment changes at daytime than nighttime. In addition, only leaf tissue was included in our diurnal expression analysis. Therefore, a higher number of cycling TF/TCs showing peak expression at daytime than the ones at night may also reflect the primary function of the leaf in capturing light and performing photosynthesis. Note that these estimates may be biased due to one extra sample collected during the day compared with the ones collected at night.

In general, all the cells in plants contain an autonomous circadian clock (McClung 2006). However, circadian oscillators within each individual cell respond to entraining signals differently, and control different physiological outputs (James et al. 2008; Endo et al. 2014). It has been shown that the plant clock is organ-specific but not organ-autonomous (James et al. 2008). Therefore, heterogeneous oscillator networks of different plant cells and organs must be coupled on a systemic level in order to produce physiologically meaningful signals. In our study, we identified 625 TFs and TCs that exhibited diurnal expression patterns in either photosynthetic or nonphotosynthetic leaf tissues, or both tissues. Among the 625 cycling TFs and TCs, 28% showed diurnal expression patterns in nonphotosynthetic leaf tissue only and 40% was identified as cycling in photosynthetic leaf tissue only, which reflect divergent regulation between photosynthetic and nonphotosynthetic leaf tissues. Among the 625 cycling TFs and TCs, 201 of them exhibited diurnal expression patterns in both photosynthetic and nonphotosynthetic leaf tissues, which offered us an opportunity to investigate coupled oscillator system between the two tissues. We found that 34% of these genes were tightly coupled. Interestingly, more than 90% of these coupled TFs and TCs displayed peak expression between 6 AM and 4 PM.

Core clock components tend to express at high-amplitude and robust cycling patterns across tissues and different environmental conditions (Doherty and Kay 2010). In our study, we found that the known core clock homologous genes were expressed at high-amplitude and enriched in the group of cycling genes that showed tightly coupled-expression patterns between photosynthetic and nonphotosynthetic leaf tissues in pineapple. Our results may suggest a new paradigm to identify new core clock genes in pineapple genome by identification of transcripts with high-amplitude expression and robust and coupled cycling patterns across tissues.

The CAM photosynthesis pathway is strictly temporally regulated by the endogenous circadian clock (Warren and Wilkins 1961). Phosphoenolpyruvate carboxylase (PEPC), the key enzyme of CAM that catalyzes the first step of the CAM

pathway, is regulated posttranslationally via reversible phosphorylation catalyzed by the PEPC kinase (PPCK) (Carter et al. 1991). And the expression of the PPCK gene is controlled by a circadian oscillator (Hartwell et al. 1999) mediated by cytosolic malate (Borland et al. 1999). The primary effect of the circadian control is on malate transport across the tonoplast, and the diurnal expression of PPCK is a secondary effect (Borland et al. 1999). However, the circadian oscillator regulating CAM activities is still unknown. In this study, we identified a pineapple homolog of AT-STOP1, Aco011214.1, whose diurnal expression pattern coincides with the diurnal oscillation of malate concentration in pineapple leaf. Furthermore, AT-STOP1 activates the expression of ALUMINUM-ACTIVATED MALATE TRANSPORTER1 (AtALMT1), which encodes a malate channel critical for aluminum resistance in *Arabidopsis* (Sawaki et al. 2009; Tokizawa et al. 2015). AtALMT1 belongs to the ALMT (aluminum-activated malate transporter) family. Besides encoding malate channels, several ALMT members are involved in stomatal movement (Palmer et al. 2016). Taken all together, our findings suggest Aco011214.1, the pineapple STOP1 homolog, may be the key circadian oscillator regulating CAM metabolism.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank M. Conway at Dole Plantation for assistance in time-course leaf sample collection, and Ratnesh Singh for technical support in data analysis. This work was supported by the United States Department of Agriculture T-START grant through the University of Hawaii to Q.Y. and R.M., the United States Department of Agriculture National Institute of Food and Agriculture Hatch Project TEX0-1-9374 to Q.Y., and the National Natural Science Foundation of China under Grant NO. 31628013. The open access publishing fees for this article have been covered in part by the Texas A&M University Open Access to Knowledge Fund (OAKFund), supported by the University Libraries and the Office of the Vice President for Research. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Literature Cited

- Arnone MI, Davidson EH. 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Dev Camb Engl*. 124: 1851–1864.
- Barak S, Tobin EM, Andronis C, Sugano S, Green RM. 2000. All in good time: the *Arabidopsis* circadian clock. *Trends Plant Sci*. 5(12): 517–522.
- Bingham J, Sudarsanam S. 2000. Visualizing large hierarchical clusters in hyperbolic space. *Bioinformatics* 16(7): 660–661.

