**Title:** Robustness of gene expression rhythmicity can be used to identify new oscillator components

**Authors**: Titouan Bonnot[1*], Desmond Cairo[2], and Dawn H Nagel[2*]

**Affiliation:** Agroécologie, INRAE, Institut Agro, Univ. Bourgogne, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France; Department of Botany and Plant Sciences, University of California, Riverside, Riverside, California 92507.

**Corresponding authors:** titouan.bonnot@inrae.fr and dawnn@ucr.edu

## Methodology

*Rhythmicity Analysis:* Results of the JTK_CYCLE and MetaCycle analyses performed in (1) (Romanowski *et al.*, 2020) and (2) (Bonnot and Nagel, 2021), respectively, were collected. BH.Q (1) and meta2d_BH.Q (2) values obtained for the 9127 and 8028 genes identified as rhythmic in (1) and (2), respectively, were selected and transformed using -log10. In (2), nine genes exhibited a meta2d_BH.Q value equal to 0, thus corresponding to an infinite value after -log10 transformation. For these genes, the -log10 value was set at 13 to facilitate comparison with other rhythmic genes. For Figure S1, amplitudes of oscillation correspond to the amplitude calculated by JTK_CYCLE, for both (1) and (2). TFs were selected according to the list of TFs and TF families described in (Pruneda-Paz *et al.*, 2014). Members of the BBX subfamily were identified from the TAIR website (https://www.arabidopsis.org/). The Venn diagram and heatmap were generated with the 'eulerr' and 'pheatmap' R packages, respectively (Kolde, 2018; Larsson, 2024). The gene network was visualized using the CYTOSCAPE software v. 3.10.1 (Smoot *et al.*, 2011). Other plots were generated with the R package 'ggplot2' (Wickham, 2016).

*Expression Data Reanalysis:* For this, we downloaded raw RNA sequencing data for misexpression lines of eight TFs without discretion for experimental conditions and samples. The raw FASTQ files were quality checked for adaptor contamination using the FastQC package and trimmed using the Fastp package if contamination was present (Chen *et al.*, 2018). The cleaned files were then aligned to the TAIR10 genome assembly using the HISAT2 alignment algorithm and further processed with SAM tools (Li *et al.*, 2009; Kim *et al.*, 2019; Yates *et al.*, 2022). To maintain consistency in data processing, double-stranded sequences were aligned using only the forward strand, which did not significantly affect alignment success. Count files were generated using summarizeOverlaps from the R package GenomicAlignments (Lawrence *et al.*, 2013). Genes with counts across samples less than 1 were filtered out, differential expression analysis was performed using the systemPipeR package (Backman and Girke, 2016). To account for the differences in experiment design across samples, comparisons were treated as independent, considering only experimental changes at the transcriptome level from WT to misexpressed lines for individual studies (Supplementary Dataset S1). Data were then filtered for differential expression of selected clock genes (Figure 1).

**Data availability**

The datasets used for rhythmicity analysis are: (1), Romanowski *et al.,* 2020, (Table S3) and (2), Bonnot and Nagel 2021, (Supplemental Data Set S2). ChIP-seq datasets were obtained from (Adams *et al.,* 2018; Nagel *et al.,* 2015; Liu *et al.,* 2013, 2016; Kamioka *et al.,* 2016; Ezer *et al.,* 2017; Nakamichi *et al.,* 2012; Huang *et al.,* 2012) and DAP-Seq datasets were obtained from O'Malley *et al.,* 2016.

The datasets used for gene expression reanalysis are: (1), Wildtype (WT, Col-0) and *35S:TZF1,* mock-treated samples (https://bigd.big.ac.cn/gsa/browse/CRA015388; (He *et al.*, 2024). (2), WT (Col-0) and *bbx28-5* (GSE217723), white light (WL), and simulated shade (SS) samples (Saura-Sánchez *et al.*, 2024). (3 and 4), WT (Col-0), *35S:PIF4* and *35S:PIF5* (GSE59699), mock-treated 16h samples, and low blue light (LBL) samples (Pedmale *et al.*, 2016). (5), WT (Col-0) and *pif4-101* mutant (GSE108162) (Ding *et al.*, 2018). (6 - 8), WT (Col-0) empty vector control (EV), *p35S:GR-CDF1*, *p35S:GR-HB6*, *p35S:GR-HYH*, and *p35S:GR-RAV1* (GSE117857), dexamethasone treated samples (Brooks *et al.*, 2019).

**References:**


**Backman TWH, Girke T**. 2016. systemPipeR: NGS workflow and report generation environment. BMC Bioinformatics **17**, 388.

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

**Brooks MD, Cirrone J, Pasquino AV, *et al.*** 2019. Network Walking charts transcriptional dynamics of nitrogen signaling by integrating validated and predicted genome-wide interactions. Nature Communications **10**, 1569.

**Chen S, Zhou Y, Chen Y, Gu J**. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**, i884–i890.

**Ding L, Wang S, Song Z-T, Jiang Y, Han J-J, Lu S-J, Li L, Liu J-X**. 2018. Two B-Box Domain Proteins, BBX18 and BBX23, Interact with ELF3 and Regulate Thermomorphogenesis in Arabidopsis. Cell Reports **25**, 1718-1728.e4.

**He S-L, Li B, Zahurancik WJ, Arthur HC, Sidharthan V, Gopalan V, Wang L, Jang J-C**. 2024. Overexpression of stress granule protein TZF1 enhances salt stress tolerance by targeting ACA11 mRNA for degradation in Arabidopsis. Frontiers in Plant Science **15**, 1375478.

**Kim D, Paggi JM, Park C, Bennett C, Salzberg SL**. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology **37**, 907–915.

**Kolde R**. 2018. pheatmap: Pretty Heatmaps. R package version 1.0.12.

**Larsson J**. 2024. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses.

**Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ**. 2013. Software for Computing and Annotating Genomic Ranges. (A Prlic, Ed.). PLoS Computational Biology **9**, e1003118.

**Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup**. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**, 2078–2079.

**Pedmale UV, Huang SC, Zander M, *et al.*** 2016. Cryptochromes Interact Directly with PIFs to Control Plant Growth in Limiting Blue Light. Cell **164**, 233–245.

**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 Reports **8**, 622–632.

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

**Saura-Sánchez M, Gomez-Ocampo G, Pereyra ME, Barraza CE, Rossi AH, Córdoba JP, Botto JF**. 2024. B-Box transcription factor BBX28 requires CONSTITUTIVE PHOTOMORPHOGENESIS1 to induce shade-avoidance response in *Arabidopsis thaliana*. Plant Physiology **195**, 2443–2455.

**Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T**. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics **27**, 431–432.

**Wickham H**. 2016. *ggplot2*. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.

**Yates AD, Allen J, Amode RM, *et al.*** 2022. Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Research **50**, D996–D1003.