https://doi.org/10.1093/plphys/kiac121 PLANT PHYSIOLOGY 2022: 190: 994–1004

Plant Physiology®

Breakthrough Tech

CAST-R: An application to visualize circadian and heat stress-responsive genes in plants

Titouan Bonnot [1], Morgane B. Gillard [1] and Dawn H. Nagel [1,*,†]

1 Department of Botany and Plant Sciences, University of California, Riverside, California, USA

*Author for correspondence: dawnn@ucr.edu
†Senior author
These authors contributed equally (T.B. and M.B.G.).
D.H.N. conceived the project, T.B. and M.B.G. designed and developed the tool, and T.B., M.B.G., and D.H.N. wrote the manuscript.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/general-instructions) is Dawn H. Nagel (dawnn@ucr.edu).

Abstract

The circadian clock helps organisms to anticipate and coordinate gene regulatory responses to changes in environmental stimuli. Under stresses, both time of day and the circadian clock closely control the magnitude of plant responses. The identification of clock-regulated genes is, therefore, important when studying the influence of environmental factors. Here, we present CAST-R (Circadian And heat STress-Responsive), a "Shiny" application that allows users to identify and visualize circadian and heat stress-responsive genes in plants. More specifically, users can generate and export profiles and heatmaps representing transcript abundance of a single or of multiple Arabidopsis (*Arabidopsis thaliana*) genes over a 24-h time course, in response to heat stress and during recovery following the stress. The application also takes advantage of published Arabidopsis chromatin immunoprecipitation-sequencing datasets to visualize the connections between clock proteins and their targets in an interactive network. In addition, CAST-R offers the possibility to perform phase (i.e. timing of expression) enrichment analyses for rhythmic datasets from any species, within and beyond plants. This functionality combines statistical analyses and graphical representations to identify significantly over- and underrepresented phases within a subset of genes. Lastly, profiles of transcript abundance can be visualized from multiple circadian datasets generated in Arabidopsis, *Brassica rapa*, barley (*Hordeum vulgare*), and rice (*Oryza sativa*). In summary, CAST-R is a user-friendly interface that allows the rapid identification of circadian and stress-responsive genes through multiple modules of visualization. We anticipate that this tool will make it easier for users to obtain temporal and dynamic information on genes of interest that links plant responses to environmental signals.

Introduction

The circadian clock is an endogenous timekeeper mechanism, which allows organisms to anticipate daily and seasonal variations of environmental factors. In diurnal conditions, transcripts between 25% and 90% show rhythmic expression in plants (Michael et al., 2008; Filichkin et al., 2011; Ferrari et al., 2019; Lai et al., 2020). Significant proportions of these genes show rhythmic transcription or translation during the day in the absence of environmental cues and therefore exhibit circadian oscillations (Mockler et al., 2007; Hsu and Harmer, 2012; Romanowski et al., 2020; Bonnot and Nagel, 2021). Thereby, thousands of genes involved in diverse biological processes are clock controlled in plants (Harmer et al., 2000; Farré, 2012; Greenham and McClung, 2015). In Arabidopsis (*Arabidopsis thaliana*), for example, circadian oscillations are observed for approximately

Received November 10, 2021. Accepted February 21, 2022. Advance access publication March 16, 2022

30%–40% of the transcriptome and translatome (Mockler et al., 2007; Hsu and Harmer, 2012; Romanowski et al., 2020; Bonnot and Nagel, 2021). Remarkably, up to three-quarters of the Brassica (*Brassica rapa*) transcriptome exhibit circadian oscillations, highlighting the strong influence of the clock on the regulation of gene expression in plants (Greenham et al., 2020).

The control of rhythmicity not only defines the timing of expression and translation of genes, but also their response to stresses. Indeed, two stresses of the same nature and strength can lead to cellular responses with different intensities if occurring at two different times of day (Grundy et al., 2015). This phenomenon, described as gating, influences one-third of the heat stress-responsive circadian translatome in Arabidopsis (Bonnot and Nagel, 2021). Due to expression rhythmicity, a gene can be upregulated in response to a stress occurring during its lowest expression level, while the changes can be nonsignificant during its peak expression (Bonnot and Nagel, 2021). Consequently, the lists of differentially regulated genes, commonly used to identify genes involved in the plant response, highly differ depending on the experimental design (e.g. the timing of stress and of samplings). Thus, it is of prime importance to consider the effect of time of day and the influence of the clock when studying the plant responses to stress.

The diurnal and circadian transcriptome datasets that have been generated over the past 15 years are great resources for biologists to identify if their genes of interest are potentially regulated by the circadian clock (Mockler et al., 2007; Filichkin et al., 2011; Hsu and Harmer, 2012; Ferrari et al., 2019; Lai et al., 2020; Romanowski et al., 2020; Bonnot and Nagel, 2021). The availability of public databases and tools to visualize these published data is also necessary to make this information easy to interpret. The web-based tool DIURNAL is particularly useful to identify the timing of expression of genes in diurnal and free-running conditions in several model plants, from multiple array experiments (http://diurnal.mocklerlab.org/; Mockler et al., 2007). In addition, chromatin immunoprecipitation-sequencing (ChIP-Seq) data for several circadian clock genes have allowed for the identification of direct binding targets of clock proteins (Table 1). Another web-based tool, ATTRACTOR, integrates the rhythmic transcriptomic (DIURNAL) and cistromic (Clock and Light Signaling ChIP-Seq) datasets to enable the inference of transcriptional control between the clock and the light signaling pathway in Arabidopsis (https://greennetwork.us.es/ATTRACTOR/; de los Reyes et al., 2020).