- Birchler JA, Bhadra U, Bhadra MP, Auger DL. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev Biol.* 234(2): 275–288.
- Birchler JA, Veitia RA. 2010. The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytol.* 186(1): 54–62.
- Bläsing OE, et al. 2005. Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell* 17(12): 3257–3281.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Borland AM, Hartwell J, Jenkins GI, Wilkins MB, Nimmo HG. 1999. Metabolite control overrides circadian regulation of phosphoenolpyruvate carboxylase kinase and CO₂ fixation in Crassulacean acid metabolism. *Plant Physiol.* 121(3): 889–896.
- Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* 296(5568): 752–755.
- Buhr ED, Yoo S-H, Takahashi JS. 2010. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science* 330(6002): 379–385.
- Carter PJ, Nimmo HG, Fewson CA, Wilkins MB. 1991. Circadian rhythms in the activity of a plant protein kinase. *EMBO J.* 10(8): 2063–2068.
- Chen W, et al. 2002. Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* 14(3): 559–574.
- Cheng X-F, Wang Z-Y. 2005. Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in *Arabidopsis thaliana*. *Plant J Cell Mol Biol.* 43(5): 758–768.
- Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. 2008. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol.* 9(8): R130.
- Crayn DM, Winter K, Smith JAC. 2004. Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. *Proc Natl Acad Sci U S A.* 101(10): 3703–3708.
- Cushman JC. 2001. Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments. *Plant Physiol.* 127(4): 1439–1448.
- Dai S, et al. 2011. Brother of lux arrhythmo is a component of the *Arabidopsis* circadian clock. *Plant Cell* 23(3): 961–972.
- Dalquen DA, Dessimoz C. 2013. Bidirectional best hits miss many orthologs in duplication-rich clades such as plants and animals. *Genome Biol Evol.* 5(10):1800–1806.
- Debernardi JM, et al. 2014. Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. *Plant J Cell Mol Biol.* 79(3): 413–426.
- Doherty CJ, Kay SA. 2010. Circadian control of global gene expression patterns. *Annu Rev Genet.* 44: 419–444.
- Dong MA, Farré EM, Thomashow MF. 2011. Circadian clock-associated 1 and late elongated hypocotyl regulate expression of the C-repeat binding factor (CBF) pathway in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 108: 7241–7246.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Endo M, Shimizu H, Nohales MA, Araki T, Kay SA. 2014. Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature* 515(7527): 419–422.
- Fang J, Miao C, Chen R, Ming R. 2016. Genome-wide comparative analysis of microsatellites in pineapple. *Trop Plant Biol.* 9(3): 117–135.
- Filichkin SA, et al. 2011. Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PLoS One* 6(6): e16907.
- Fornara F, et al. 2009. *Arabidopsis* DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev Cell* 17(1): 75–86.
- Fowler SG, Cook D, Thomashow MF. 2005. Low temperature induction of *Arabidopsis* CBF1, 2, and 3 is gated by the circadian clock. *Plant Physiol.* 137(3): 961–968.
- Franco-Zorrilla JM, et al. 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci U S A.* 111: 2367–2372.
- Freeling M, Thomas BC. 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* 16(7): 805–814.
- Frietze S, Farnham PJ. 2011. Transcription factor effector domains. *Subcell Biochem.* 52: 261–277.
- Fulton DL, et al. 2006. Improving the specificity of high-throughput ortholog prediction. *BMC Bioinformatics* 7: 270.
- Gangappa SN, Botto JF. 2014. The BBX family of plant transcription factors. *Trends Plant Sci.* 19(7): 460–470.
- Gao G, et al. 2006. DRTF: a database of rice transcription factors. *Bioinformatics* 22(10): 1286–1287.
- Gaudinier A, Tang M, Kliebenstein DJ. 2015. Transcriptional networks governing plant metabolism. *Curr Plant Biol.* 3–4: 56–64.
- Gendron JM, et al. 2012. *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc Natl Acad Sci U S A.* 109(8): 3167–3172.
- Goethe JWV. 1790. Versuch die Metamorphose der Pflanzen zu erklären. C. W. Ettinger: Gotha.
- Goodspeed D, Chehab EW, Min-Venditti A, Braam J, Covington MF. 2012. *Arabidopsis* synchronizes jasmonate-mediated defense with insect circadian behavior. *Proc Natl Acad Sci U S A.* 109(12): 4674–4677.
- Greenham K, McClung CR. 2015. Integrating circadian dynamics with physiological processes in plants. *Nat Rev Genet.* 16(10): 598–610.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3): 307–321.
- Hall T. 2011. BioEdit: an important software for molecular biology. *GERF Bull Biosci.* 2: 60–61.
- Hartwell J, et al. 1999. Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *Plant J Cell Mol Biol.* 20(3): 333–342.
- Haydon MJ, Bell LJ, Webb AAR. 2011. Interactions between plant circadian clocks and solute transport. *J Exp Bot.* 62(7):2333–2348.
- Hsu PY, Harmer SL. 2014. Wheels within wheels: the plant circadian system. *Trends Plant Sci.* 19(4): 240–249.
- Huang J, Zhao X, Weng X, Wang L, Xie W. 2012. The rice B-box zinc finger gene family: genomic identification, characterization, expression profiling and diurnal analysis. *PLoS One* 7(10): e48242.
- Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA. 2005. FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in *Arabidopsis*. *Science* 309(5732): 293–297.
- Ishihama A, Shimada T, Yamazaki Y. 2016. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Res.* 44(5): 2058–2074.
- Iuchi S, et al. 2007. Zinc finger protein STOP1 is critical for proton tolerance in *Arabidopsis* and coregulates a key gene in aluminum tolerance. *Proc Natl Acad Sci U S A.* 104(23): 9900–9905.
- James AB, et al. 2008. The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science* 322(5909): 1832–1835.
- Jin J, Zhang H, Kong L, Gao G, Luo J. 2014. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 42(Database issue): D1182–D1187.
- Juan D, Pazos F, Valencia A. 2008. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc Natl Acad Sci U S A.* 105(3): 934–939.

