

Doctoral Thesis

Multi-omics Characterization  
of the Responses to  
Seasonal Variations in Diel Cycles  
in the Marine Picoeukaryote  
*Ostreococcus tauri*

Ana Belén  
Romero-Losada

Sevilla, 2024





Doctoral thesis

# Multi-omics Characterization of the Responses to Seasonal Variations in Diel Cycles in the Marine Picoeukaryote *Ostreococcus tauri*

*Dissertation presented by Ana Belén Romero Losada to obtain  
the PhD Degree by Universidad de Sevilla.*

*Supervisors:*

*Prof. Francisco José Romero Campero  
Prof. Mercedes García González*



*This work was supported by the research projects MINOTAUR (BIO2017-84066-R) and BLOOM (RTC-2017-6080-5) from the Spanish Ministry of Science and Innovation.*



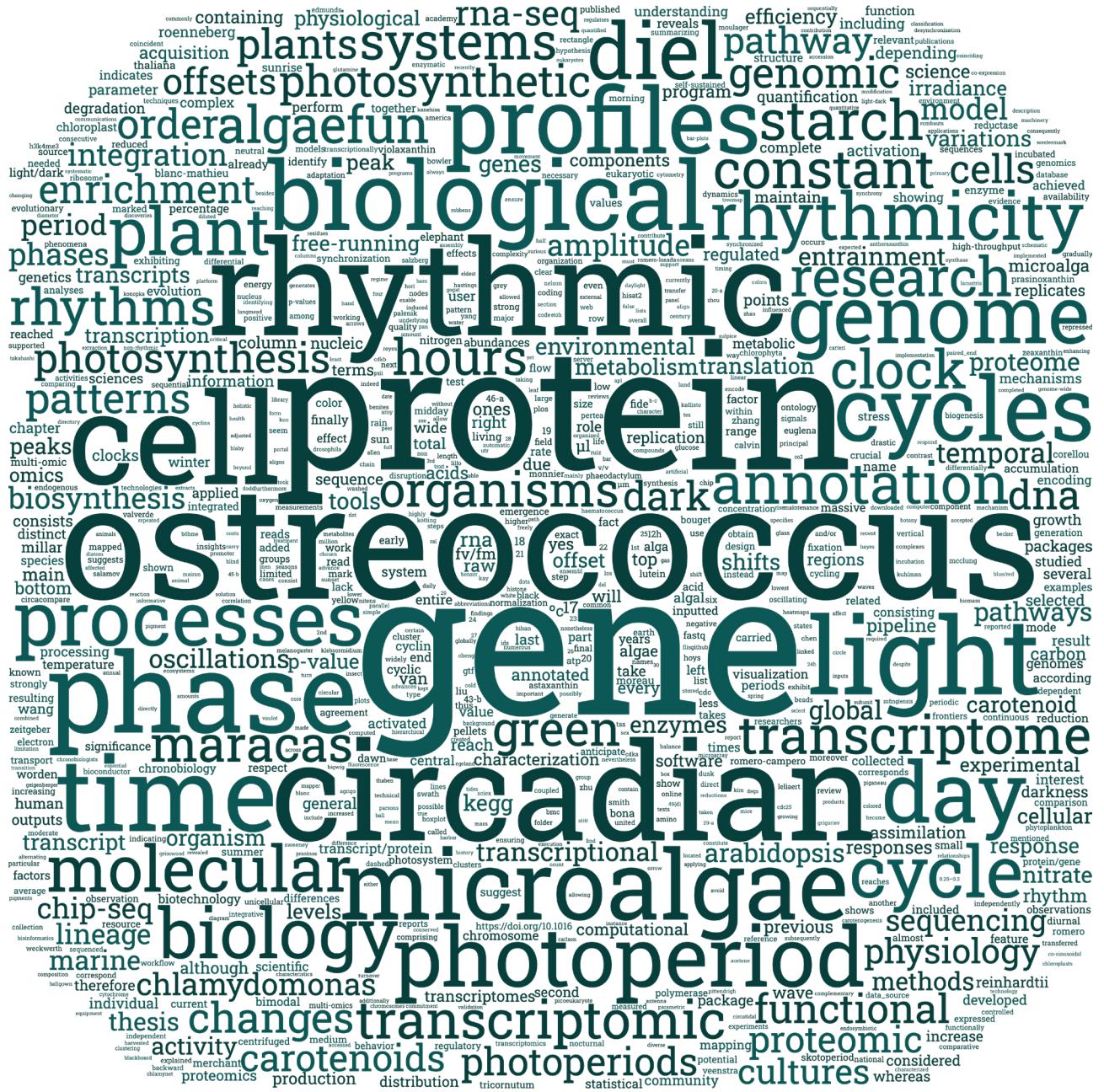
## Acknowledgments

Cuando finalmente me senté a escribir este trabajo, me sentía como una hormiga mirando a una montaña: sola y con vértigo. La montaña era tan alta que no conseguía ver el final. Suerte la mía, que me he rodeado de personas que han sabido guiarme con paciencia y cariño.

Innumerables veces he entrado en el despacho de Mercedes, directora de esta tesis, con una nube negra en la cabeza y he salido de allí, horas más tarde, con determinación y un camino claro para seguir. Gracias por el tiempo dedicado, los innumerables consejos y el apoyo, te admiro mucho. Por otro lado, Fran, el otro director de esta tesis, ha sido la persona en quien me he mirado desde que lo conocí hace 8 años. El trato humano que tiene con sus estudiantes siempre será un ejemplo a seguir para mí. Por ello, después de todo el tiempo que hemos compartido (que ha sido muchísimo) solo nos quedan batallitas que ahora recordamos y nos hacen reír. Gracias por esas risas, por los empujones después de cada fallo y por saber comprenderme. Ambos, en conjunto, me han otorgado una visión científica que me abrirá muchas puertas y, en lo personal, un inicio de mi carrera científica lleno de cariño y buen hacer. Gracias también al resto de personas que a día de hoy forman parte de su laboratorio, mis compañeros/as. Seguid apoyándenos entre vosotros y buscando ratitos de diversión, yo os voy a echar mucho de menos. Sé por experiencia cómo era todo antes de que llegarais y os aseguro que lo hacéis más fácil. Sin embargo, Pablo siempre ha estado, en la puerta de al lado y fuera. “He hecho cocido, ¿quieres?”. Gracias también al servicio de secuenciación del CABIMER, de citometría del IBIS y proteómica, cromatografía y microscopía del IBVF. Es un placer poder contar con grandes profesionales tan cerca. Thanks to van Ooijen’s lab for having me during those months. The discussions we had were refreshing and nourishing for me.

Fuera de esta burbuja científica he tenido a personas que me han preguntado mil veces cómo lo llevo, aunque no tuvieran muy claro qué es esto que estoy haciendo en la universidad. Aun así, saben que es importante para mí y me han escuchado pacientemente. Entre ellas están mis amigas, mis grandes confidentes, compartir tiempo con ellas ha sido y será la forma más efectiva de recargarme de energía. Quien ha convivido más de cerca con esta etapa ha sido JD. En los llantos y celebraciones siempre has estado tú. Gracias por espantar a los miedos y recordarme mi valor cuando me asaltan las dudas. Por otro lado, están las personas que más cuidan: mi familia. El amor incondicional que les haría recorrer el mundo para venir a darme un abrazo si se lo pido. Lo sé porque lo han hecho, y sus “Te llevo coliflores y molletes” o “Eso te lo arreglo yo en un momento” me han sacado a flote innumerables veces. Cada vez veo en mí más cosas que reconozco tuyas y me llena de alegría, pues admiro la valentía y resiliencia que les caracteriza. Gracias por dármelo todo, durante estos años de tesis y siempre.







# Index

<b>Abstract .....</b>	<b>13</b>
<b>Introduction.....</b>	<b>17</b>
Chronobiology.....	19
Circadian research.....	24
<i>Ostreococcus tauri</i> .....	28
Molecular Systems Biology.....	34
<b>Materials and Methods.....</b>	<b>41</b>
Organism and culture conditions.....	43
Organism and growth medium	
Continuous culture in photochemostats	
Experimental design .....	44
Transcriptomic analysis.....	46
Sample collection	
Cell disruption	
RNA extraction	
RNA purification	
RNA sequencing and processing	
Proteomic analysis.....	47
Sample collection	
Cell disruption	
Proteins extraction	
Proteins digestion	
SWATH acquisition.....	48
<i>Equipment and data acquisition method</i>	
<i>Library construction</i>	
<i>SWATH runs.....</i>	49
<i>Data processing</i>	
Cell cycle analysis.....	49
Sample collection and cell fixation method	
Cell staining method	
Data acquisition and processing	
Confocal fluorescence imaging.....	50
Determination of photosynthetic activity.....	50
Sample collection	
Data acquisition	
Analytical determinations.....	51
Sample collection	
Starch content	
<i>Cell disruption</i>	
<i>Starch solubilization and digestion</i>	
<i>Spectrophotometric quantification</i>	
Carotenoid content.....	52
<i>Cell disruption</i>	
<i>Carotenoids extraction</i>	
<i>Carotenoids determination and quantification</i>	
Enzyme assay.....	53
Rhythmic patterns analysis .....	53
Rhythmic patterns detection	
Rhythmic patterns comparison.....	54
Data and code availability.....	54
Hypothesis and Objectives.....	57