The above-mentioned datasets and tools visualize transcriptomic data under normal growth conditions continue to be an invaluable resource. In our recent study, we examined the circadian-regulated changes at the translation level in Arabidopsis. We profiled both the transcriptome and translatome in response to heat stress (Bonnot and Nagel, 2021). We noticed that a tool that incorporates large-scale circadian-regulated expression datasets at multiple levels of gene regulation and in response to environmental stresses is lacking. Here, we introduce CAST-R, an R package "Shiny" application that allows users to quickly identify and visualize (1) individual and groups of genes exhibiting circadian oscillations at the transcriptome and/or translatome levels, and responding to heat stress, (2) genes that are targeted by clock proteins, and (3) phases (timing of peak expression) that are over-represented in a list of genes of interest. Most of the information used in this Shiny application comes from datasets generated in Arabidopsis, but CAST-R also provides a functionality that allows users to compare circadian oscillations between datasets and plant species. Graphical representations— that include dot plots, heatmaps, interactive networks, and circular bar plots—can be exported in multiple formats, along with raw data. This first version of CAST-R, therefore, provides the plant biology research community with a user-friendly interface to graphically represent circadian and heat stress-responsive genes in plants. CAST-R can be accessed at https://nagellab.shinyapps.io/CASTR-v1/.

Table 1 Targets of clock proteins identified in ChIP-Seq experiments in Arabidopsis, at normal growth temperatures

Clock protein	Number of targets	References
CCA1	1,991	Nagel et al. (2015) Kamioka et al. (2016)
LHY	722	Adams et al. (2018)
PRR5	1,024	Nakamichi et al. (2012)
PRR7	113	Liu et al. (2013)
PRR9	132	Liu et al. (2016)
TOC1	772	Huang et al. (2012)
LUX	27	Ezer et al. (2017)

Results and discussion

Web application content

The first tab of CAST-R, "Introduction", presents the application and helps users to navigate through and to use the different tabs. Two tabs named "Single genes" and "Multiple genes" allow the visualization of circadian and heat stress-responsive genes at the transcriptome and translatome levels in Arabidopsis, from a single-gene locus or from multiple loci, respectively. In another tab "Network", connections between clock proteins and their selected downstream targets, identified in published ChIP-Seq data, can be visualized in an interactive network. These first three modules ("Single genes", "Multiple genes", and "Network") allow a rapid identification of clock-regulated genes in Arabidopsis. In addition, the joint graphical representation of the gene response and recovery to heat stress provides useful information on how time of day and the timing of gene expression and translation affect the plant responses to temperature stresses.

When using large-scale omics approaches to study the plant response to stresses, a common strategy is to identify differentially regulated genes and group them based on their pattern of response, through clustering or network analyses. It can then be inferred that grouped or connected genes are potentially acting together to coordinate the appropriate

regulatory response. When Arabidopsis plants are exposed to cold stress, the majority of upregulated genes peak in the afternoon, while most genes downregulated by cold peak around dawn (Grundy et al., 2015). Thus, significant proportions of genes responding to stress can be expressed at the same time during the day and therefore be co-regulated. Identifying over-representation of genes peaking at the same time during the day is, therefore, very helpful to better define co-expressed gene modules. We implemented in the tab "Phase enrichment" a functionality that allows for the identification and visualization of over-represented phases within a list of genes.

Lastly, a tab named "Multispecies circadian oscillations" allows users to compare the timing of transcript accumulation between published circadian datasets, and in different plant species: Arabidopsis, *B. rapa*, barley (*Hordeum vulgare*), and rice (*Oryza sativa*). Although the first version of CAST-R is limited to four species, future versions could include other plant species.

Module 1—Circadian oscillations and heat stress response of single genes

In the "Single genes" tab, users can enter an Arabidopsis AGI locus code (example: *AT3G47500*, i.e. the *CYCLING DOF FACTOR 3* [*CDF3*] gene, Figure 1) and click on "Submit". This will generate plots from the Arabidopsis transcriptome and translatome datasets published in Bonnot and Nagel (2021). All data were obtained from 12-d-old seedlings that were grown in light (12 h) and dark cycles (12 h) at 22°C for 10 days and then transferred to constant light for 2 d before sampling (see Bonnot and Nagel, 2021 for further details).

The first box shows the normalized transcript abundance of the gene during a 24-h time course at 22°C. Above the plots, a table summarizes the information related to the circadian oscillations: the "Phase", the "Adjusted *P*-value" corresponding to the BH.Q value of the Metacycle analysis, and a column "Cycling" indicating if the transcript is significantly cycling or not, based on the criteria used in Bonnot and Nagel (2021).

In a second box, the plots represent the time course obtained at 22°C and the heat stress response of the gene at different times of day, from early morning (ZT48) to end of night (ZT69). At each time point, the heat stress condition corresponds to a 1-h treatment at 37°C. Below the plots, users can visualize if the difference between the two temperatures is significant (false discovery rate [FDR] < 0.05, green dots) or not (FDR > 0.05, gray dots). In addition, heatmaps represent as color gradients the $\log_2$ fold change values (37°C versus 22°C). Of note, genes were considered as significantly differentially regulated when the FDR < 0.05 and the $\log_2$ fold change $> |1|$ in Bonnot and Nagel (2021). In our example, the gene *CDF3* is significantly upregulated at ZT66 at the transcriptome level and at ZT63 and ZT66 at the translatome level, based on these criteria (Figure 1).

In a third box, users can visualize as colored dots and heatmaps the gene recovery following a 1-h heat stress (37°C) applied in the middle of the day (from ZT53 to ZT54, i.e. from 5 to 6 h after dawn). Below the transcript profile, colored dots and heatmaps highlight the comparison between the stress and the control conditions, and represent FDR and log2 fold change values, respectively. For example, *CDF3* is significantly downregulated at 6 h of recovery at the transcriptome level (Figure 1).

Module 2—Circadian oscillations and heat stress response of multiple genes

The "Multiple genes" tab uses the same information as the "Single Genes" tab, but enables the user to visualize the results for multiple AGI codes at once. In the different boxes ("Time course at 22°C", "Heat stress response" and "Recovery following heat stress"), heatmaps are generated and allow the users to identify genes responding to heat stress, and to compare results between the transcriptome and the translatome. This functionality is particularly useful to analyze groups of genes with a similar pattern of expression or of response to stress, or that belong to the same gene family. These plots also allow users to rapidly visualize and identify circadian genes and the peak of expression (phase). As an example, we provided in CAST-R the lists of transcription factor (TF) families identified by Pruneda-Paz et al. (2014). Users can select any of the families and generate heatmaps. For example, if the "C2C2-DOF' TF family is selected, 30 members are represented on the heatmaps (Figure 2). Eight and 19 of them are significantly cycling at the transcriptome and at the translatome levels, respectively. Thus, 11 members exhibit circadian oscillations specifically at the translatome level, based on our data published by Bonnot and Nagel (2021). Others show circadian rhythmicity at both levels of regulation, such as five of the six members of the *CDF* subfamily (Figure 2). Users can either sort genes based on their phase, their average heat stress response over the day, their average heat stress recovery (between 1 h and 6 h of recovery following heat stress), or keep the same order as the gene list that they provide to CAST-R.

