**LARGE-SCALE BIOLOGY ARTICLE**

**Multiomic responses to seasonal and diurnal cycles in the marine picoeukaryote *Ostreococcus tauri***

**Short title:** Seasonal and diurnal cycles in Ostreococcus

Ana B. Romero-Losadaa,b, Christina Arvanitidoua,b, María Morales-Pinedaa, M. Elena García-Gómeza, Mercedes García-Gonzáleza,1, Francisco J. Romero-Camperoa,b,1

a Institute for Plant Biochemistry and Photosynthesis, Universidad de Sevilla – Consejo Superior de Investigaciones Científicas, Av Américo Vespucio 49, 41092 Seville (Spain)

b Department of Computer Science and Artificial Intelligence, Universidad de Sevilla, Av Reina Mercedes s/n, 41012 Seville (Spain)

1 Corresponding Authors, email [fran@us.es](mailto:fran@us.es) , [mggonza@us.es](mailto:mgg@us.es)

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 (www.plantcell.org) is: Francisco J. Romero-Campero ([fran@us.es](mailto:fran@us.es))

**ABSTRACT**

**INTRODUCTION**

To anticipate periodic changes induced by Earth rotation and the annual translation around the sun, many organisms have evolved and endogenous clock that coordinate biological processes including cell division, photosynthesis and protein synthesis, called circadian rhythms

**RESULTS**

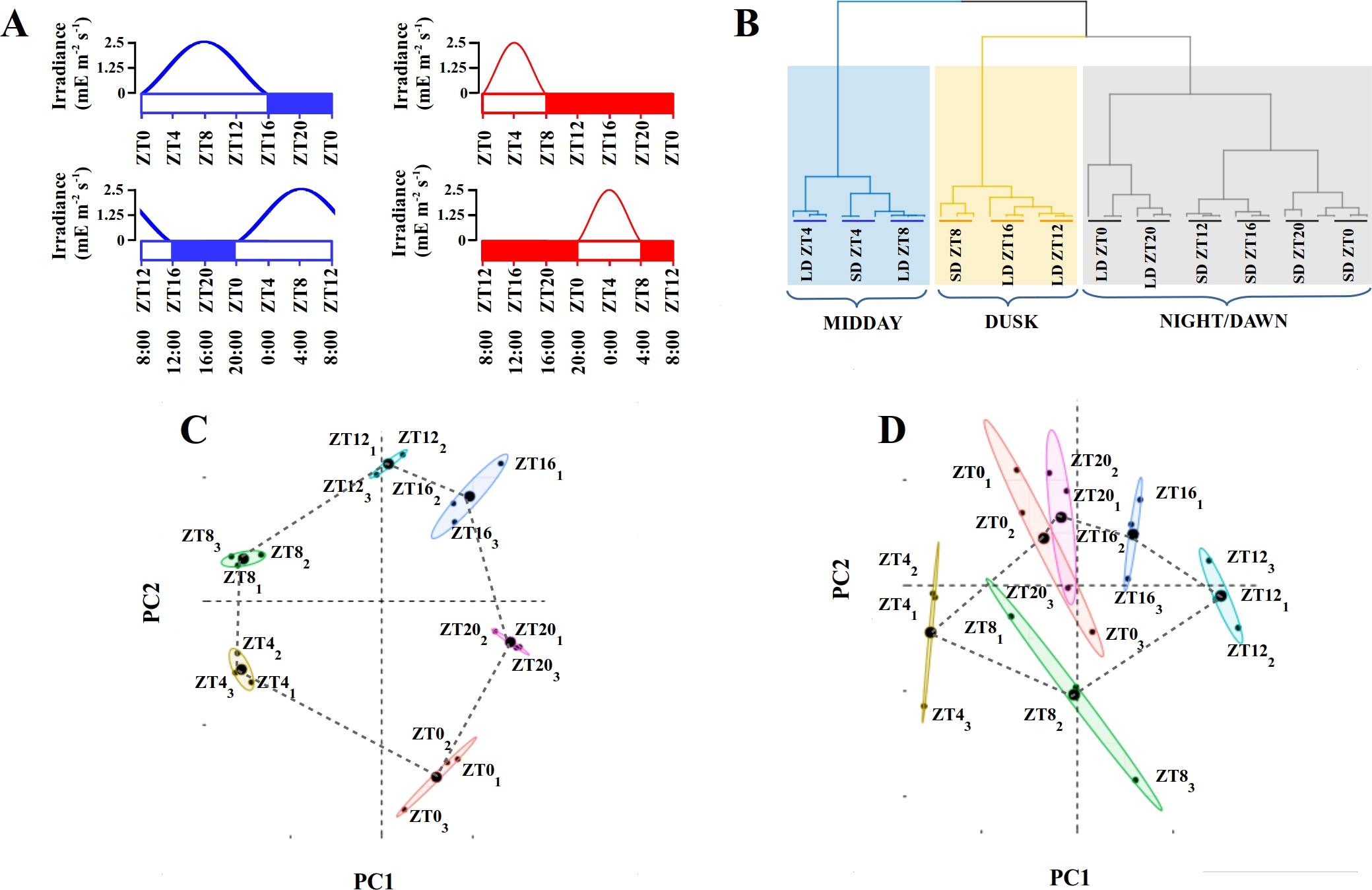
**An optimized and automatically controlled cultivation system ensures a high level of circadian synchronization**