## **Results.....61**

<b>Chapter 1. ALGAEFUN with MARACAS, microALGAE FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq Analysis.....63</b>
Implementation.....66
<i>Integration of different microalgae databases.....66</i>
<i>Development of functional annotation and genomic packages.....69</i>
<i>MARACAS implementation: high-throughput sequencing data processing.....69</i>
<i>ALGAEFUN implementation: functional annotation analysis.....75</i>
Case study 1: from RNA-seq raw sequencing data to biological processes and pathways.....80
Case study 2: From ChIP-seq raw sequencing data to marked genes.....82
Contribution of ALGAEFUN with MARACAS to the field.....83
<b>Chapter 2. Transcriptomic analysis of diel and seasonal cycles in <i>Ostreococcus tauri</i>.....85</b>
Transcriptomic characterization of diel expression profiles.....89
<i>Global rhythmicity analysis of long day and short day transcriptomes.....89</i>
<i>Identification of bona fide circadian genes.....90</i>
<i>Rhythmicity analysis of transcriptomes corresponding to free-running conditions.....95</i>
Transcriptomic characterization of seasonal effects over gene expresion profiles.....98
<i>Analysis of rhythmic profiles with 12 h period.....100</i>
<i>Characterization of temporal transcriptional programmes during diel cycles.....103</i>
<b>Chapter 3. Proteomic analysis of seasonal variations in diel cycles in <i>Ostreococcus tauri</i>.....109</b>
Proteomic characterization of diel rhythmic abundance profiles.....113
Correlation analysis of transcript-protein phase offsets.....115
<b>Chapter 4. Diel and seasonal multiomic integration with physiological data.....121</b>
Cell Division Cycle (CDC) of <i>Ostreococcus tauri</i> under seasonal variations in diel cycles.....123
<i>Temporal programme of cell division cycle under summer and winter photoperiod.....124</i>
<i>Integration of CDC programme with transcriptomic and proteomic data.....127</i>
Diel and seasonal rhythm of photosynthesis in <i>Ostreococcus tauri</i> .....131
<i>Rhythmic oscillations of photosynthetic efficiency under summer and winter photoperiods.....133</i>
<i>Integration of photosynthesis efficiency rhythmic profile with multi-omics data.....134</i>
<i>Integration of starch content diel oscillations with multiomic data.....138</i>
Other metabolic pathways of <i>Ostreococcus tauri</i> showing periodic oscillations under diel and seasonal cycles.....140
<i>Carotenoids biosynthesis in <i>Ostreococcus tauri</i> under diel and seasonal cycles.....140</i>
<i>Integration of multi-omics data with oscillations described by carotenoids content in <i>Ostreococcus tauri</i> under diel and seasonal cycles.....141</i>

<i>Nitrate assimilation under diel and seasonal cycles in Ostreococcus tauri.....</i>	144
<i>Integration of key enzyme activities from nitrate assimilation pathway with multi-omic data.....</i>	145
<b>Conclusions.....</b>	<b>149</b>
<b>Bibliography.....</b>	<b>155</b>
<b>Publications.....</b>	<b>193</b>
<b>Appendix.....</b>	<b>199</b>



# **Abstract**



## **Abstract**

Earth tilted rotation and translation around the Sun produce one of the most pervasive periodic environmental signals on our planet giving rise to seasonal variations in diel cycles. Marine phytoplankton undergo significant alterations in response to these signals and, although it plays a key role on ecosystems, their response to these rhythms remains largely unexplored. In this work, the marine picoalga *Ostreococcus tauri* is chosen as model organism grown under summer long days, winter short days, constant light and constant dark conditions to characterize these responses. The 80 % of its transcriptome present diel rhythmicity and genes with robust self-sustained rhythmic expression profiles are identified. A drastic reduction in proteome rhythmicity with respect to transcriptome rhythmicity is observed with 25 % of the proteins oscillating. Seasonally specific rhythms are found in key physiological processes such as cell cycle, photosynthesis, carotenoid biosynthesis, starch accumulation and nitrate assimilation. A global orchestration between transcriptome, proteome and physiological dynamics is observed with specific seasonal temporal offsets between gene expression, protein abundance levels and physiological activity.

## **Resumen**

Los movimientos de rotación y traslación de la Tierra alrededor del Sol producen una de las señales ambientales periódicas más dominantes en nuestro planeta: ciclos de luz y oscuridad con variaciones estacionales. El fitoplancton marino experimenta alteraciones significativas en respuesta a estas señales y, aunque desempeñan un papel clave en los ecosistemas, su respuesta a estos ciclos a lo largo de las diferentes estaciones del año permanece en gran medida inexplorado. En este trabajo, se ha elegido la picoalga marina *Ostreococcus tauri* como organismo modelo cultivado en condiciones de días largos de verano, días cortos de invierno, luz constante y oscuridad constante para caracterizar estas respuestas. El 80 % del transcriptoma presenta ritmidad cíclica y pudieron identificarse genes cuyos perfiles de expresión cílicos eran mantenidos ante cualquier condición externa. Se observa una drástica reducción de la ritmidad del proteoma con respecto a la del transcriptoma, con un 25 % de las proteínas oscilando. Además, existen ritmos estacionales específicos en procesos fisiológicos clave como el ciclo celular, la fotosíntesis, la biosíntesis de carotenoides, la acumulación de almidón y la asimilación de nitrato. En general, se describe una orquestación global entre el transcriptoma, el proteoma y la fisiología con desfases temporales estacionales específicos entre la expresión génica, los niveles de abundancia de proteínas y la actividad fisiológica.

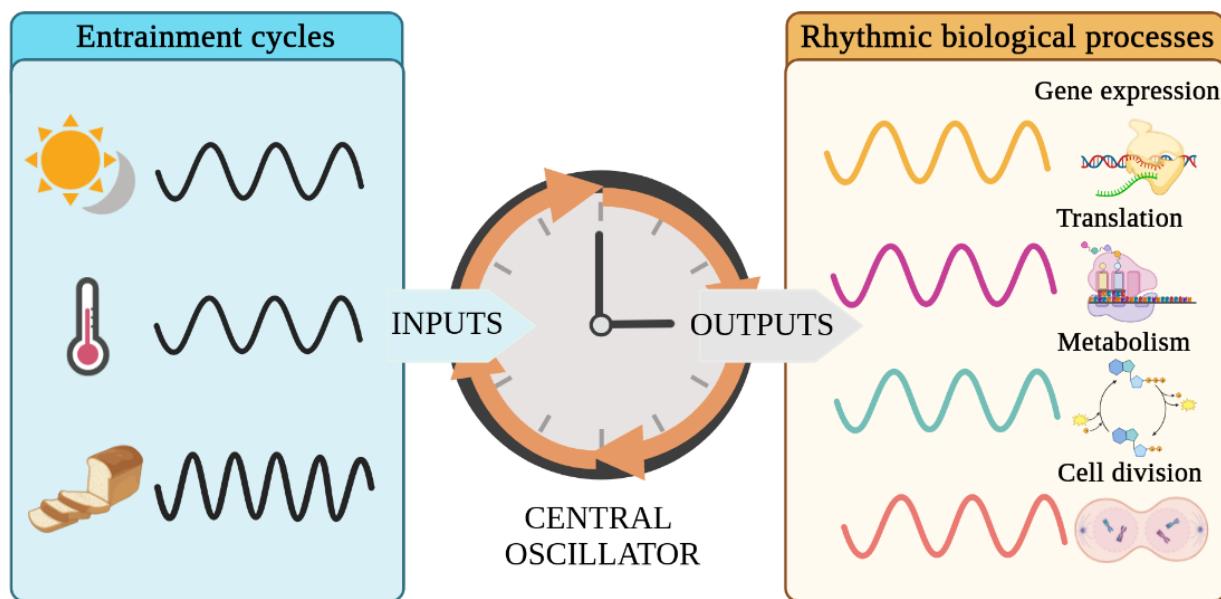


# Introduction



# Chronobiology

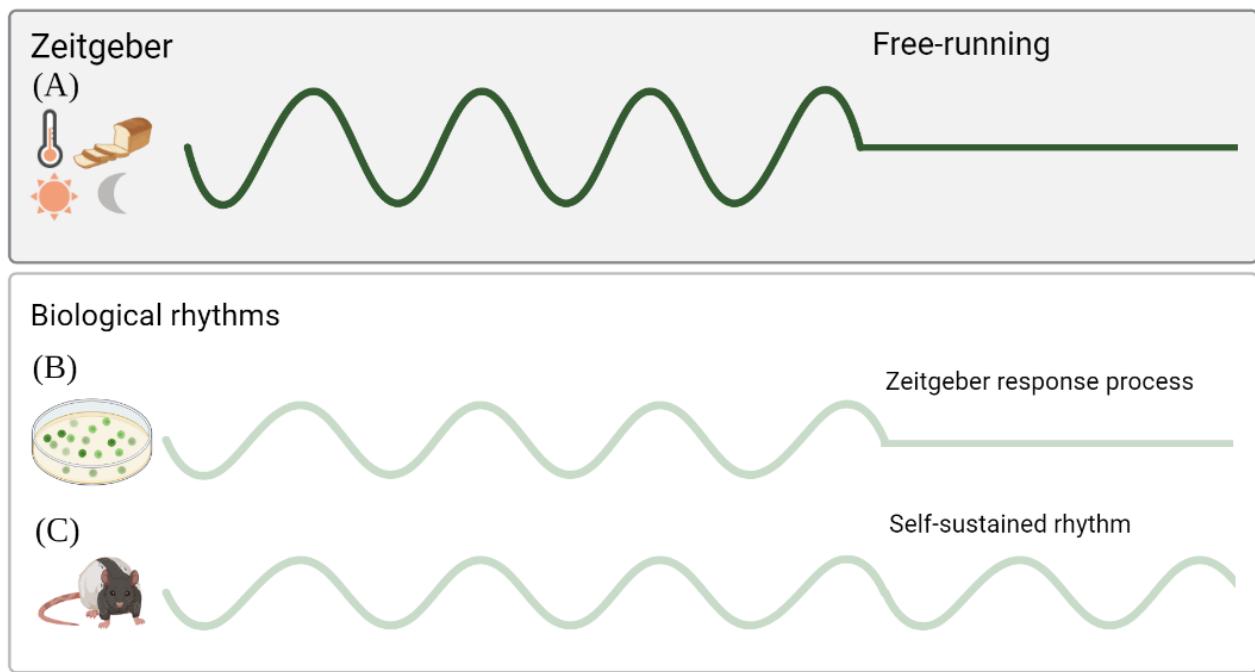
Simple direct observations of our surroundings easily unveil many cyclic processes. We are so familiar with them that we are mainly unconscious about their drastic influences over our lives and bodies, in particular, and life on Earth, in general. Four environmental cycles are the most apparent ones on Earth: tides (which repeat every 12 and half hours), lunar cycles (lasting 28.5 days), days (every 24 hours) and years (every 365.25 days). The biological processes governed by these cycles are called circatidal, circalunar, circadian, or circannual (Numata et al., 2015). These cycles arise from gravitational forces exerted by the Sun and the Moon over the Earth. Their key characteristic is full predictability. For example, it is possible to determine the exact time of a high tide months in advance. We constantly see how full moons are always followed by new moons, and everybody goes to bed certain that the sun will rise next morning. The exact date and even time for the next solar equinox, the moment when Earth's rotational axis is perpendicular to the line linking the Sun and the Earth, is known and written in all calendars all over the world. Therefore, sequential and periodic seasonal changes are never a surprise. Since these four cycles affect the Earth with an overwhelming precision, earthlings have evolved endogenous molecular systems, referred to as molecular clocks, that allow them to anticipate those cyclic changes (Fig. 1).