- Kenyon WH, Severson RF, Black CC. 1985. Maintenance carbon cycle in crassulacean acid metabolism plant leaves: source and compartmentation of carbon for nocturnal malate synthesis. *Plant Physiol.* 77(1): 183–189.
- Kim JH, Choi D, Kende H. 2003. The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in *Arabidopsis*. *Plant J Cell Mol Biol.* 36(1): 94–104.
- van der Knaap E, Kim JH, Kende H. 2000. A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. *Plant Physiol.* 122(3): 695–704.
- Kolmos E, Chow BY, Pruneda-Paz JL, Kay SA. 2014. HsfB2b-mediated repression of PRR7 directs abiotic stress responses of the circadian clock. *Proc Natl Acad Sci U S A.* 111(45): 16172–16177.
- Lechner M, et al. 2014. Orthology detection combining clustering and synteny for very large datasets. *PLoS One* 9(8): e105015.
- Lin J-J, Yu C-P, Chang Y-M, Chen SC-C, Li W-H. 2014. Maize and millet transcription factors annotated using comparative genomic and transcriptomic data. *BMC Genomics* 15(1): 818.
- Lin R-C, Park H-J, Wang H-Y. 2008. Role of *Arabidopsis* RAP2.4 in regulating light- and ethylene-mediated developmental processes and drought stress tolerance. *Mol Plant.* 1(1): 42–57.
- Lou P, et al. 2012. Preferential retention of circadian clock genes during diploidization following whole genome triplication in *Brassica rapa*. *Plant Cell* 24: 2415–2426.
- Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154(1): 459–473.
- Lynch M, O’Hely M, Walsh B, Force A. 2001. The Probability of Preservation of a Newly Arisen Gene Duplicate. *Genetics* 159(4): 1789–1804.
- Makino S, et al. 2000. Genes encoding pseudo-response regulators: insight into His-to-Asp phosphorelay and circadian rhythm in *Arabidopsis thaliana*. *Plant Cell Physiol.* 41: 791–803.
- McClung CR. 2006. Plant circadian rhythms. *Plant Cell* 18(4): 792–803.
- Michael TP, et al. 2008. Network discovery pipeline elucidates conserved time-of-day—specific cis-regulatory modules. *PLoS Genet.* 4(2): e14.
- Michael TP, McClung CR. 2003. enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis*. *Plant Physiol.* 132(2): 629–639.
- Ming R, et al. 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet.* 47(12): 1435–1442.
- Mizoguchi T, et al. 2002. LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in *Arabidopsis*. *Dev Cell* 2(5): 629–641.
- Mockler TC, et al. 2007. The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb Symp Quant Biol.* 72: 353–363.
- Nakamichi N. 2011. Molecular mechanisms underlying the *Arabidopsis* circadian clock. *Plant Cell Physiol.* 52(10): 1709–1718.
- Nakamichi N, et al. 2010. PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the *Arabidopsis* circadian clock. *Plant Cell* 22(3): 594–605.
- Nakamichi N, et al. 2009. Transcript profiling of an *Arabidopsis* PSEUDO RESPONSE REGULATOR arrhythmic triple mutant reveals a role for the circadian clock in cold stress response. *Plant Cell Physiol.* 50(3): 447–462.
- Nakano T, Suzuki K, Fujimura T, Shinshi H. 2006. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol.* 140(2): 411–432.
- Nakashima K, Ito Y, Yamaguchi-Shinozaki K. 2009. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. *Plant Physiol.* 149(1): 88–95.