Culture synchrony was optimized in our experimental design to ensure an accurate identification of genes exhibiting diurnal rhythmic expression patterns under distinct photoperiods corresponding to different seasons. In order to achieve this, we optimized a culturen system consisting of photochemostats operating in continuous regime. These photochemostats were controlled by accurate automatic devices regulating temperature, pH, , flow rate of media addition and irradiance (Supplemental Figure1) (del Río et al., 2015). They are bubble columns with an external jacket where flowing water from a cooling device, maintaining the culture temperature exactly at 20ºC. Culture homogenization was ensured by continuous air sparging from the bottom of the column. On demand, injection of CO2 into the air stream entering the culture precisely maintained pH at 8.0. Cell density and nutrient availability was maintained at stationary state by the continuous addition of fresh media to the photochemostats using a peristaltic pump at a rate of 45 mL h-1 during the light period (dilution rate 0.3 d-1). Excess culture volume was removed by the overflow at the same rate. *Ostreococcus tauri* cultures were grown under long day (LD) condition (16 hours light: 8 hours dark), simulating a summer day, and under short day (SD) condition (8 hours light: 16 hours dark), simulating a winter day. Instead of abrupt dark-to-light and light-to-dark transitions, progressive increase and decrease incident light irradiance were used to simulate solar daylight cycles as described in the method sections, Figure 1A. For each experiment, two different photochemostats were used with a time offset of 12 hours in the illumination program. Samples were collected every four hours and labeled as Zeitgeber time (ZT) N indicating the time point N hours after lights are switched on simulating dawn. The time offset between the two photochemostats allowed us to collect samples ZT0/ZT12, ZT4/ZT16 and ZT8/ZT20 simultaneously at 8:00, 12:00 and 16:00 from the first and second photochemostat respectively. In this way we completed the sample collection corresponding to an entire day/night cycle in eight hours, Figure 1A.

Samples collection was performed during three consecutive days under either LD or SD conditions. At the end of the third day the illumination program was set to continuous light (LL) in half of the chemostats and continuous darkness (DD) in the other half. During the next two days samples were collected under these free running conditions (Supplemental Figure2). In overall 84 samples were generated, 18 samples corresponding to three days under LD conditions, 12 samples under LL after LD, 12 samples under DD after LD, 18 samples corresponding to three days under SD conditions, 12 samples under LL after SD and 12 samples under DD after SD. Additionally, in this study we reanalyze microarray data previously published (Monnier et al., 2010) generated under neutral day (ND) conditions (12 hours light: 12 hours dark) that could be used to simulate a spring or autumn day.