Module 3—Interactive network of Arabidopsis clock proteins and their targets

The "Network" functionality of CAST-R takes advantage of identified interactions between clock proteins and their targets in Arabidopsis to build an interactive network. Targets of CIRCADIAN CLOCK ASSOCIATED 1 (CCA1), LONG HYPOCOTYL (LHY), PSEUDO RESPONSE REGULATOR 5 (PRR5), PRR7, PRR9, LUX ARRHYTHMO (LUX), and TIMING OF CAB EXPRESSION 1 (TOC1) proteins were determined in published ChIP-Seq data (Table 1). These connections are represented in the network as edges from the clock proteins to the targeted genes (Figure 3). Similar to the "Multiple genes" tab, users can either select a TF family or paste a list of Arabidopsis AGI codes. From the selected or provided list, CAST-R identifies genes that are known to be targeted

Figure 1 Circadian oscillations, heat stress response and heat stress recovery of single gene visualized in the CAST-R tool. The gene *CDF3* (AT3G47500) is represented as an example. In the box "Time course at 22°C", the adjusted *P*-value corresponds to the BH.Q value obtained in the Metacycle analysis (for details, see Bonnot and Nagel, 2021). Gray areas represent the subjective night. Time of day is referred to as Zeitgeber Time (ZT) and corresponds to the hours after moving seedlings into constant conditions (light and temperature). ZT48 and ZT60 correspond to dawn and the beginning of the subjective night, respectively. The color scales correspond to $\log_2$ fold change values (37°C versus 22°C).

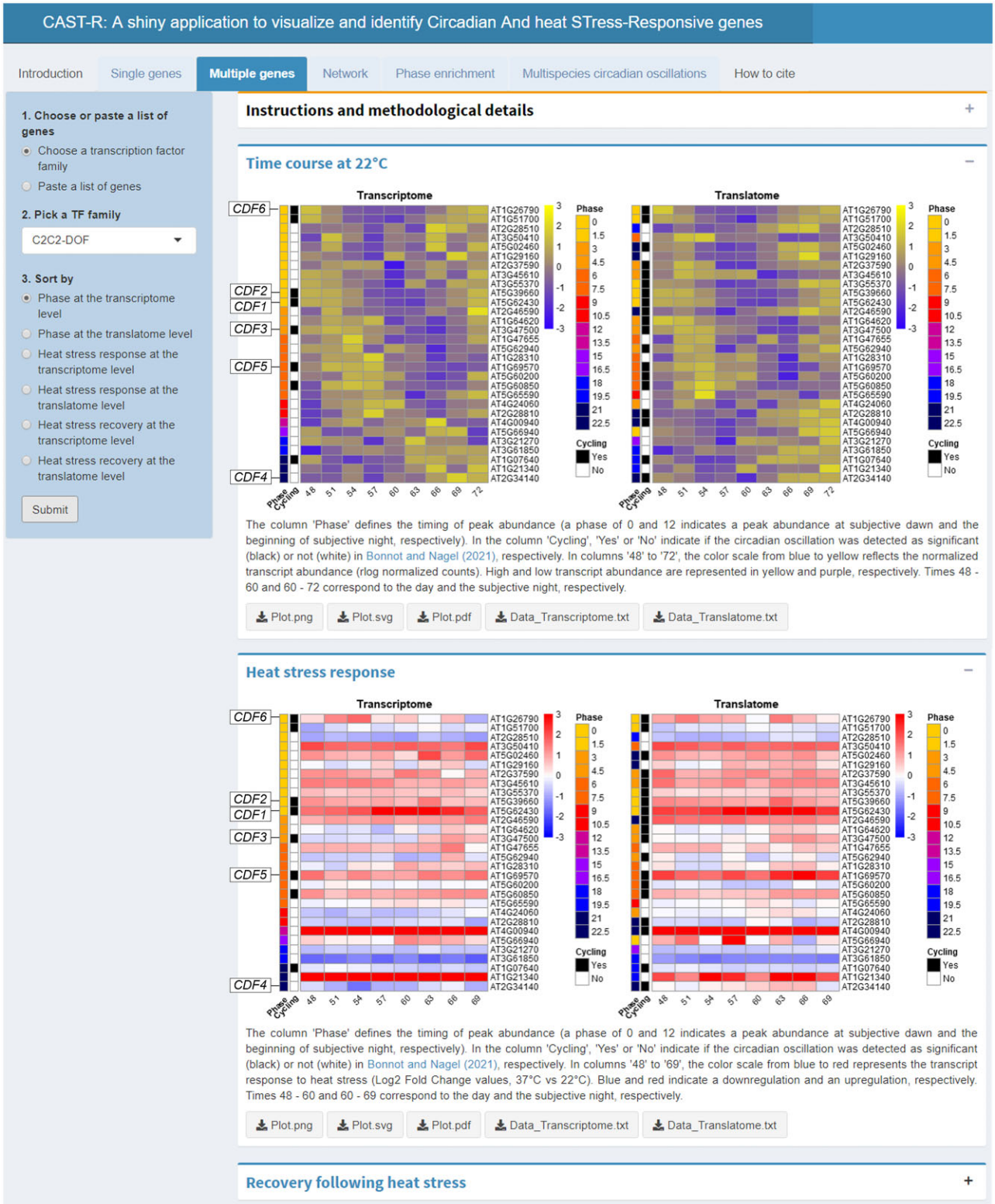


Figure 2 Circadian oscillations, heat stress response and heat stress recovery of multiple genes. The C2C2-DOF family is represented as an example. Members of the *CDF* subfamily are highlighted and have been manually annotated on the figure. For each heatmap, the column "Phase" indicates the transcript phase (timing of peak abundance), the color gradient from yellow to dark blue representing the subjective dawn and the end of the subjective night, respectively. The column "Cycling" indicates if the circadian oscillation was detected as significant (black) or not (white) in Bonnot and Nagel (2021), respectively. In the "Time course at 22°C" box, color gradients from purple to yellow represent the normalized transcript abundance (rlog normalized counts) during a 24 h time course. In the "Heat stress response" and "Recovery following heat stress" boxes, color gradients from blue to red correspond to $\log_2$ fold change values (37°C versus 22°C). For size constraints, heatmaps of the "Recovery following heat stress" box is not represented on this figure.