**Figure 1: Diagram of the central oscillator known as clock.** External cyclic inputs, such as light/dark cycles or photoperiod, temperature changes and nutrients availability, are represented in the blue panel (left). Rhythmic biological processes as outputs of the clock (central oscillator) are represented in the yellow panel (right). Figure created using Biorender.

Molecular clocks (Fig. 1) can be seen as central oscillators susceptible to be entrained by external environmental cyclic inputs (rhythmic environmental changes) that, in turn, generate rhythmic outputs controlling a myriad of biological processes. This has motivated the emergence of chronobiology, a research area that studies how rhythmic environmental changes affect responses and adaptions of living organisms (Edmunds, 1983; Kuhlman et al., 2018). Chronobiologists have found biological rhythms regulated by the clock in a wide range of spatial scales, ranging from the molecular level (transcription, translation, protein degradation, metabolites biosynthesis, etc.), to the cellular and tissue level (cell division, synaptic connections, apoptosis, etc.) or even at the level of whole organisms, populations and ecosystems (Edmunds, 1983; Merrow et al., 2005; Sharma et al., 2022). A distinctive and defining characteristic of clock-regulated biological rhythms is self-sustainability under constant non cycling environmental signals (Pittendrigh, 1960; Roenneberg & Merrow, 2005). Typically, when the response of an organism to rhythmic environmental changes (light/dark cycles, food availability, temperature changes, etc.) is studied (Fig. 2-A), multiple biological processes are found showing rhythmic profiles (Fig. 2-B, C). Some of these processes would not maintain their rhythmicity under constant conditions indicating that their rhythmic profiles are direct responses to changes in the environment rather than being generated by endogenous oscillating systems or molecular clocks. Therefore, when environmental signals are constant, the profiles of such biological processes stop oscillating and remain constant as well (Fig. 2-B). However, a subset of biological processes would maintain a rhythmic profile under constant environmental conditions, revealing that their rhythmicity is self-sustained and thus, regulated by endogenous oscillating systems acting as molecular clocks (Fig. 2-C).

Accordingly, chronobiology experiments are commonly designed as follows: a temporal series consisting of consecutive days when the organism is exposed to a rhythmic environmental condition (called zeitgeber, which is used as a synchronizer, literally a time giver) are followed by several consecutive days when the organism is exposed to constant non cycling conditions termed as free-running conditions (Fig. 2-A) (Kuhlman et al., 2018). Data are collected with a specific time interval, every few hours, minutes or seconds depending on the complexity of the data. Specifically, in the case of circadian experiments, as the ones performed in this thesis, the zeitgeber is the alternation of light-dark cycles and the free-running conditions consist of constant light and constant dark periods.



**Figure 2: Different patterns of response to rhythmic environmental changes.**

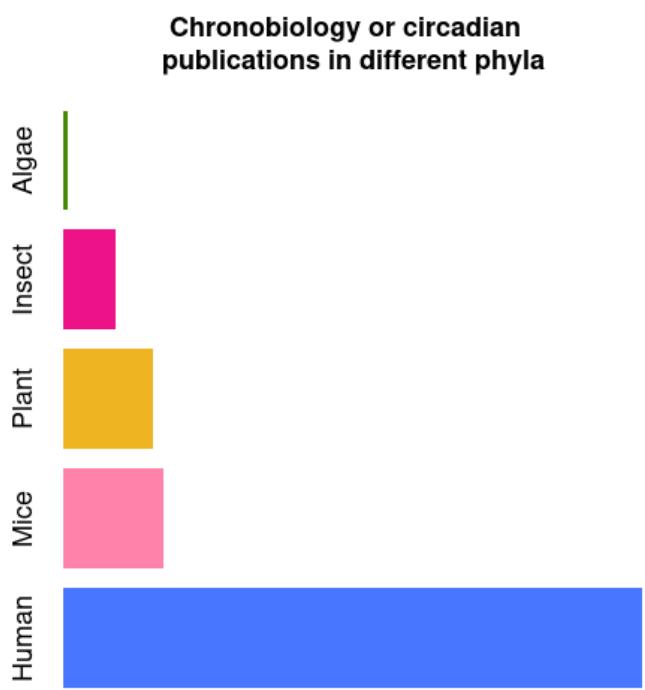
(A) Common experimental design used in chronobiology studies, consisting of consecutive days under rhythmic environmental conditions and several consecutive days under free-running conditions; (B) Figurative oscillation corresponding to a zeitgeber responding process; (C) Figurative oscillation corresponding to a self-sustained process regulated by an endogenous clock. Figure created using Biorender.

Following this experimental design, chronobiologists have described self-sustained biological rhythms reacting to the four different previously described environmental cycles acting as zeitgebers. In particular, some marine organisms have been shown to produce self-sustained circatidal rhythms when they are kept in laboratory tanks without its zeitgeber, in this case, tidal rhythms (Rock et al., 2022). As an example, the marine diatom *Hantzschia amphioxys* descends to the sand at high tides and rises to the surface at low tides (Fauré-Fremiet, 1951). In contrast, self-sustained circalunar processes are yet more unknown (Andreatta & Tessmar-Raible, 2020). An eminent example is the larvae of the insect called Ant Lion, which builds small holes in the sand as traps for insects. Scientists found that the size of the traps changes showing a circalunar profile that is maintained under constant conditions (Youth & Moran, 1969).

However, scientific studies about circadian or circannual rhythms are much more numerous, and they have been found in a wide range of organisms (Merrow et al., 2005; Pfeuty et al., 2012; Roenneberg & Merrow, 2005). In animals, circadian rhythms studies focus mainly on activity-rest cycles (Roenneberg et al., 2022; Zee & Abbott, 2020). Nevertheless, a myriad of

other parameters from ethology to gene expression also show circadian profiles. For example, the olfactory discrimination in mice is higher at night, even at subjective nights under free-running conditions (Granados-Fuentes et al., 2006). Also fungi show circadian rhythmic phenomena, for instance, *Neurospora crassa* generates asexual spores according to a rhythm of 24 h that is maintained under constant darkness (Correa & Bell-Pedersen, 2002). Organisms in the green lineage react in many different ways to circadian cycles (Noordally & Millar, 2015), one of the most known ones for microalgae is the 24 h cyclic movement in the water column adjusted to their metabolism requirements (Lebert et al., 1999), or the described rhythmic leaf movement, growth rate and stomatal opening in plants (Merrow et al., 2005). Day length or photoperiod is a crucial signal for the circannual timing system. In fact, seasonal changes in photoperiod have been strongly connected to reproduction. Many animal reproductive systems are related to seasons, from gene expression profiles to anatomical structures. Hamsters kept in short day condition have 10-fold smaller testes than the ones kept in long day condition (Klante & Steinlechner, 1994; Nishiwaki-Ohkawa & Yoshimura, 2016). But also in plants, flowering and seed production are photoperiod-regulated (Brandoli et al., 2020; Serrano-Bueno et al., 2017).

Since light serves as the primary source of energy for photosynthetic organisms, they are highly synchronized with cyclic environmental changes, such as circadian rhythms and photoperiods, as previously discussed. However, the genetics and molecular techniques used to identify clock components in other taxa have not been widely applied to microalgae yet (Noordally & Millar, 2015). Consequently, chronobiology in microalgae and picoalgae remains barely studied compared with other photosynthetic organisms, despite representing one of the largest polyphyletic groups in the eukaryotic domain (Fig. 3). This thesis aims to contribute to the chronobiology community by developing tools specifically designed for microlgae and describing, for the first time, circadian rhythms emerging under different seasonal variations in diel cycles in the model marine planktonic picoalga *Ostreococcus tauri*, considered the evolutionary eldest sister in the green lineage.