As described in the methods section, RNA was extracted, purified and processed for high-throughput transcriptome sequencing producing approximately 10 million short reads per sample (Supplemental Table 1). This allowed us to accurately estimate gene expression levels measured as FPKM (Fragments Per Kilobase of exon per Million reads mapped) in the transcriptomes corresponding to each data point of our time series (Supplemental Table 2). Indeed, out of the 7668 genes currently annotated in the *Ostreococcus tauri* genome (Palenik et al., 2007; Blanc-Mathieu et al., 2014), only 260 genes never exceeded an expression level of ten FPKM. This shows that practically the entire *Ostreococcus tauri* genome is expressed under seasonal and diurnal cycles. First, we focus in the 36 transcriptomes corresponding to the time points taken during three days under LD and SD conditions and perform a hierarchical clustering analysis. The transcriptomes corresponding to the same time points during the three different days tend to cluster together, Figure 1B. This indicates a high circadian synchronization in our cultures. Moreover, these 36 transcriptomes assemble together into three different groups, Figure 1B. The first cluster corresponds to midday. The transcriptomes at time points ZT4 and ZT8 under LD and ZT4 under SD constitute this cluster. These time points correspond to the moments of maximal incident light irradiance under both LD and SD conditions, Figure 1A. The second cluster conforms the dusk group. Here the transcriptomes at time points ZT12 and ZT16 under LD and ZT8 under SD are found. These time points coincide with the end of the light period in both LD and SD conditions when incident light irradiance is low. The third cluster represents night/dawn and comprises the transcriptomes at time points ZT20, ZT0 under LD and ZT12, ZT16, ZT20 and ZT0 under SD. The transcriptomes at time points in the LD and SD nights or dark periods constitute two distinct groups suggesting noticeable differences in the transcriptomic responses during the night under LD and SD conditions. It is also noteworthy the higher similarity between the dusk, night/dawn transcriptomes when compare to the midday one, Figure 1B. In order to obtain a deeper understanding of the underlying structure in our data we performed principal components analysis separately over the LD and SD transcriptomes, Figure 1C and 1D respectively. Under LD conditions, we observed that the transcriptomes corresponding to the same time point in the three different days tightly cluster together globally constituting a circular structure. Nonetheless, under SD conditions more variability is observed and the time point transcriptomes form a structure resembling an ellipse. This could indicate that whereas in LD conditions gene expression in globally cycling precisely with a similar period a more complex behavior is expected under SD conditions. Also, it is remarkable the high similarity between the transcriptomes corresponding to ZT0 and ZT20 under SD conditions that is not present under LD conditions. This suggests that the transcriptomic response at the end of a SD night is already preparing all molecular systems for the incoming light availability at dawn whereas this anticipation is not as clearly observed under LD conditions. In overall, these results support that our experimental design grants a high level of synchronization in our data allowing us to proceed to the identification and comparison of genes exhibiting rhythmic expression patterns under LD and SD conditions.