- Narlikar L, Ovcharenko I. 2009. Identifying regulatory elements in eukaryotic genomes. *Brief Funct Genomic Proteomic.* 8(4): 215–230.
- Niehrs C, Pollet N. 1999. Synexpression groups in eukaryotes. *Nature* 402(6761): 483–487.
- Nohales MA, Kay SA. 2016. Molecular mechanisms at the core of the plant circadian oscillator. *Nat Struct Mol Biol.* 23(12): 1061–1069.
- Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. *Nat Genet.* 32(2): 261–266.
- Omidbakhshfard MA, Proost S, Fujikura U, Mueller-Roeber B. 2015 2015. Growth-Regulating Factors (GRFs): a small transcription factor family with important functions in plant biology. *Mol Plant.* 8(7): 998–1010.
- Palmer AJ, Baker A, Muench SP. 2016 2016. The varied functions of aluminium-activated malate transporters—much more than aluminium resistance. *Biochem Soc Trans.* 44(3): 856–862.
- Para A, et al. 2007. PRR3 is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell* 19(11): 3462–3473.
- Paranjpe DA, Sharma VK. 2005. Evolution of temporal order in living organisms. *J Circadian Rhythms* 3(1): 7.
- Paull RE, et al. 2016. Carbon flux and carbohydrate gene families in pineapple. *Trop Plant Biol.* 9(3): 200–213.
- Pérez-Rodríguez P, et al. 2010. PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38(Database issue): D822–D827.
- Pruneda-Paz JL, Breton G, Para A, Kay SA. 2009. A functional genomics approach reveals CHE as a component of the *Arabidopsis* circadian clock. *Science* 323(5920): 1481–1485.
- Rainha N, et al. 2016. Leaf malate and succinate accumulation are out of phase throughout the development of the CAM plant *Ananas comosus*. *Plant Physiol Biochem.* 100: 47–51.
- Rawat R, et al. 2009. REVELLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. *Proc Natl Acad Sci U S A.* 106(39): 16883–16888.
- Reinke H, et al. 2008. Differential display of DNA-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes Dev.* 22(3): 331–345.
- Reyes JC, Muro-Pastor MI, Florencio FJ. 2004. The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol.* 134(4): 1718–1732.
- Riño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B. 2007. PlnTFDB: an integrative plant transcription factor database. *BMC Bioinformatics* 8: 42.
- Riechmann JL, et al. 2000. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290(5499): 2105–2110.
- Romanel EAC, Schrago CG, Couñago RM, Russo CAM, Alves-Ferreira M. 2009. Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PLoS One* 4: e5791.
- Sawaki Y, et al. 2009. STOP1 regulates multiple genes that protect *Arabidopsis* from proton and aluminum toxicities. *Plant Physiol.* 150(1): 281–294.
- Schaffer R, et al. 1998. The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* 93(7): 1219–1229.
- Scott MP. 2000. Development: the natural history of genes. *Cell* 100(1): 27–40.
- Shiu S-H, Shih M-C, Li W-H. 2005. Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol.* 139(1): 18–26.
- Singh R, Ming R, Yu Q. 2016. Comparative analysis of GC content variations in plant genomes. *Trop Plant Biol.* 9(3): 136–149.
- Smith SM, et al. 2004. Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in *Arabidopsis* leaves. *Plant Physiol.* 136(1): 2687–2699.