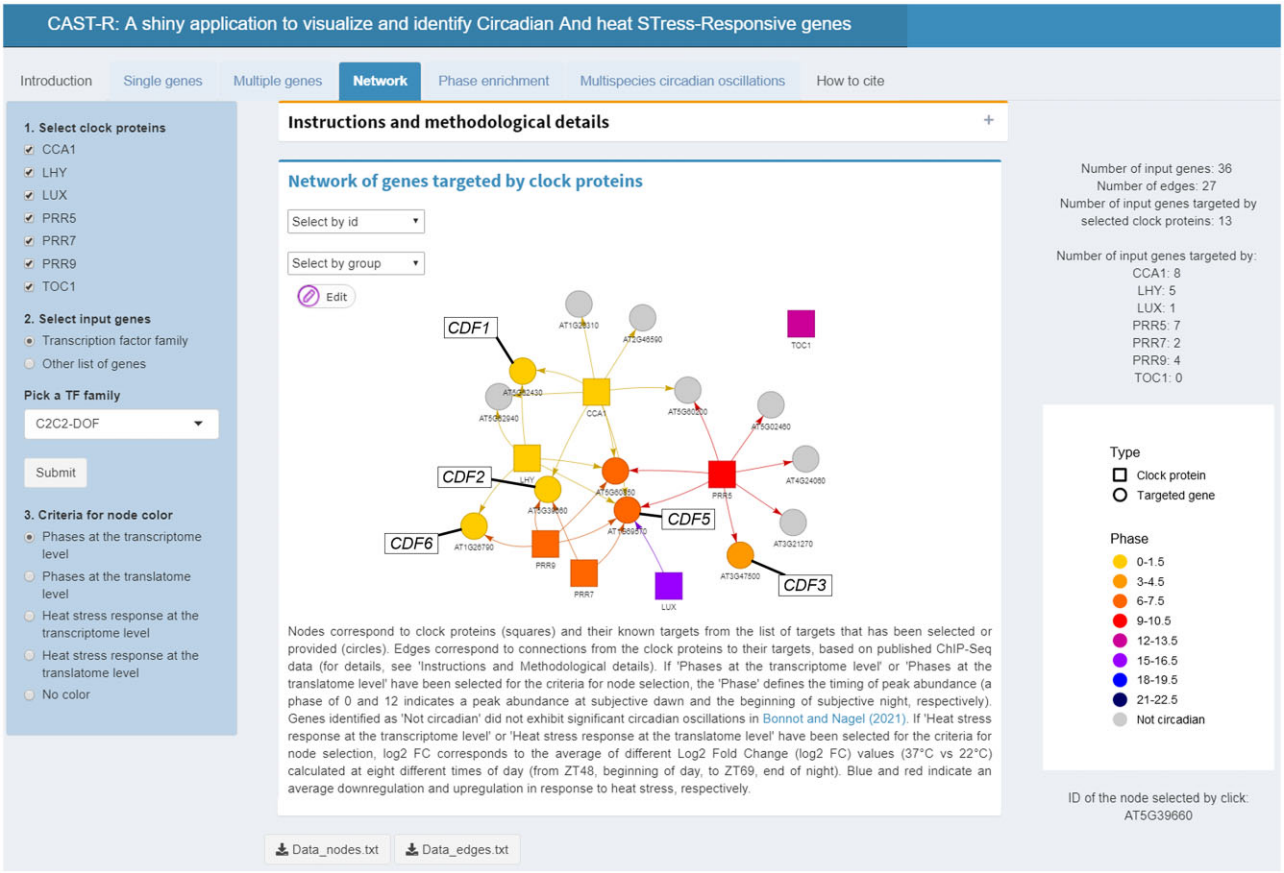


Figure 3 Interactive network of genes targeted by clock proteins. The "C2C2-DOF' TF family is represented as an example. Members of the *CDF* subfamily are highlighted and have been manually annotated for figure preparation. Users can either select a TF family or paste a list of Arabidopsis AGI codes to generate a network. Nodes correspond to clock proteins (squares) and their known targets from the list of input genes (circles). Edges from clock proteins to target genes were identified in published ChIP-Seq data (see Table 1). Nodes are colored based on the selected criteria (i.e. the phases at the transcriptome or translatome level, the heat stress response at the transcriptome or translatome level). Here, "Phases at the transcriptome level" is selected.

by clock proteins. For example, 13 members of the "C2C2-DOF" family are direct targets of clock proteins, including five *CDFs* (Figure 3). The generated network is interactive and allows users to move nodes and to add new nodes and connections in the network. Specific nodes and their connections can be highlighted in the network by directly pointing or clicking on the nodes, or by scrolling in the list of IDs and selecting a gene.

To connect this network representation with the information provided in the "Single genes" and "Multiple genes" tabs, we implemented an option to color nodes based on their phase or on their heat stress response, from our published transcriptome and translatome datasets (Bonnot and Nagel, 2021). On the graphical network visualization, an option allows to highlight genes by group (i.e. timing of expression/translation or level of downregulation or upregulation under heat stress). Of note, the gene IDs for clock proteins and targets are shown under the "Select by id" drop down box. Hovering on a node also provides its function, assigned to each locus ID using BioMart (https://plants.ensembl.org/biomart/martview/). In addition, clicking on a node makes accessible its AGI to copy and paste in the "Single gene" module for example.