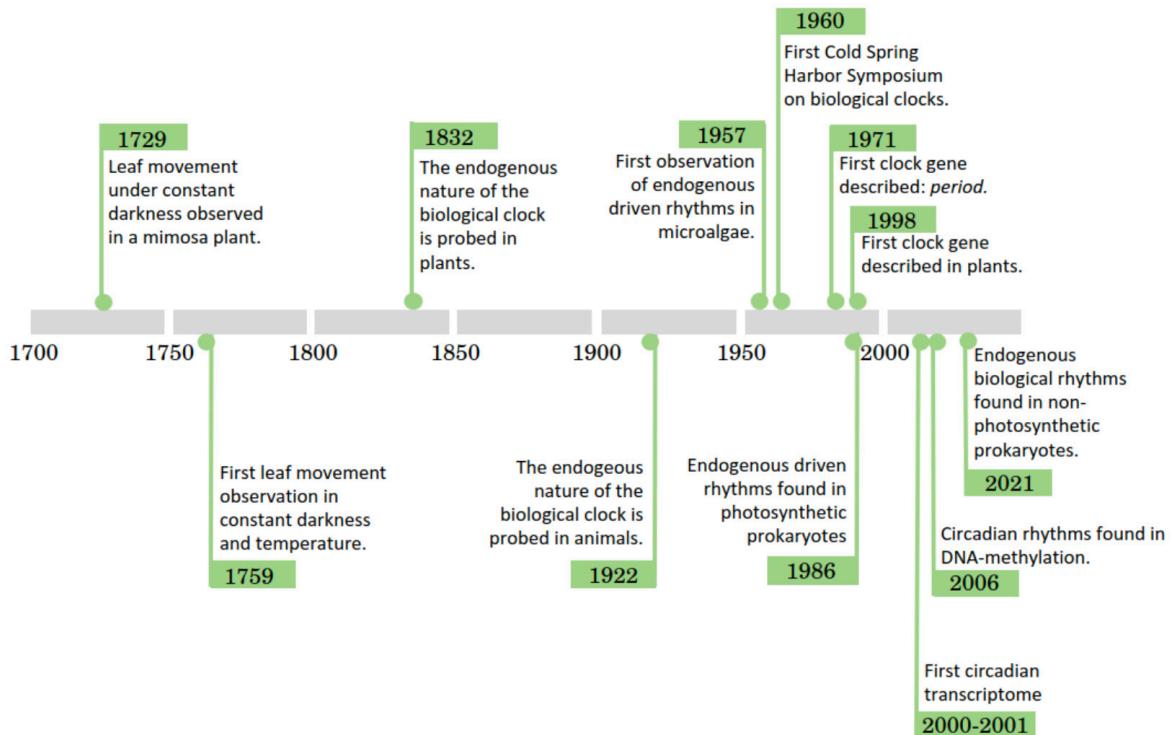
**Figure 3: Number of publications found in PubMed using “chronobiology” or “circadian” keywords in their abstracts in July of 2023.** For the algae group (using the generic term “algae”, “microalgae”, “*Ostreococcus tauri*”, “*Chlamydomonas reinhardtii*” and “*Euglena*”), only 493 publications were found. For the insect group (including the generic term “insect” and “*Drosophila melanogaster*”), 5717 publications were found. For plants (including the generic term “plant” and “*Arabidopsis thaliana*”), 9734 publications were found. For mice and human, 10771 and 62305 publications were found, respectively.

## Circadian research

Our society is worldwide structured in a 24 h / 7 days system. People from different countries and cultures are experiencing jet-lags, shift work, exposure to artificial light and lack of outdoor activities. Circadian clock research is gaining relevance since this phenomenon has a crucial impact on human health, behavior and quality of life (Mermet et al., 2017; Roenneberg et al., 2019, 2022; Roenneberg & Merrow, 2016). Understanding the physiology, genetics and epigenetics (Ripperger & Merrow, 2011) at a laboratory level using different model organisms besides mice (such as plants, fungi and microalgae) will greatly contribute to advance our comprehension of the mechanisms and functioning of circadian clocks in living organisms on Earth beyond humans.

The synchrony that exists between sunrise / twilight and organisms functioning have been so obvious for scientists that the underlying molecular mechanisms remained ignored and unexplored for centuries. The first observation indicating that daily rhythms were programmed took place in the 18th century. Jean Jacques d'Ortous De Mairan, a French astronomer, described in less than 350 words (nowadays, less than two tweets) how a mimosa plant inside a closet maintained its daily leaf movement (De Mairan, 1729). For some scientists, it was a clear proof that leaf movement was not controlled by the alternation of light and dark cycles indicating the existence of an endogenous oscillating system. De Mairan invited botanists to investigate his discovery to confirm that leaf movement was also maintained when temperature changes were absent. However, it took 30 years to confirm his observations taking into account temperature for the first time (Duhamel du Monceau, 1759; Kuhlman et al., 2018; Merrow et al., 2005). The history of circadian research was sparse in results with large gaps between discoveries during its early stages (Fig. 4). For instance, the physiology of the endogenous nature of the clock in plants was not studied until 1832 (De Candolle, 1832), despite De Mairan's observations being reported in 1729. Similar phenomena in animals took another century to be identified (Richter, 1922), and 50 years more were needed to find them in humans (Kuhlman et al., 2018; McClung, 2006; Roenneberg & Merrow, 2005). Although circadian research and chronobiology were born in the 18th century, they only became relevant and coherent in the second half of 20th century. One of the breaking points in circadian research history was the international conference in Cold Spring Harbor in 1960, where 157 pioneering scientists in the field such as Colin Pittendrigh, Patricia Decourse and Franz Halberg met together for the first time (Evans, 1961). From all the data shared,

Pittendrigh summarized the qualities of circadian clocks in 16 generic empirical features that he predicted to be true in all organisms (Pittendrigh, 1960). Nowadays, 62 years later, these empirical features are still useful in spite of the huge development in the field.



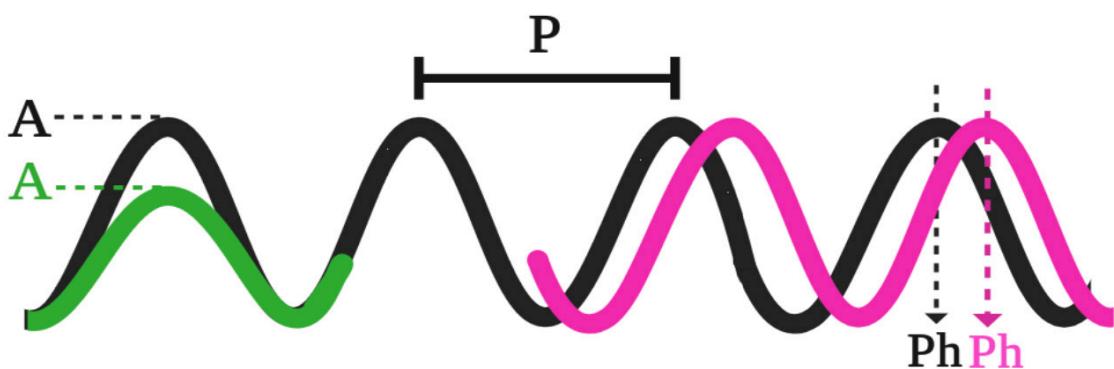
**Figure 4: Timeline of circadian research. The main circadian discoveries have been listed in chronological order. (De Mairan, 1729; Duhamel du Monceau, 1759; De Candolle, 1832; Richter, 1922; Hastings & Sweeney, 1957; Evans, 1961; Konopka & Benzer, 1971; Mitsui et al., 1986; Somers et al., 1998; Akhtar et al., 2002; Eissenberg & Elgin, 2006; Eelderink-Chen et al., 2021; Kuhlman et al., 2018; Roenneberg & Merrow, 2005)**

Recent technological approaches have enabled a more controlled experimentation and data analysis in chronobiological studies (Fig. 4). Chronobiologists have already identified circadian rhythms in organisms spanning almost the entire tree of life (Hastings & Sweeney 1957; Mitsui et al., 1986), even non-photosynthetic prokaryotes (Eelderink-Chen et al., 2021). Also the genetics behind the circadian clock have been broadly studied in a wide range of phyla, since the first clock gene was described in 1971 by Seymour Benzer and Ronald Konopka using mutant screening in *Drosophila melanogaster* (Konopka & Benzer, 1971; Takahashi, 2021) and the first clock gene in plants in 1998 (Somers et al., 1998). These discoveries about individual genes directly involved in the circadian clock started the path for massive genetics approaches, obtaining in the early 2000s the first circadian transcriptome (Akhtar et al., 2002).

followed by the discovery of DNA methylations following circadian rhythms (Eissenberg & Elgin, 2006). As it can be observed in the timeline, circadian rhythms discoveries are very diverse. Chronobiologists usually have strong roots in other fields such as anatomy, physiology, molecular biology, genetics, ecology, mathematics and/or computer science. The knowledge obtained from each field have been shared and integrated in order to obtain a precise characterization of circadian rhythms (Klante & Steinlechner, 1994; Merrow et al., 2005; Nishiwaki-Ohkawa & Yoshimura, 2016). In recent years, mathematics and computer science have strongly influenced the entire field of research by proposing a paradigm in which biological rhythms are studied as waves. Wave patterns repeat themselves periodically, maintaining several features that define their profiles. These features can be modelled as wave parameters and used to quantitatively compare different waves. Cosinusoidal parametrizations of the following form are commonly applied in circadian research to model rhythmic phenomena:

$$Y = m + A \cos\left(\frac{2\pi}{\tau}(t - Ph)\right)$$

where Y represents the measurements of the phenomenon under study such as gene expression or protein abundance in this thesis; m is the mesor or wave mean value around which oscillations take place; A is the wave amplitude or range between the maximum and minimum wave values; P is the wave period or the time length between successive repetition of the same wave pattern which is fixed to 24 h in circadian research and Ph is the wave phase or time point when the maximum value is reached (Fig. 5).