- Strayer C, et al. 2000. Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. *Science* 289(5480): 768–771.
- Takahashi N, Hirata Y, Aihara K, Mas P. 2015. A hierarchical multi-oscillator network orchestrates the *Arabidopsis* circadian system. *Cell* 163(1): 148–159.
- Takata N, Saito S, Saito CT, Uemura M. 2010. Phylogenetic footprint of the plant clock system in angiosperms: evolutionary processes of pseudo-response regulators. *BMC Evol Biol.* 10: 126.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. 1999. Systematic determination of genetic network architecture. *Nat Genet.* 22(3): 281–285.
- Tokizawa M, et al. 2015. Sensitive to proton rhizotoxicity1, calmodulin binding transcription activator2, and other transcription factors are involved in aluminum-activated malate transporter1 expression. *Plant Physiol.* 167(3): 991–1003.
- Trapnell C, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 7(3): 562–578.
- Troein C, Locke JCW, Turner MS, Millar AJ. 2009. Weather and seasons together demand complex biological clocks. *Curr Biol.* 19(22): 1961–1964.
- Wai CM, Powell B, Ming R, Min XJ. 2016a. Analysis of alternative splicing landscape in pineapple (*Ananas comosus*). *Trop Plant Biol.* 9(3): 150–160.
- Wai CM, Powell B, Ming R, Min XJ. 2016b. Genome-wide identification and analysis of genes encoding proteolytic enzymes in pineapple. *Trop Plant Biol.* 9(3): 161–175.
- Walhout AJM. 2006. Unraveling transcription regulatory networks by protein–DNA and protein–protein interaction mapping. *Genome Res.* 16(12): 1445–1454.
- Walter W, et al. 2014. Improving the accuracy of expression data analysis in time course experiments using resampling. *BMC Bioinformatics* 15(1):352.
- Wang W, et al. 2011. Timing of plant immune responses by a central circadian regulator. *Nature* 470(7332): 110–114.
- Wang ZY, et al. 1997. A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis* Lhcb gene. *Plant Cell* 9(4): 491–507.
- Wardlaw IF. 1990. Tansley Review No. 27 The control of carbon partitioning in plants. *New Phytol.* 116(3): 341–381.
- Warren DM, Wilkins MB. 1961. An endogenous rhythm in the rate of dark-fixation of carbon dioxide in leaves of *Bryophyllum fedtschenkoi*. *Nature* 191(4789): 686–688.
- West-Eberhard MJ, Smith JAC, Winter K. 2011. Photosynthesis, reorganized. *Science* 332(6027): 311–312.
- Wilkins O, Bräutigam K, Campbell MM. 2010. Time of day shapes *Arabidopsis* drought transcriptomes. *Plant J Cell Mol Biol.* 63(5): 715–727.
- Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. 2009. Expansion and diversification of the Populus R2R3-MYB family of transcription factors. *Plant Physiol.* 149(2): 981–993.
- Wray GA et al. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.* 20(9): 1377–1419.
- Xie Q, et al. 2014. LNK1 and LNK2 are transcriptional coactivators in the *Arabidopsis* circadian oscillator. *Plant Cell* 26(7): 2843–2857.
- Yilmaz A, et al. 2009. GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol.* 149(1): 171–180.
- Zhang X, Liang P, Ming R. 2016. Genome-wide identification and characterization of nucleotide-binding site (NBS) resistance genes in pineapple. *Trop Plant Biol.* 9(3): 187–199.
- Zheng Y, et al. 2016. Identification of microRNAs, phasiRNAs and their targets in pineapple. *Trop Plant Biol.* 9(3): 176–186.

Associate editor: Wen-Hsiung Li