Module 4—Phase enrichment analysis

The "Phase enrichment" tab combines statistical analysis and graphical representations to identify and represent under- and over-represented timing of peak expression (phase) within a list of genes. To do so, users need to select an existing reference (Table 2), which corresponds to all circadian transcripts that were identified in published datasets, or to provide their own reference. Existing references provided in CAST-R were established from Arabidopsis seedlings grown in free-running conditions (constant light and temperatures, necessary to identify circadian oscillations) after entrainment with thermocycles and/or photocycles (Table 2). For example, the reference "Bonnot and Nagel_Transcriptome" corresponds to all 8,028 circadian transcripts identified at the transcriptome level by Bonnot and Nagel (2021). Users then need to provide a list of AGI codes of interest, to see if any phases are enriched within this group. CAST-R will generate three plots and a table. Two circular bar plots represent the

Table 2 Description of the datasets that can be used in the application as the reference for the phase enrichment analysis

Dataset	Reference name	Tool used to detect oscillations	Cutoff	Number of circadian transcripts	References
1	Bonnot and Nagel_Transcriptome_LL_LDHH	Metacycle (combining JTK_CYCLE and Lomb-Scargle)	BH.Q < 0.01 and transcripts with 0.01 < BH.Q < 0.05 and overlapping with circadian transcripts in datasets 3–8	8,028	Bonnot and Nagel (2021)
2	Bonnot and Nagel Translatome_LL_LDHH			10,657	
4	Michael et al_LL_LDHC	Phaser	Correlation > 0.7	8,909	Michael et al. (2008)
5	Michael et al_LL_LLHC			7,955	
6	Covington and Harmer_LL_LDHH			7,858	Covington and Harmer (2007)
7	Edwards et al_LL_LDHH			9,940	Edwards et al. (2006)
8	Hsu and Harmer LL_LDHH	COSOPT JTK_CYCLE	pMMCβ < 0.05 $P < 0.01$ and $q < 0.05$	7,124	Hsu and Harmer (2012)
9	Romanowski_LL_LDHH	JTK_CYCLE	$P < 0.01$ and $q < 0.01$	9,128	Romanowski et al. (2020)

phase distribution within the selected reference and within the pasted list of AGI codes, respectively (Figure 4). Within the user subset, only circadian genes are used for this analysis. CAST-R then compares the proportions of each phase between the reference and the user subset of genes. Under-represented (fold enrichment < 1) and over-represented (fold enrichment > 1) phases in the subset of genes are represented on a circular dot plot. Significant differences are assessed at $P < 0.05$ using Chi-squared tests and significant phases are highlighted on the dot plot and in a table that summarizes the analysis (Figure 4).

As lists of circadian transcripts can substantially differ between published datasets, we highly encourage the users to perform the analysis using different references and to compare the results. The rapidity of the analysis makes it very useful to determine if a pattern of peak expression is particularly represented in a list of genes of interest. In addition, this phase enrichment functionality is very flexible and not restricted to Arabidopsis nor to the references that are provided. We provide an option that allows users to use their own reference (not listed or not published yet) of circadian transcripts and perform the same analysis (formatting details are specified when "Use your own reference" is selected). Moreover, as long as a rhythmic omics dataset is provided as a reference, and the information of phase for each measured variable is provided, phase enrichment can be performed. Thus, phase enrichment analyses with CAST-R can be performed in multiple organisms other than plants and with diverse types of omics data.

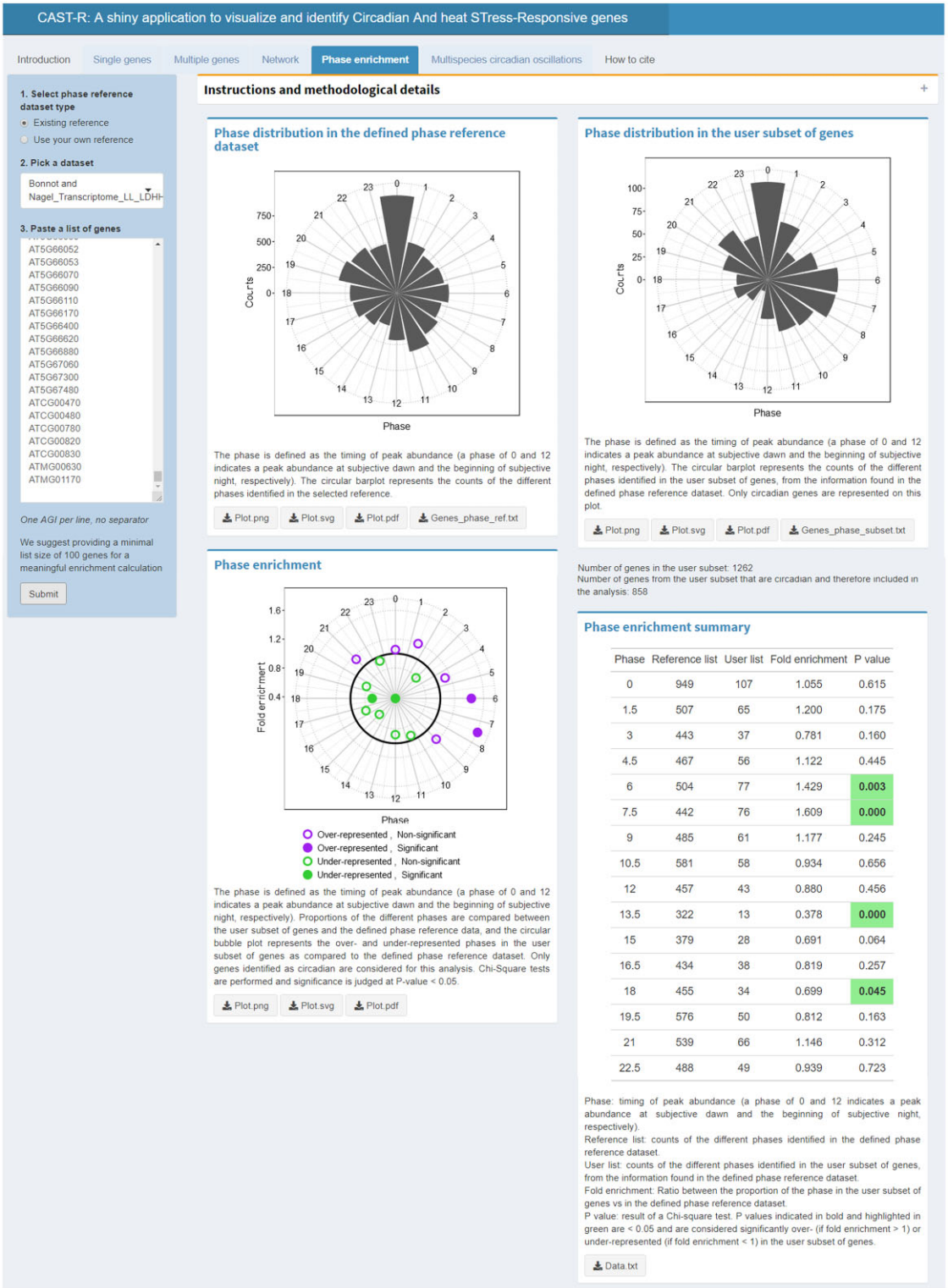
Module 5—Comparison of circadian oscillations between Arabidopsis and rice orthologs