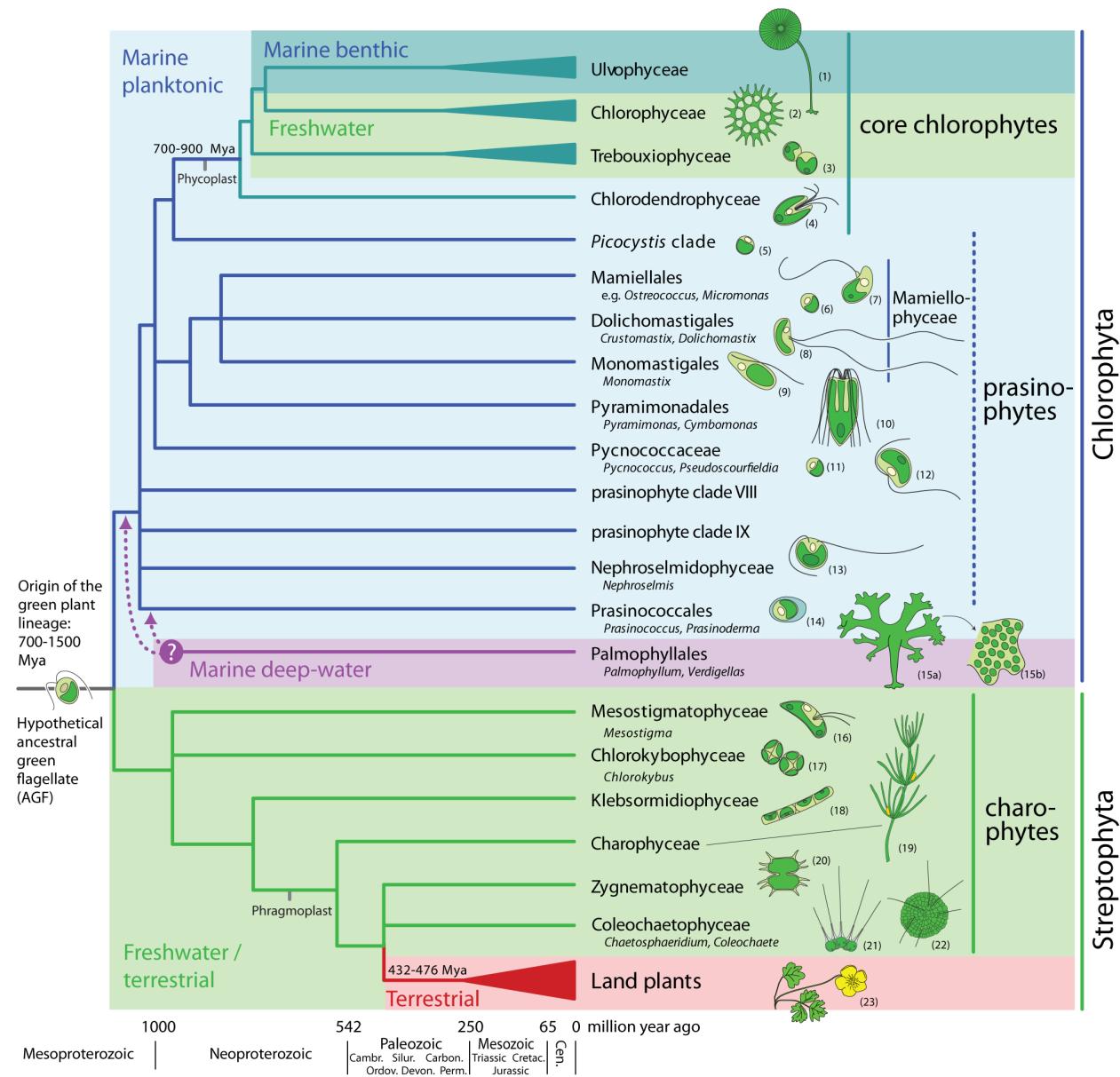
**Figure 5: Graphical representation of cosinusoidal wave model parameters.** “P” stands for period, which corresponds to the time interval between consecutive repetitions of the wave pattern. “A” stands for amplitude, it corresponds to the difference between the highest and lowest points the wave reaches. The green oscillation represents a wave with a different amplitude when compared to the black wave. “Ph” stands for phase, which corresponds to the time point where the wave reaches its maximum value. The pink oscillation represents a wave with different phase compared to the black wave.

These parameters can be estimated from experimental data applying methods such as non-linear least squares for different rhythmic phenomena and the statistical significance of the differences between them can be assessed (McClung, 2006; Parsons et al., 2020). A major drawback of the previous parametric models consists in their ability to detect only symmetric waves failing to identify more complex rhythmic patterns. Therefore, non-parametric models have been proposed to complement these limitations detecting rhythmic patterns of arbitrary form. Robust non-parametric models have been developed based on rank tests for umbrella alternatives (Mack & Wolfe, 1981; Thaben & Westermark, 2014). Specifically, in this thesis these two complementary approaches, parametric and non-parametric, have been applied to obtain a characterization of circadian rhythms at three different molecular levels (transcriptome, proteome and physiology) in *Ostreococcus tauri* subjected to photoperiodic seasonal variations of diel cycles.

## ***Ostreococcus tauri***

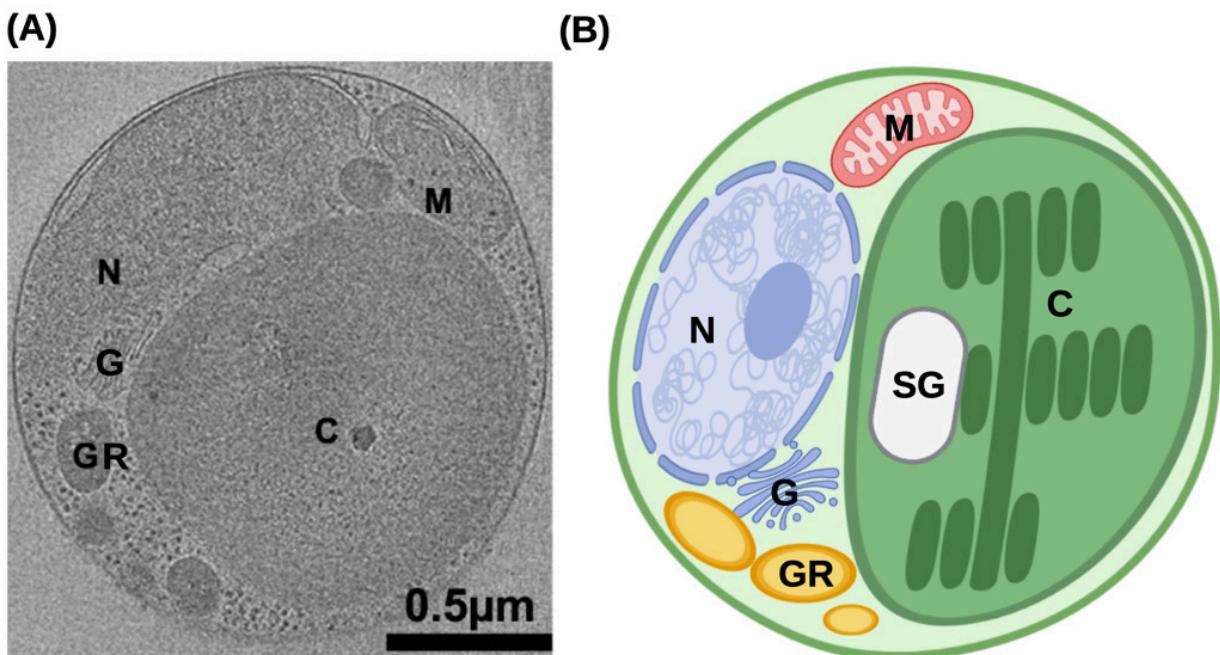
The green lineage (*Viridiplantae*) comprehends two of the most important groups of oxygen photosynthetic eukaryotes: green microalgae and their descendants, terrestrial plants. Microalgae are a very diverse group of photosynthetic microorganisms of special interest due to their plastic physiology and their biotechnological applications. They can be found in different water-based habitats, from freshwater to oceans, as well as land ecosystems, including deserts. Microalgae grow under a broad range of temperatures, salinities, pH and light intensities. More than 5000 species of microalgae have been identified in the oceans accounting for the production of 50 % of the oxygen necessary to sustain life on Earth (Field et al., 1998). Microalgae play a central ecological role as primary producers of biomass establishing the base of aquatic trophic chains. They have also been of great interest for the scientific community due to the large and yet increasing number of their biotechnological applications. Microalgae have been described as a high yield source of carbon compounds and good candidates to mitigate CO<sub>2</sub> emissions. The fixation of CO<sub>2</sub> is coupled with growth and biosynthesis of compounds of interest such as polysaccharides, lipids, vitamins and antioxidants. They are currently large-scale cultured to produce biostimulants in agriculture, health supplements, pharmaceuticals and cosmetics, as well as, successfully applied in wastewater treatment coupled with the fixation of atmospheric CO<sub>2</sub> (Abinandan et al., 2018; Borowitzka, 2013; Chen et al., 2019; García-Cubero et al., 2018).

Recently, the development of high-throughput sequencing is contributing to the clarification of the evolutionary history of the green lineage (Bachy et al., 2022; Becker & Marin, 2009; Benites et al., 2021; Leliaert et al., 2012; Merchant et al., 2007). Two clades are described to constitute *Viridiplantae*: *Streptophyta*, including *Embryophyta* or land plants and *Charophyta* (land plant's closest algal ancestors or evolutionary eldest algal sisters); and *Chlorophyta*, comprising the core chlorophytes (*Chlorophyceae* and *Trebouxiophyceae*) and the evolutionary eldest sisters in this clade and the entire green lineage: *Prasinophyta*, which include *Mamiellales* as *Ostreococcus* (Bachy et al., 2022; Leliaert et al., 2012; Tragin & Vaulot, 2019) (Fig. 6).



**Figure 6: Phylogenetic relationships among the orders in the green lineage or Viridiplante.** Figure obtained from (Leliaert et al., 2011). The tree topology is a composite on accepted relationships based on molecular phylogenetic evidence. Uncertain phylogenetic relationships are indicated by polytomies. The divergence times are rough approximations based on the fossil record and molecular clock estimates. Drawings illustrate representatives of each lineage: (1) *Acetabularia*, (2) *Pediastrum*, (3) *Chlorella*, (4) *Tetraselmis*, (5) *Picocystis*, (6) *Ostreococcus*, (7) *Micromonas*, (8) *Crustomastix*, (9) *Monomastix*, (10) *Pyramimonas*, (11) *Pycnococcus*, (12) *Pseudoscourfieldia*, (13) *Nephroelmis*, (14) *Prasinococcus*, (15) *Verdigellas* (a: general habit, b: individual cells in a gelatinous matrix), (16) *Mesostigma*, (17) *Chlorokybus*, (18) *Klebsormidium*, (19) *Chara*, (20) *Xanthidium*, (21) *Coleochaete*, (22) *Chaetosphaeridium*, (23) *Ranunculus*.

Initial biodiversity studies on marine phytoplankton posited the domination of oceans by the so-called red lineage conformed by diatoms and dinoflagellates while the green lineage was thought to have less importance in marine ecosystems, being confined to freshwaters and terrestrial environments (Worden et al., 2004). However, in the last decade, metabarcoding studies have unveiled the relevance of the green lineage in marine waters completing previous studies based solely on microscopic and traditional molecular techniques. Consequently, the cosmopolitan distribution of *Prasinophyta* has been unveiled, especially for the order *Mamiellales* (Collado-Fabbri et al., 2011; Demir-Hilton et al., 2011; Leconte et al., 2020; Tragin & Vaulot, 2019; Worden et al., 2004). *Ostreococcus tauri* is classified as a mamiellale picoalga (Fig. 6) and its contribution to the marine phytoplankton is crucial in a wide range of oceans and seas all over the world (Benites et al., 2021; Bolaños et al., 2020). As examples of *Ostreococcus* cosmopolitan character, it is worth mentioning that different strains have caused blooms in the Atlantic and Pacific Oceans (O'Kelly et al., 2003; Worden et al., 2004) and that a specific *Ostreococcus* strain has been found to be the most prevalent *Mamiellale* in the Mediterranean Sea (Tragin & Vaulot, 2019).