**Figure 1.** Experimental design and circadian synchronization in our cell cultures.



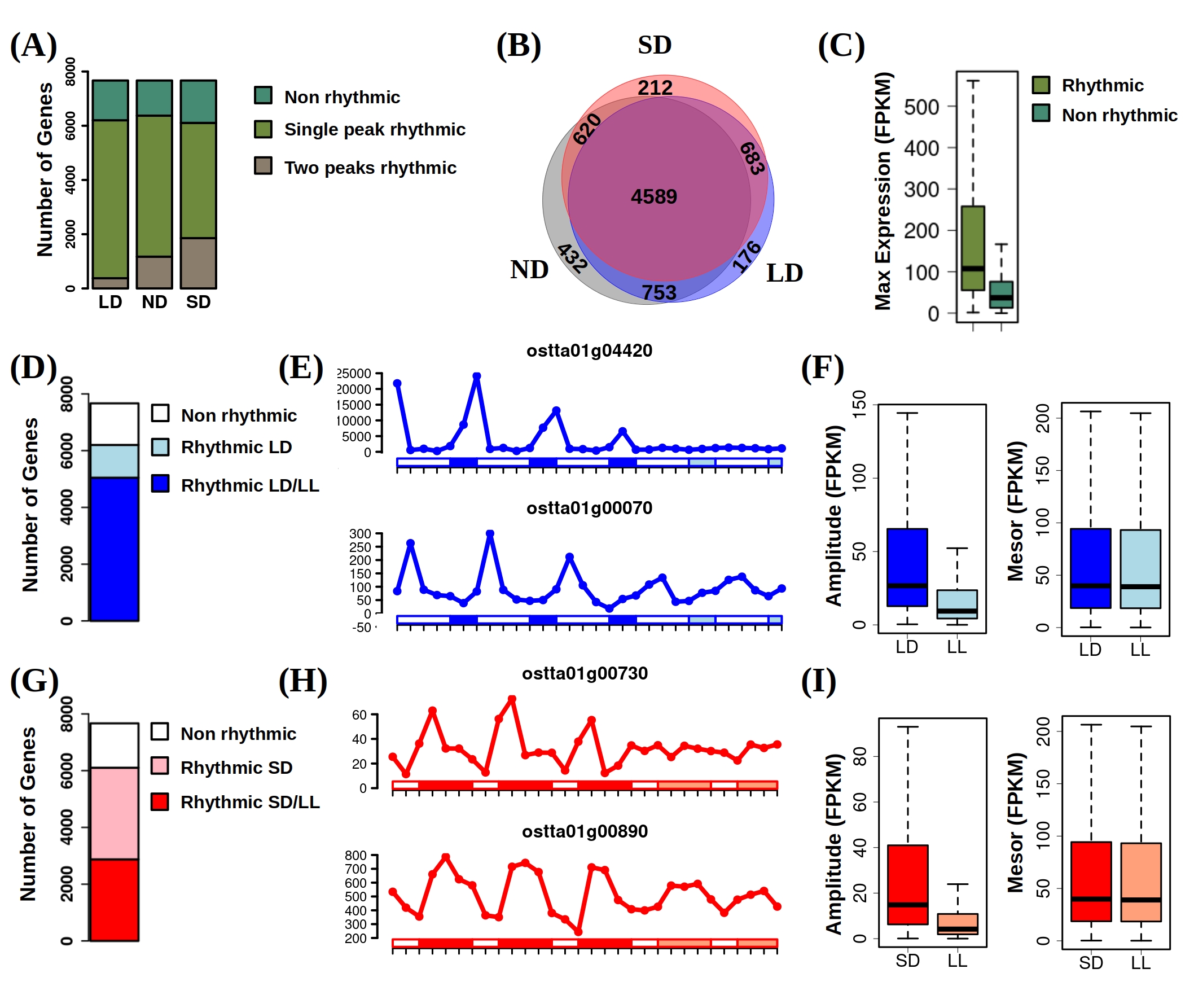
**Most genes in *Ostreococcus tauri* present diurnal rhythmic expression patterns under different photoperiods that are maintained in different proportions under free running conditions depending on the photoperiod.**

We used the bioconductor R package RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods) (Thaben and Westermark, 2014) to identify genes exhibiting diurnal rhythmic expression patterns under the three distinct seasonal conditions represented by different photoperiods in this study. We used both the RNA-seq data generated in this study for LD and SD conditions representing summer and winter together with previously published microarray data for ND conditions representing spring/autumn (Monnier et al., 2010). Specifically, we used time series consisting of three days with alternating light / dark periods. Independently from the photoperiod, more than 6000 genes comprising approximately 80% of the entire *Ostreococcus* genome present diurnal periodic rhythmic expression patterns, Figure 2A. The specific rhythmic genes under each photoperiod are practically coincident, Figure 2B (Supplemental Table 3). This result is in agreement with previous studies in *Ostreococcus tauri* and other microalgae such as *Chlamydomonas reinhardtii* (Monnier et al., 2010; Zones et al. 2015). We searched for simple diurnal rhythmic patterns with a single peak during a day by setting the period to 24 hours. More complex periodic patterns exhibiting two or three peaks per day were identified by fixing the period to 12 or 8 hours respectively. Under LD conditions most rhythmic genes show an expression profile with a single peak per day. Only 376 genes present a more complex rhythmic expression pattern cycling twice per day with two expression peaks and an apparent period of 12 hours. Remarkably, an increasing number of genes exhibit this more complex behavior under ND and SD conditions. Precisely, 1171 genes in ND and 1855 genes in SD representing 15.27% and 24.19% of the entire genome respectively, Figure 2A. This suggests that longer dark periods or nights in the different annual seasons induce the emergence of complex periodic gene expression patterns with two peaks during a complete day. This type of rhythmic genes is present in the transcriptome of other microalgae not being restricted only to *Ostreococcus tauri*. We reanalyzed previously published RNA-seq data generated in *Chlamydomonas* reinhardtii under ND conditions (Zones et al. 2015) and found that 2190 genes comprising 14.02% of the entire genome exhibit two expression peaks per day with an apparent period of 12 hours. The number of genes with expression patterns cycling three times in a day with an apparent period of 8 hours was negligible under the three different photoperiods comprising less than 1% of the genome.