In the last tab of CAST-R, entitled "Multispecies circadian oscillations", users can plot the transcript abundance of individual genes in Arabidopsis, brassica, barley, and rice over different time courses performed in free-running conditions (Edwards et al., 2006; Covington and Harmer, 2007; Mockler et al., 2007; Michael et al., 2008; Filichkin et al., 2011; Greenham et al., 2020; Müller et al., 2020; Romanowski et al., 2020; Bonnot and Nagel, 2021). The aim of this CAST-R functionality is to propose a multi-view tab where users can plot multiple gene expression profiles in different plant species. If the orthologous genes are known, this allows them to rapidly look for potentially conserved profiles between species. For example, *AtLHY* (*AT1G01060*) peaks at ZT0 and ZT24 in the "Diurnal LL_LDHC" Arabidopsis dataset, and *BrLHY* (*BraA10g01800R*) peaks at ZT24, ZT46, and ZT68 in the "Greenham LL_LLHC" Brassica dataset (Figure 5).

We emphasize that this CAST-R functionality does not indicate if the genes are significantly cycling or not. To validate circadian oscillations of specific genes, we encourage users to either check the lists of identified circadian genes in the corresponding studies, and select a significance cutoff, or to download the data from CAST-R and to perform a detection of rhythmicity using appropriate tools such as Metacycle (Wu et al., 2016). This "Multispecies circadian oscillations" functionality also offers the possibility to compare the expression patterns of individual genes between datasets, within the same species. Although the provided datasets were obtained from similar conditions (i.e. light and temperature conditions), slight differences in experimental design (seedling age, media composition, etc.) might lead to differences in expression rhythmicity.

Summary and future directions

Through the diverse modules proposed in the first version of CAST-R, users can determine if their Arabidopsis genes of interest have been identified as circadian and heat stress-responsive genes. The main objective is to help biologists formulate new hypotheses on the regulation of their candidate genes by considering the influence of the time of day and the clock. Importantly, circadian expression and time of day heat responses of individual genes can be compared with other members within a gene family or genes that



Phase	Reference list	User list	Fold enrichment	P value
0	949	107	1.055	0.615
1.5	507	65	1.200	0.175
3	443	37	0.781	0.160
4.5	467	56	1.122	0.445
6	504	77	1.429	**0.003**
7.5	442	76	1.609	**0.000**
9	485	61	1.177	0.245
10.5	581	58	0.934	0.656
12	457	43	0.880	0.456
13.5	322	13	0.378	**0.000**
15	379	28	0.691	0.064
16.5	434	38	0.819	0.257
18	455	34	0.699	**0.045**
19.5	576	50	0.812	0.163
21	539	66	1.146	0.312
22.5	488	49	0.939	0.723

Figure 4 Phase enrichment analysis and visualization. Here, a list of 1,262 genes differentially regulated under heat stress is used as an example. The list of 8,028 circadian transcripts identified at the transcriptome level in Bonnot and Nagel (2021) is used as the reference ("Bonnot and Nagel_Transcriptome_LL_LDHH"). The phase distribution (i.e. the number of transcripts per phase) is represented in this reference and in the subset of genes provided by the user. In this example, 858 of the 1,262 genes provided by the user overlap with the selected reference (and therefore exhibit circadian oscillations at the transcriptome level) and are represented in the box "Phase distribution in the user subset of genes". The phase enrichment dot plot shows the over- (purple) and under-represented (green) phases in the subset of 858 genes as compared to the 8,028 genes in the reference. In this example, four phases are significantly differentially represented (filled circles, $P < 0.05$, Chi-squared test) in the subset of 858 genes: 6 and 7.5 (over-represented, fold enrichment > 1) and 13.5 and 18 (under-represented, fold enrichment < 1). A summary of the phase distribution in the reference and the user subset of genes is indicated in the "Phase enrichment summary" table, as well as the fold enrichment and the *P*-value for each individual phase.