**Figure 7: *Ostreococcus tauri*.** (A) Tomographic 2D section from an *Ostreococcus tauri* cell from (Henderson et al., 2012). (B) An illustrated diagram of an *Ostreococcus tauri* cell (created using BioRender) based on the previous 2D segmentation. Labels: nucleus N, chloroplast C, Golgi body G, mitochondrion M, starch granule SG and granules GR.

Beyond its ecologically important role (Chapman, 2013; Worden et al., 2004) and cosmopolitan presence in marine environments, *Ostreococcus tauri* key feature consists in being considered

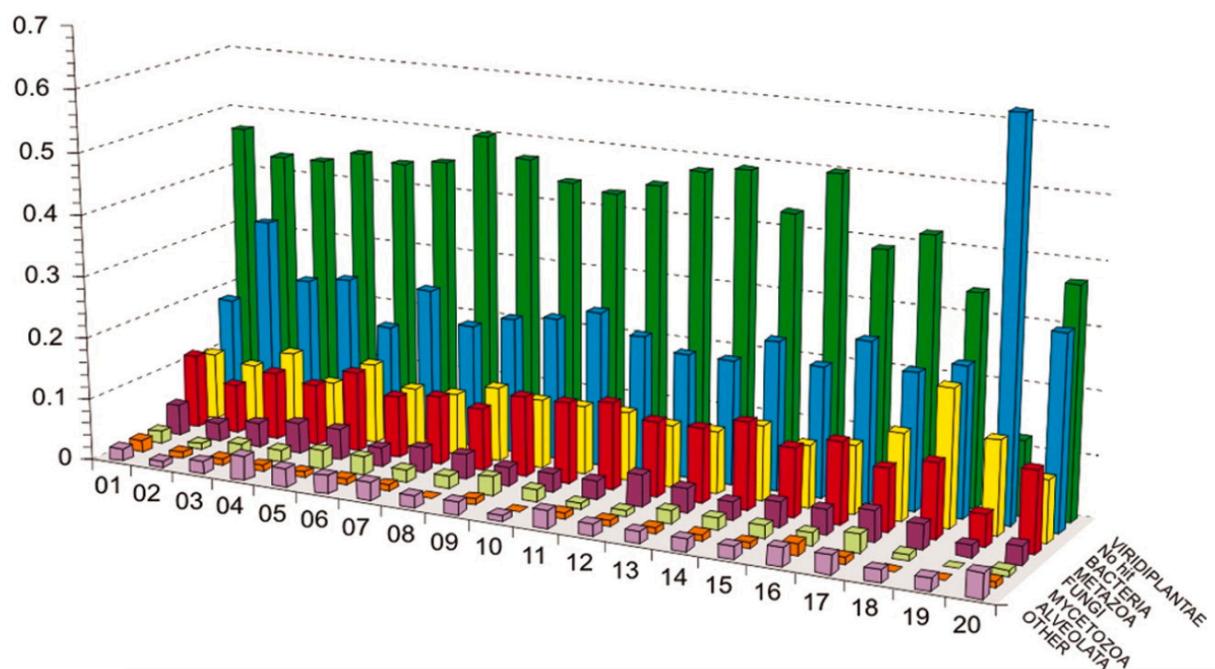
the world's smallest free-living eukaryote known to date (around 1 µm). Due to its small size, *Ostreococcus tauri* have been “invisible” to field researchers for a long time. It was first detected by flow cytometry as a rounded microalgae in a bloom that took place on the Mediterranean French lagoon, Étang de Thau, known for its semi-intensive farming of oysters. Hence, it was referred to as an “oysters coccus from the Thau lagoon” or *Ostreococcus tauri*. Its subcellular structure was resolved consisting of a nucleus, a single mitochondrion, a single chloroplast containing one starch granule and a very reduced or almost non-existent cytoplasmic compartment (Fig. 7) (Moreau, et al., 1995).

For a very long time, genome sequencing studies to understand marine phytoplankton were mainly focused on prokaryotes (Berube et al., 2018; Palenik et al., 2003). *Ostreococcus tauri* genome was not sequenced for the first time until 2006 (Derelle et al., 2006). From that point on, an increasing number of picoeukaryotes genomes were sequenced, including a second version of *Ostreococcus tauri* genome (Blanc-Mathieu et al., 2014) with 12.56 Mb arranged into 20 chromosomes. These contributions allowed the identification of unique genomic features, showing that its simplicity goes beyond its reduced size and cell structure. With a similar genome size, *Saccharomyces cerevisiae* has around 6275 protein coding genes, while *Ostreococcus tauri* has 8166 (Blanc-Mathieu et al., 2014; Derelle et al., 2006; Engel et al., 2014), making it the most gene dense free-living eukaryote known to date (Table 1).

**Table 1: Genomic features of different green lineage model organisms.** Their genome size and protein coding genes (Blaby et al., 2014; Blanc-Mathieu et al., 2014; Craig et al., 2021; Derelle et al., 2006; Hori et al., 2014; Lamesch et al., 2012; Swarbreck et al., 2008; M. Yang et al., 2018) are compared with their number of predicted transcription factors. (Rayko et al., 2010; Zheng et al., 2016)

	Genome size (Mb)	Number of protein coding genes.	Number of transcription factors.
<i>Arabidopsis thaliana</i>	135	27474	1779
<i>Klebsormidium nitens</i>	104	17055	286
<i>Chalmydononas reinhardtii</i>	110	14000	279
<i>Phaeodactylum tricornutum</i>	27.4	10567	212
<i>Ostreococcus tauri</i>	12.56	8166	102

This simplicity is also clear when genome size, number of genes and transcription factors in *Ostreococcus tauri* is compared to other photosynthetic species such as *Arabidopsis thaliana* (*Embryophyta*), *Klebsormidium nitens* (*Charophyta*), *Chlamydomonas reinhardtii* (*Chlorophyta*) and *Phaeodactylum tricornutum* (*Heterokontophyta*) (Table 1). Other key genomic characteristics in *Ostreococcus tauri* are its short intergenic regions and reduced gene family sizes comprising mostly single genes or even merging different genes into a single one. These features contribute to its intense degree of genome compaction, with the only exception of a long internal duplication on chromosome 19, hypothesized to be of recent origin due to its lack of divergence. There are more special characteristics in chromosomes 19 and 2. They contain 77 % of the transposable elements of the genome, have lower G+C content and a different codon usage (Blanc-Mathieu et al., 2014; Derelle et al., 2006). The first hypothesis about these chromosomes were that they had a different origin than the rest of the genome being the product of horizontal gene transfer events. Currently, only chromosome 19 is considered an “alien” chromosome since most of its protein coding genes are not related to the green lineage. However, chromosome 2 protein coding genes are essential housekeeping genes not duplicated and related to the green lineage (Fig. 8).



**Figure 8: Taxon distribution of best hits for genes from each *Ostreococcus tauri* chromosome.** Figure from (Derelle et al., 2006). The number corresponding to each chromosome is represented in the x axis, the percentage of best hits is represented in y axis. Each color stands for: viridiplantae, green; no hit, blue; bacteria, yellow; metazoa, red; fungi, purple; mycetozoa, light yellow; alveolates, orange; others, light purple.

Alternatively, this chromosome has recently been considered as a sex-related or mating-type chromosome (Benites et al., 2021; Blanc-Mathieu et al., 2014). Although sex is now accepted as a ubiquitous and ancestral feature of eukaryotes (Sekimoto, 2017; Swanson et al., 2011), direct observation of sex is still missing in most unicellular eukaryotic lineages. The genomic regions, so-called mating-type, involved in sex have been characterized in other *Chlorophyta* (Sekimoto, 2017) and, recently, in *Ostreococcus tauri*, which appears to encode two highly divergent haplotypes. These *Mamiellales* mating-type candidate regions are likely to be the oldest mating-type loci described to date (Benites et al., 2021; Leconte et al., 2020). Overall, *Ostreococcus tauri* is proposed as a novel model organism due to its structural and genomic features. In addition, inside the green lineage there is a wide diversity which makes difficult to find a model organism able to represent the entire lineage (Cock & Coelho, 2011). *Ostreococcus tauri* taxonomical classification as the eldest sister in *Prasinophyta* makes it an ideal organism to potentially represent the green lineage eldest ancestor (Derelle et al., 2006; Leliaert et al., 2012). Therefore, *Ostreococcus tauri* molecular systems could be considered as representative of the fundamental systems in the green lineage and the knowledge gained from them could be somehow extrapolated to a wide range of photosynthetic organisms.

Research in Molecular Systems Biology often faces challenges produced by the complexity of the model organisms used in the corresponding field and the high economical and computational costs associated to the generation and analysis of the necessary massive amounts of omics data (De Keersmaecker et al., 2006; Jamers et al., 2009; Joyce & Palsson, 2006; Weckwerth, 2011). In this respect, due its genomic simple features *Ostreococcus tauri* is an ideal model organism in Molecular Systems Biology when compared with other models in the green lineage such as *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* (Table 1) (de los Reyes et al., 2017; Derelle et al., 2006; Krumholz et al., 2012; Le Bihan et al., 2011; Lelandais et al., 2016). Indeed, in this thesis, *Ostreococcus tauri* is used to generate massive multiomic data that is analyzed and interpreted to characterize responses to seasonal variations in diel cycles in potentially fundamental systems in the green lineage.

# Molecular Systems Biology

There is an ancient Indian fable that illustrates how a holistic view and ontological reasoning contributes to knowledge, it is called “Blind men and an elephant”. A poem of John Godfrey Saxe is one of the most famous written versions of it:

“ It was six men of Indostan  
To learning much inclined,  
Who went to see the Elephant  
(Though all of them were blind),  
That each by observation  
Might satisfy his mind.