Independently from the photoperiod, our analysis found that around 20% of the *Ostreococcus* genome is not significantly expressed following a rhythmic pattern. In order to further explore these non rhythmic genes we compared their maximum or highest expression values under LD and SD conditions with the ones reached by rhythmic genes. The maximum expression level of rhythmic genes is three times higher than the corresponding one for non rhythmic genes. This difference was significant according to a p-value of 1.45×10-4 computed using Mann-Whitney-Wilcoxson test. The current methods for detecting rhythmic gene expression are known to perform optimally only for highly expressed genes (Laloum and Robison-Rechavi, 2020). Therefore, the genes identified as non rhythmic in this study could indeed be rhythmic although their low expression level have prevented our methods from detecting them.

Circadian genes are defined as those rhythmic genes that maintained their cyclic gene expression under free running cond

**Figure 2.** Identification of genes exhibiting diurnal rhythmic expression patterns under distinct photoperiods corresponding to different seasons.



**DISCUSSION**

**METHODS**

**Culture Conditions**

*Ostreococcus tauri* sequenced strain RCC4221 was used for all experiments. The growth media was prepared using Artificial Sea Water (24.55g NaCl, 0.75g KCl, 4.07g MgCl2·6H2O, 1.47g CaCl2 ·2H2O, 6.04g MgSO4·7H2O and 0.21g NaHCO3 per 1L distilled water) supplemented with 1 mL of Solution I (100g NaNO3 in 250mL distilled water), 1mL of Solution II (700mg Na2HPO4 and 2.5g K2HPO4 in 250mL distilled water), 1mL of Solution III with trace metals (2.68g NH4Cl, 5.2g Fe-EDTA, 37.2g Na2 -EDTA, 23mg ZnSO4, 14mg CoSO4 , 7.89mg Na2 MoO4 ·2H2O, 2.5mg CuSO4 , 1.7mg H2SeO3 and 180mg MnCl2·4H2O in 500mL distilled water) and 1mL of f/2 vitamin solution.

RCC4221 is routinely maintained at 20ºC under LD conditions in flasks illuminated with white light providing approximately 50 μE m-2s-1 during the light period. When starting a new experiment one of such flasks was used to inoculate a roux culture bottle up to approximately 500 ml. This batch culture was grown at 20ºC under continuous white light (50 μE m-2s-1 ) and continuously sparged with air supplemented with 1% CO2. After approximately two weeks, the culture from the roux bottle was used to inoculate two photochemostats. These consists of water jacketed bubble columns with 2 L capacity (7 cm diameter, 50 cm height) containing 1.8 L of cell suspension continuously sparged with air to ensure culture homogenization. The flow of water from the jacket to an external cooling device () precisely controlled temperature at 20ºC. A pH probe is submerged into the culture and connected to a pH meter serving as input to a LabJack that controls an electrovalve allowing the on demand injection of CO2 into the air stream entering the culture to maintain the pH at 8.0. Each photochemostat is illuminated during the corresponding light periods using six Phillips PL-32 W/840/4p white-light fluorescent lamps. Instead of abrupt dark-to-light and light-to-dark transitions our illuminating system controlled by a LabJack ... simulates the progressive increase and decrease observed in solar daylight cycles according to the following formula:

Each photochemostat is kept within a wooden case and covered by a completely opaque fabric to ensure that the illumination is only provided by our systems.

Initially after inoculation with the culture from the roux bottle the photochemostats were operated in batch mode for about 3-4 days with incident irradiance set to either LD or SD conditions depending on the experiment with maximal irradiance value fix to 1.5 mE m-2s-1. The illuminating program was set in each photochemostat with a time offset of 12 hours in order to facilitate the collection of samples corresponding to entire day/night cycles in 8 hours approximately, Figure 1A. Subsequently, incident irradiance was progressively increased every day until reaching 2.5 mE m-2s-1. After approximately one week, the photochemostats were switched to continuous mode with a peristaltic pump adding fresh media continuously to the photochemostats during the light period at a flow rate of 45 mL h-1 (dilution rate 0.3 d-1). Excess culture was removed at the same rate by the overflow in order to keep a constant volume. Once the culture reached stationary phase entrainment under LD or SD conditions was maintained up to four weeks before sample collection.