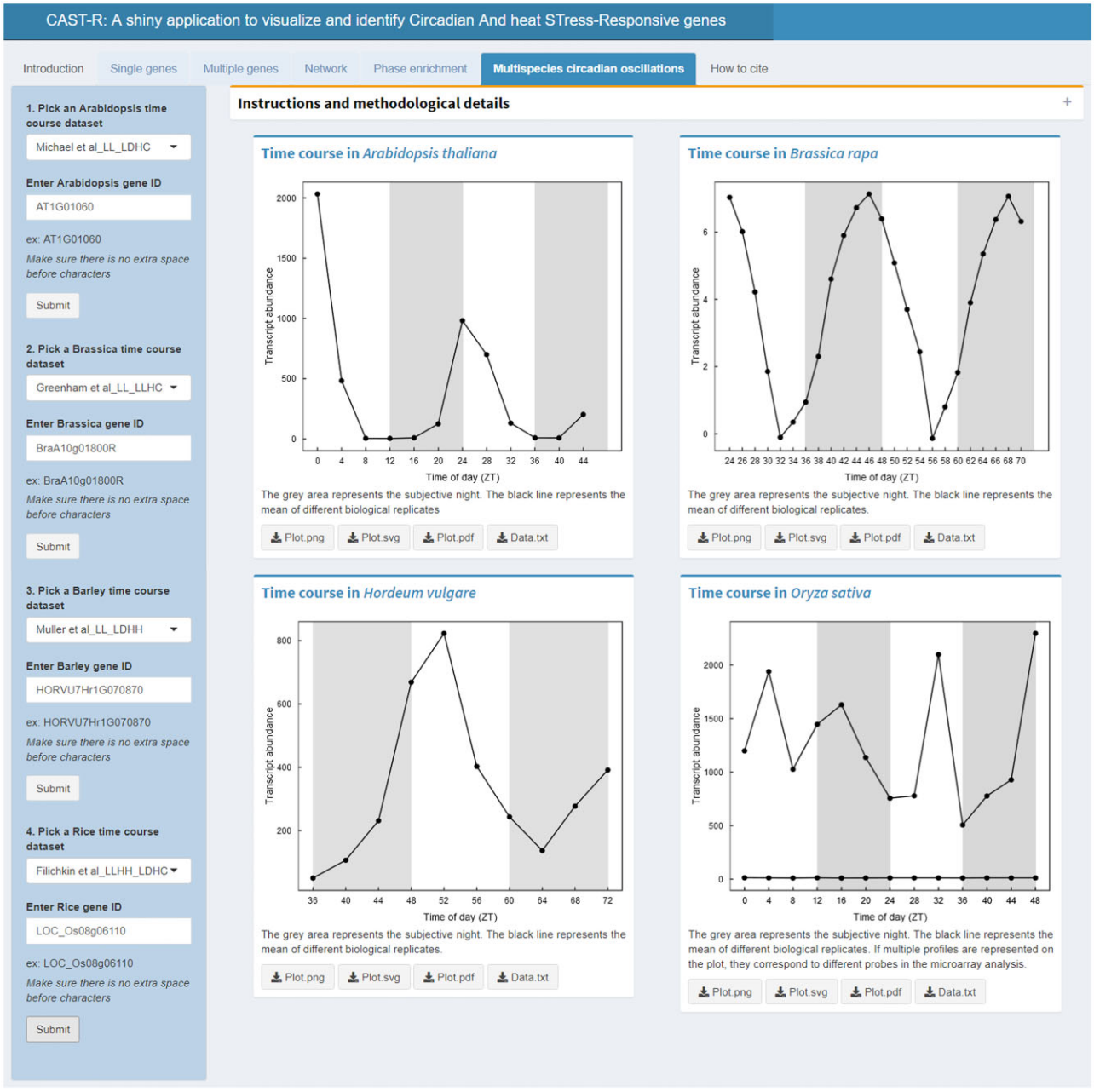


Figure 5 Multispecies circadian oscillations. Here, *LHY* is represented as an example. The conditions "Michael et al_LL_LDHC", "Greenham et al_LL_LLHC", "Muller et al_LL_LDHH", and "Filichkin et al_LLHH_LDHC" are used for the *Arabidopsis (A. thaliana)*, *brassica (B. rapa)*, barley (*H. vulgare)*, and rice (*O. Sativa)* time course datasets, respectively. Two profiles are represented on the plot in the 'Time course in *Oryza sativa*' box, corresponding to different probes in the microarray analysis (for details, see Filichkin et al., 2011).

function in a similar biological pathway. By producing plots that can easily be exported in multiple formats, CAST-R also represents a useful platform to generate publication-quality figures. For example, phase enrichment circular plots have been used in our recent paper (Bonnot and Nagel, 2021), and can be produced with CAST-R from data obtained in any organism, as long as a rhythmic dataset is provided as a reference. Future developments will aim at improving comparisons between conditions and plant species. They will include rhythmic datasets obtained in diverse plants in addition to Arabidopsis, brassica, barley, and rice present in this first version, and in both free-running and diurnal conditions. In the network analysis, other connections determined in DNA Affinity Purification-Sequencing or interactome data could be added to build more complete gene regulatory networks. In the long term, CAST-R may represent a major platform to identify clock regulation of abiotic stress responses and conservation between plant species.

Materials and methods

The first version of CAST-R was built from R (R Core Team, 2020) with the package "Shiny" (Chang et al., 2020) and was tested on Firefox 89.0.2, Google Chrome 91.0.4472.106, Opera 77.0.4054.90, and Safari 14.1.1 on Windows 10 home, Mac OS Catalina and Pop!_OS 20.10. The large heatmaps are generated using the R package "pheatmap" (Kolde, 2019). The interactive network is built with the R package "visNetwork" (Almende and Thieurmel, 2019). All other plots are made using the R package "ggplot2" (Wickham, 2016).

All datasets used in the development and design of this web application were described previously (Edwards et al., 2006; Covington and Harmer, 2007; Mockler et al., 2007; Michael et al., 2008; Filichkin et al., 2011; Hsu and Harmer, 2012; Huang et al., 2012; Nakamichi et al., 2012; Liu et al., 2013; Pruneda-Paz et al., 2014; Nagel et al., 2015; Kamioka et al., 2016; Liu et al., 2016; Ezer et al., 2017; Adams et al., 2018; Romanowski et al., 2020; Greenham et al., 2020; Müller et al., 2020; Bonnot and Nagel, 2021). Within each CAST-R tab, an "Instructions and methodological details" window gives information about these datasets and provides links to the corresponding papers.

In the 'Phase enrichment" tab, proportions of phases within the user subset of genes are compared with those of the selected reference (i.e. all circadian genes identified in the corresponding study). The fold enrichment corresponds to the ratio between the proportion of the phase in the user subset of genes versus the proportion of the same phase in the selected reference. A fold enrichment <1 and >1 indicates an under-representation and an over-representation of the phase in the user subset of genes as compared to the reference, respectively. Chi-square tests are then performed and significance is judged at $P < 0.05$. For this analysis, as mentioned on the control panel of the application, we suggest to provide a minimal list size of 100 genes for a meaningful enrichment calculation.

Accession numbers

CDF3, AT3G47500; AtLHY, AT1G01060; BrLHY, BraA10g01800R; CCA1, AT2G46830; PRR5, AT5G24470; PRR7, AT5G02810; PRR9, AT2G46790, LUX, AT3G46640; TOC1, AT5G61380.