The First approached the Elephant,  
And happening to fall  
Against his broad and sturdy side,  
At once began to bawl:  
“God bless me! but the Elephant  
Is very like a wall!”

The Second, feeling of the tusk,  
Cried, “Ho! what have we here  
So very round and smooth and sharp?  
To me ‘tis mighty clear  
This wonder of an Elephant  
Is very like a spear!”

The Third approached the animal,  
And happening to take  
The squirming trunk within his hands, Thus  
boldly up and spake:  
“I see,” quoth he, “the Elephant  
Is very like a snake!”

The Fourth reached out an eager hand,  
And felt about the knee.  
“What most this wondrous beast is like  
Is mighty plain,” quoth he;  
“This clear enough the Elephant  
Is very like a tree!”

The Fifth, who chanced to touch the ear,  
Said: “E’en the blindest man  
Can tell what this resembles most;  
Deny the fact who can  
This marvel of an Elephant  
Is very like a fan!”

The Sixth no sooner had begun  
About the beast to grope,  
Than, seizing on the swinging tail  
That fell within his scope,  
“I see,” quoth he, “the Elephant  
Is very like a rope!”

And so these men of Indostan  
Disputed loud and long,  
Each in his own opinion  
Exceeding stiff and strong,  
Though each was partly in the right,  
And all were in the wrong! (...)"

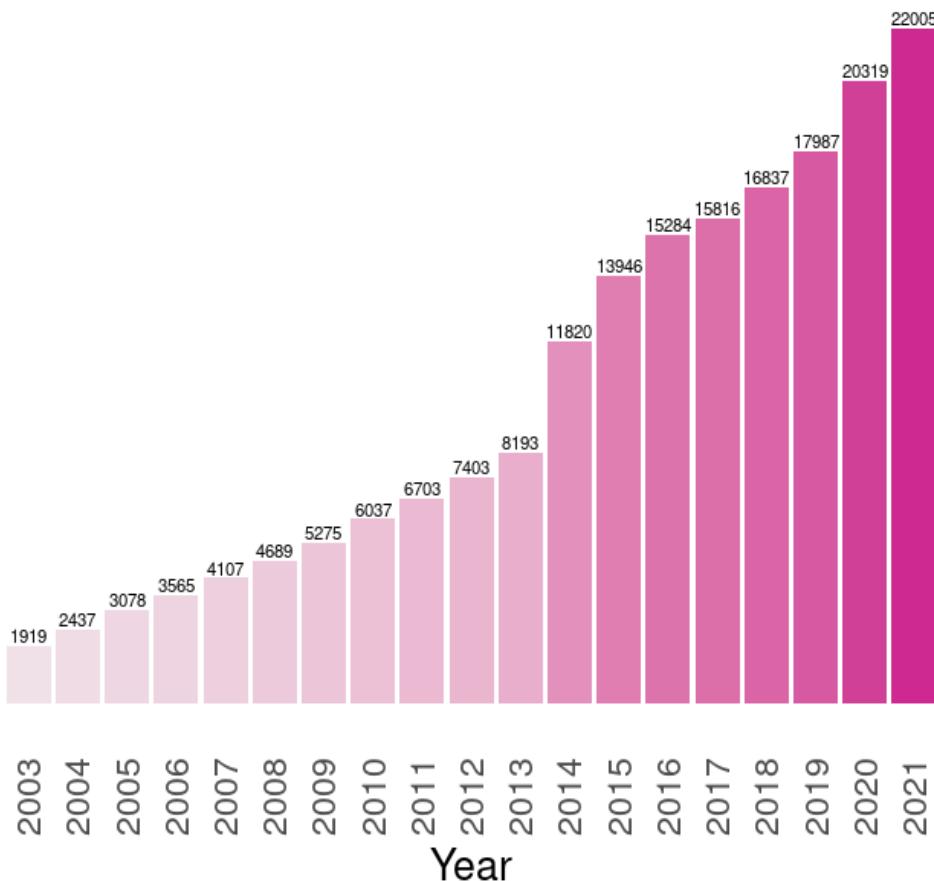
After touching different parts of the animal, one of the blind wise men concluded that the elephant was like a snake, he was touching only the trunk of the elephant. Another one concluded that it was like a fan since he was only touching the ear of the animal, and so on with the other blind wise men. Each one of them was confident about their individual findings and reaching an agreement between them was impossible. The subject of study was

so massive that they individually were only able to collect information from very specific regions that were treated as independent and contradictory truths instead of considering them as complementary parts of a whole.

Until recently, studies producing holistic views of living organisms were impossible to carry out due to the lack of suitable technologies. The development of omics technologies together with advances in computational science and artificial intelligence have made possible such systemic approaches in biology (Ideker et al., 2001; Karahalil, 2016; Veenstra, 2021; Weckwerth, 2011). In the last 50 years, sequencing methods have exponentially improved. First generation DNA sequencing methods were developed around 45 years ago shortly anticipating the development of one of the biggest international research projects in History, the Human Genome Project, which started about 20 years ago (Ideker et al., 2001; Veenstra, 2021). In this project an extremely ambitious goal was pursued: the sequencing of the whole human genome. However ambitious, its impact in science was beyond expected transcending the initial purpose. Different groups cooperated all over the world to sequence the complete human genome which promoted the development of new sequencing methods in order to make the process easier and cheaper. These methods are now applicable to study the molecular systems codified in the genome of living organism spanning the entire tree of Life (Abascal et al., 2020). During the Human Genome Project, 13 years were needed to sequence a single complete human genome and the overall costs were 2.7 billion dollars. Currently, with the emergence of Next Generation Sequencing (NGS) methods, around 100000 human genomes have been sequenced in the last 6 years. The Illumina HiSeq System generates around 500 gigabase sequences per second, dropping the cost of sequencing a complete human genome to \$1500 in only 10 years (Prendergast et al., 2020; Veenstra, 2021). Ultimately, these methods have changed how scientific projects are approached in biology. These new approaches have been applied to sequence, identify and quantify complete transcriptomes (total mRNA transcripts from samples). This is generating unprecedented massive amounts of biological information. Although, transcriptomic studies are extremely informative they only focus on one layer of the complexity in biological systems leaving out the actual molecular components that perform most cellular functions, the proteins. This has promoted the development of methodologies aiming to the identification and quantification of the complete proteome (total set of proteins from biological samples). Moreover, in order to obtain a more complete picture of biomolecular systems, methods identifying and quantifying specific metabolites need to be explored developing targeted metabolomics. Although these molecular components are

identified and quantified independently by different methods, they all highly interact with each other establishing different positive and negative regulatory loops underpinning cellular physiology. Researchers in Life Sciences are beginning to approach the study of living organisms and processes as systems with interacting modules comprising genes, transcripts, proteins and metabolites that constitute large molecular networks establishing the field of Molecular Systems Biology (Joyce & Palsson, 2006; Veenstra, 2021; Weckwerth, 2011). The typical methodology in Molecular Systems Biology starts, thus, with the generation of massive amounts of omic data and its subsequent integrative analysis. The results obtained from computational analysis based on statistical and Artificial Intelligence methods must be validated by independent wet lab experiments. The final expected result is the development of predictive models enabling researchers to anticipate the physiological response of the system under specific conditions (Jammers et al., 2009; Veenstra, 2021; Weckwerth, 2011; Zurbriggen et al., 2012).

In this sense, this thesis aims at providing a molecular systems biology characterization of the responses in the prasinophyceae microalgae *Ostreococcus tauri* to seasonal variations in diel cycles. Accordingly, generation and integrative analysis of massive amounts of data from two different omics techniques, transcriptomics and proteomics, have been performed. Specifically, the transcriptomic method used in this thesis was RNA sequencing, RNA-seq, the up to date method used for this omic. RNA-Seq analyses identify and quantify a wide variety of different classes of RNAs without requiring transcript-specific probes (Ditz et al., 2021; Veenstra, 2021; Wang et al., 2009), which has almost driven obsolete microarrays, the previous method used in transcriptomic analyses. Respectively, the field of proteomics have been directly connected with the development of mass spectrometry (MS) technology. During the first part of this century, there were two lines of development working separately for what would be known as proteomics today: identification (2-D electrophoresis gel) and quantification (using isotope tags) of proteins. Nowadays, proteomic technologies have enabled the development of proteins identification and quantification methods so the number of proteins that can be quantified/identified today is around several thousands (Shen et al., 2022; Veenstra, 2021). Specifically, the proteomic technology applied in this thesis was SWATH-MS, Sequential Window Acquisition of all Theoretical fragment ion spectra Mass Spectrometry. SWATH-MS is a data independent acquisition (DIA) method in which liquid chromatography is coupled with two runs of mass spectrometry (MS) for protein identification and label free relative quantification (Chen et al., 2021; Ludwig et al., 2018).



**Figure 9: Publications on systems biology.** Barplot representing the number of publications using the term “systems biology” in PubMed since the year when the Genome Human Project was completed (2003). The number on the top of each bar corresponds to the exact number of publications of that year.

While this holistic view of biological systems is gaining strength growing steadily during the first decades of the XXI century (Fig. 9), traditional reductionist methods focusing on individual molecular components are still between the most popular ones because scientific projects in Molecular Systems Biology are limited by the time, economic resources and the development of software tools to analyze and integrate massive amounts of biological data (Karahalil, 2016; Mazzocchi, 2012; Veenstra, 2021). However, the sequencing of an increasing number of microalgae genomes (Blanc et al., 2012; Bowler et al., 2008; Cheng et al., 2019; Derelle et al., 2006; Hori et al., 2014; Merchant et al., 2007; Morimoto et al., 2020; Prochnik et al., 2010) has allowed the application of omics technologies: transcriptomics (de los Reyes et al., 2017; Ditz et al., 2021; Hoys et al., 2021; Serrano-Pérez et al., 2022), proteomics (Le Bihan et al., 2011; Weckwerth, 2011) and metabolomics (Hoys et al., 2021; Serrano-Pérez et al., 2022; Weckwerth, 2011) in microalgae research, although here is a lack of specific software tools for microalgae to analyze and interpret omics data. In order to

contribute to resolve this deficiency in the microalgae research community, a web app is developed in this work ([Romero-Losada et al., 2022](#)), allowing researchers without expertise in computational data analysis to process omic raw data.