**Sample Collection, RNA Extraction and Purification**

For each time point 50 mL of culture were collected for RNA extraction. Cells were pelleted by centrifugation (4 min, 5000 RCF, 4ºC). The supernant was discarded and the pelleted cells were quickly resuspended in Phosphate-buffered saline solution and pelleted again by centrifugation (1 min, 5000 RCF, 4ºC). After removal of the supernatant, pelleted cells were immediately flash frozen in liquid nitrogen and stored at -80ºC.

Further RNA purification was performed using the Isolate II RNA Plant Kit (Bioline). Washing, DNase treatment and elution were carried out following the manufacturer instructions. The eluted RNA concentration and integrity were measured using a bioanalyzer 2100 (Agilent RNA 6000 Nano Kit).

**RNA-seq data analysis**

Library preparation was carried out following the manufacturer instructions and sequencing was performed on the Illumina NextSeq500 sequencer. Approximately 10 million 75nt long single end reads were generated for each sample. The high quality of the sequencing data was assessed using the software package FASTQC. The *Ostreococcus tauri* genome sequence and annotation v3.0 (<https://mycocosm.jgi.doe.gov/Ostta4221_3/Ostta4221_3.home.html>) were used as reference genome (Blanc-Mathieu et al., 2014). Reads were mapped to this reference genome with HISAT2 (Kim et al., 2019). Transcript assembly and gene expression estimation measured as FPKM (fragments per kilobase of exon and million of mapped reads) were performed using StringTie (Kovaka et al., 2019) and the bioconductor R package ballgown (Frazee et al., 2015). Principal components analysis and hierarchical clustering were performed using the R package FactoMineR (Lê et al., 2008).

**Identification of genes exhibiting rhythmic expression patterns**

The bioconductor R package RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods) (Thaben and Westermark, 2014) was used to identify genes exhibiting rhythmic expression patterns. A p-value threshold equal to 0.05 was used in all the cases under study. Simple rhythmic expression patterns with a single peak over a diurnal cycle were identified by setting the period parameter to 24 hours. More complex rhythmic expression patterns exhibiting two or three peaks over a diurnal cycle were identified by setting the period parameter to 12 and 8 hours respectively. Three complete diurnal cycles were used to determine rhythmic genes under LD, ND and SD conditions. In order to detect circadian genes in LD and SD conditions the two last days of alternating light/dark periods and the two days under continuous light were considered in order to ensure a similar pattern was maintained in both light regimes and to prevent a bias towards any of the two conditions.

The significant difference between the maximum expression level in rhythmic and non rhythmic genes was performed using the Mann-Whitney-Wilcoxon non parametric test implemented in the R function wilcox.test.

**Statistical comparison between different rhythmic expression patterns**

Rhythmic gene expression profiles identified previously using RAIN were fitted to a cosinusoidal curve in order to be characterized according to three widely used parameters. Namely, mesor or the mean expression level of a gene around which its rhythmic expression profile oscillates; amplitude or half the difference between the highest and lowest expression level of a gene; and phase or time point at which a rhythmic gene reaches its highest expression value. The statistical significance of the differences in mesor, amplitude and phase between different groups of rhythmic gene expression patterns was performed using the R package CircaCompare (Parsons et al., 2020). A p-value threshold of 0.05 was used to determine statistically differences. The significance of the global differences in the three rhythmic parameters was performed using the Mann-Whitney-Wilcoxon non parametric test implemented in the R function wilcox.test.

**Functional Annotation of Gene Clusters**

**Data and Code Availability**

RNA-seq data generated in this study is freely available from the Gene Expression Omnibus (GEO) database under the accession number GSE155535. Previously published microarray data analyzed in this study is also available from the GEO database under the accession number GSE16422. The data analysis code developed using the statistical programming language R is freely available from the following GitHub repository: <https://github.com/fran-romero-campero/SANDAL>.

**AUTHORS CONTRIBUTION**

**ACKNOWLEDGMENTS**

**REFERENCES**

**Blanc-Mathieu, R., et al.** (2014) An improved genome of the model marine alga Ostreococcus tauri unfolds by assessing Illumina de novo assemblies. BMC Genomics **15:** 1103.