Acknowledgments

We thank members of the Nagel lab, Emily J Blair and Sabrina Gilmour for constructive feedback and discussions about the Shiny App. Thanks to both Dr Julia Bailey-Serres and Nagel lab members for beta testing the Shiny App and for assistance in improving the usability of the application.

Funding

This work was supported by an National Science Foundation Division of Integrative Organismal Systems Early Career Award #1942949 to D.H.N.

Conflict of interest statement. The authors declare no competing interests.

References

Adams S, Grundy J, Veflingstad SR, Dyer NP, Hannah MA, Ott S, Carré IA (2018) Circadian control of abscisic acid biosynthesis and signalling pathways revealed by genome-wide analysis of LHY binding targets. New Phytol **220**: 893–907

Almende BV, Thieurmel BTR (2019) Vis network: network visualization using "vis.js" library. R package version, 2(9)

Bonnot T, Nagel DH (2021) Time of day prioritizes the pool of translating mRNAs in response to heat stress. Plant Cell **33**: 2164–2182

Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2020) Shiny: web application framework for R. R package Version 1.4.0.

Covington MF, Harmer SL (2007) The circadian clock regulated auxin signaling and responses in Arabidopsis. PLoS Biol **5**: e222

de los Reyes P, Romero-Losada AB, Romero-Campero FJ (2020) ATTRACTOR, Arabidopsis thaliana transcriptional circadian network v1.0. doi: 10.5381/zenodo.3780022

Edwards KD, Anderson PE, Hall A, Salathia NS, Locke JCW, Lynn JR, Straume M, Smith JQ, Millar AJ (2006) FLOWERING LOCUS C mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. Plant Cell **18**: 639–650

Ezer D, Jung JH, Lan H, Biswas S, Gregoire L, Box MS, Charoensawan V, Cortijo S, Lai X, Stöckle D, et al. (2017) The evening complex coordinates environmental and endogenous signals in *Arabidopsis*. Nat Plants **3**: 17087

Farré EM (2012) The regulation of plant growth by the circadian clock. Plant Biol **14**: 401–410

Ferrari C, Proost S, Janowski M, Becker J, Nikoloski Z, Bhattacharya D, Price D, Tohge T, Bar-Even A, Fernie A, et al. (2019) Kingdom-wide comparison reveals the evolution of diurnal gene expression in Archaeplastida. Nat Commun **10**: 737

Filichkin SA, Breton G, Priest HD, Dharmawardhana P, Jaiswal P, Fox SE, Michael TP, Chory J, Kay SA, Mockler TC (2011) Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. PLoS One **6**: e16907

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

Greenham K, Sartor RC, Zorich S, Lou P, Mockler TC, McClung CR (2020) Expansion of the circadian transcriptome in *Brassica rapa* and genome-wide diversification of paralog expression patterns. Elife **9**: 1–26

Grundy J, Stoker C, Carré IA (2015) Circadian regulation of abiotic stress tolerance in plants. Front Plant Sci **6**: 648

Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA (2000) Orchestrated Transcription of Key Pathways in Arabidopsis by the Circadian Clock. Science **290**: 2110–2113

Hsu PY, Harmer SL (2012) Circadian phase has profound effects on differential expression analysis. PLoS One **7**: e49853

Huang W, Pérez-García P, Pokhilko A, Millar AJ, Antoshechkin I, Riechmann JL, Mas P (2012) Mapping the core of the *Arabidopsis* circadian clock defines the network structure of the oscillator. Science **336**: 75–79

Kamioka M, Takao S, Suzuki T, Taki K, Higashiyama T, Kinoshita T, Nakamichi N (2016) Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED1 in the *Arabidopsis* circadian clock. Plant Cell **28**: 696–711

Kolde R (2019) Pheatmap: Pretty Heatmaps. *R Packag. version 1.0*, *8*.

Lai X, Bendix C, Yan L, Zhang Y, Schnable JC, Harmon FG (2020) Interspecific analysis of diurnal gene regulation in panicoid grasses identifies known and novel regulatory motifs. BMC Genomics **21**: 428

Liu T, Carlsson J, Takeuchi T, Newton L, Farré EM (2013) Direct regulation of abiotic responses by the *Arabidopsis* circadian clock component PRR7. Plant J **76**: 101–14

Liu TL, Newton L, Liu MJ, Shiu SH, Farré EM (2016) A G-box-like motif is necessary for transcriptional regulation by circadian pseudo-response regulators in Arabidopsis. Plant Physiol **170**: 528–539

Michael TP, Mockler TC, Breton G, McEntee C, Byer A, Trout JD, Hazen SP, Shen R, Priest HD, Sullivan CM, et al. (2008) Network discovery pipeline elucidates conserved time-of-day–specific cis-regulatory modules. PLoS Genet **4**: e14

Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay SA, Chory J (2007) The diurnal project: diurnal and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harb Symp Quant Biol **72**: 353–363

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

Nagel DH, Doherty CJ, Pruneda-Paz JL, Schmitz RJ, Ecker JR, Kay SA (2015) Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis. Proc Natl Acad Sci USA **112**: E4802–E4810

Nakamichi N, Kiba T, Kamioka M, Suzuki T, Yamashino T, Higashiyama T, Sakakibara H, Mizuno T (2012) Transcriptional repressor PRR5 directly regulates clock-output pathways. Proc Natl Acad Sci USA **109**: 17123–17128

Pruneda-Paz JL, Breton G, Nagel DH, Kang SE, Bonaldi K, Doherty CJ, Ravelo S, Galli M, Ecker JR, Kay SA (2014) A Genome-scale resource for the functional characterization of *Arabidopsis* transcription factors. Cell Rep **8**: 622–632

R Core Team (2020) R: A Language and Environment for Statistical Computing. R Core Team, Vienna, Austria

Romanowski A, Schlaen RG, Perez-Santangelo S, Mancini E, Yanovsky MJ (2020) Global transcriptome analysis reveals circadian control of splicing events in *Arabidopsis thaliana*. Plant J **103**: 889–902

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4

Wu G, Anafi RC, Hughes ME, Kornacker K, Hogenesch JB (2016) MetaCycle: an integrated R package to evaluate periodicity in large scale data. Bioinformatics **32**: 3351–3353