**Frazee, A.C., Pertea, G., Jaffe, A.E., Langmead, B., Salzberg, S.L., and Leek, J.T.** (2015) Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat. Biotechnol. **33:** 243–246.

**Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L.** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37:** 907–915.

**Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M.** (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. **20:** 278.

**Laloum, D., Robinson-Rechavi, M.** (2020) Methods detecting rhythmic gene expression are biologically relevant only for strong signal. PLoS Comput. Biol. **16:** e1007666.

**Lê, S., Josse, J., and Husson, F.** (2008) FactoMineR: an R package for multivariate analysis. J. Stat. Softw. **25:** 1–18.

**Monnier, A., Liverani, S., Bouvet, R., Jesson, B., Smith, J.Q., Mosser, J., Corellou, F., and Bouget, F.** (2010) Orchestrated transcription of biological processes in the marine picoeukaryote Ostreococcus exposed to light/dark cycles. BMC Genomics **11:** 192.

**Palenik, B., et al.** (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. USA **104:** 7705–7710.

**Parsons, R., Parsons, R., Garner, N., Oster, H., and Rawashdeh, O.** (2020) CircaCompare: a method to estimate and statistically support differences in mesor, amplitude and phase, between circadian rhythms. Bioinformatics **36**: 1208–1212.

**Thaben, P.F., Westermark, P.O.** (2014) Detecting rhythms in time series with RAIN, J. Biol. Rhythms **29**: 391–400.

**Zones, J.M., Blaby, I.K., Merchant, S.S., and Umen, J.G.** (2015) High-resolution profiling of a synchronized diurnal transcriptome from Chlamydomonas reinhardtii reveals continuous cell and metabolic differentiation. Plant Cell **27**:2743 –2769.

**FIGURE LEGENDS**

**Figure 1.** Experimental design and circadian synchronization in our cell cultures.

**(A)** Left in blue the experimental design for long day (LD) conditions (16 h light: 8 h dark). Two different photochemostats were used with a time offset of 12 hours in their illumination program. Samples at ZT0/ZT12, ZT4/ZT16 and ZT8/ZT20 were collected simultaneously from photochemostat 1 and 2 respectively. Similarly, right in red the short day (SD) conditions (8 h light: 16 h dark) experimental design. **(B)** Hierarchical clustering of the different time point transcriptomes under LD and SD conditions. Three clusters are identified: Midday (LD ZT4, SD ZT4 and LD ZT8), Dusk (SD ZT8, LD ZT16 and LD ZT12) and Night/Dawn (LD ZT0, LD ZT20, SD ZT12, SD ZT16, SD ZT20 and SD ZT0). **(C)** Principal Components Analysis of the different time point transcriptomes under LD conditions. The transcriptomes corresponding to the same time points from the three different days cluster together constitute a global structure resembling a circle. **(D)** Principal Components Analysis of the different time point transcriptomes under SD conditions. More variability is observed and a global structure resembling an ellipse is apparent.

**Figure 2.** Identification of genes exhibiting diurnal rhythmic expression patterns under distinct photoperiods corresponding to different seasons.  **(A)** Number of rhythmic genes with a single or two peaks in a day and non rhythmic genes depending on the photoperiod. **(B)** Comparison between rhythmic genes under different photoperiods. **(C)** Highest expression value of rhythmic and non rhythmic genes under LD and SD conditions. **(D)** Number of rhythmic genes exclusively under LD conditions and under both LD and LL conditions. **(E)** Top an example of rhythmic gene under LD conditions losing rhythmicity under LL conditions. Bottom an example of rhythmic genes under LD conditions maintaining rhythmicity under LL conditions. **(F)** Effect over the amplitude and mesor of rhythmic gene expression in LD when transferred to LL conditions. **(G)** Number of rhythmic genes exclusively under SD conditions and under both SD and LL conditions. **(H)** Top an example of rhythmic gene under SD conditions losing rhythmicity under LL conditions. Bottom an example of rhythmic genes under SD conditions maintaining rhythmicity under LL conditions. **(F)** Effect over the amplitude and mesor of rhythmic gene expression in SD when transferred to LL conditions.

**TABLES**