# **Materials and Methods**



# Organism and culture conditions

## Organism and growth medium

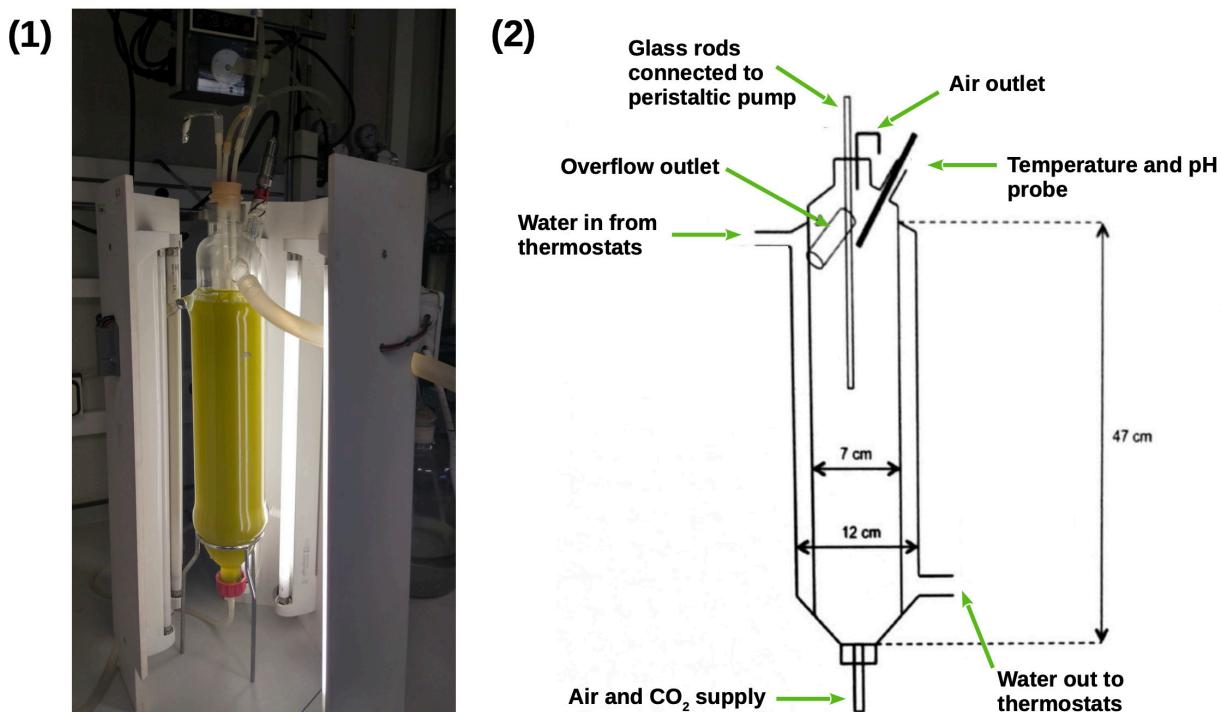
The sequenced strain of *Ostreococcus tauri*, RCC4221, was used in this study. Cells were grown photoautotrophically on sterilized artificial sea water (ASW) (Kester et al., 1967) supplemented with nitrates, phosphates, trace metal and vitamins. Components and concentrations in the medium are described in Table 2.

**Table 2: Components supplementing ASW as *Ostreococcus tauri* culture medium.**

Concentration in medium (mg L <sup>-1</sup> )
222.2 NaNO <sub>3</sub>
1.6 Na <sub>2</sub> HPO <sub>4</sub>
5.6 K <sub>2</sub> HPO <sub>4</sub>
3.0 NH <sub>4</sub> Cl
5.8 Fe-EDTA
41.3 Na <sub>2</sub> -EDTA
2.6·10 <sup>-2</sup> ZnSO <sub>4</sub>
1.6·10 <sup>-2</sup> CoSO <sub>4</sub>
8.9·10 <sup>-2</sup> Na <sub>2</sub> MoO <sub>4</sub> · 2H <sub>2</sub> O
2.8·10 <sup>-3</sup> CuSO <sub>4</sub>
1.9·10 <sup>-2</sup> H <sub>2</sub> SeO <sub>3</sub>
2.0·10 <sup>-2</sup> MnCl <sub>2</sub> · 4H <sub>2</sub> O
0.2 Thiamin-HCl
1.7·10 <sup>-3</sup> Biotin
1.7·10 <sup>-3</sup> Cyanocobalamin

## Continuous culture in photochemostats

Continuous culture was performed in 2.0 L capacity (0.07m diameter, 0.50 m height) jacketed sterilized photochemostat (bubble columns) containing 1.8 L of cells suspension, continuously sparged with air (1L (L culture<sup>-1</sup>) h<sup>-1</sup>) (Fig. 10). Culture conditions (pH, dilution rate and illumination regime) were constantly measured and computationally controlled by a LabJack. Temperature was maintained at 20°C, and pH at 8 by on demand injection of pure CO<sub>2</sub> into the air stream entering the culture.



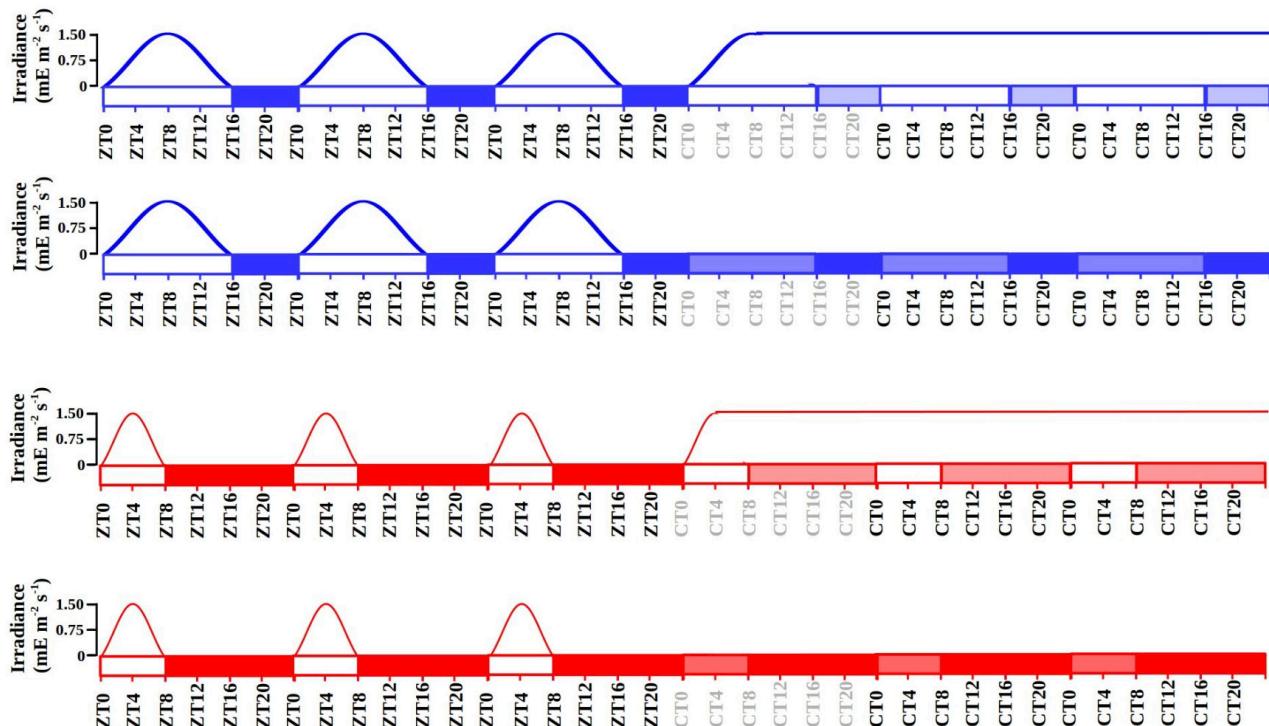
**Figure 10: Growth system used for continuous cultures.** (1) Picture of one of the photochemostats used; (2) Schematic detailed design of photochemostats.

Six Phillips PL-32W/840/4p white-light lamps, which provided a maximal incident irradiance of  $1500 \mu\text{E m}^{-2} \text{s}^{-1}$  on the reactor surface, were used as light inputs. Initially, the reactors were inoculated with batch-grown cells and operated on batch mode for about 3–4 days, with incident irradiance being progressively increased, until the culture reaches the stationary phase. Then, the reactors were switched to operate on continuous mode, fresh medium being continuously fed during the light period at a flow rate of  $45 \text{ mLh}^{-1}$  (dilution rate ( $D$ ),  $0.3 \text{ d}^{-1}$  under long day condition and  $0.15 \text{ d}^{-1}$  under short day condition), with withdrawal of culture at the same rate. Once steady state was achieved, samples were collected and analytical determinations were performed. Light was gradually increased by the LabJack controlled system until the maximal irradiance was reached, simulating the natural photoperiod (Fig. 11). Furthermore, photochemostats were surrounded by a wooden box and a completely opaque fabric to avoid any external light input.

## Experimental design

Our experiments were focused on two extreme photoperiods: summer long day condition (LD, 16 h light : 8 h dark) and winter short day condition (SD, 8 h light : 16 h dark). The experimental design (Fig. 11) consisted of three consecutive

days under each photoperiod, LD or SD. Cells were harvested at specific times during diel cycles, represented as zeitgeber time (ZT), where ZT0 corresponds to dawn, ZT4 to 4 h after dawn and so on. Samples were taken every 4 h (from ZT0 to ZT20) during the three days of alternating light/dark cycles. Therefore, a total of 6 samples for each day of sampling were generated. *Bona fide* circadian genes are considered those presenting rhythmicity under both LD and SD as well as maintaining their rhythmic expression profiles under free-running conditions. For the identification of *bona fide* circadian genes, three days under free-running conditions (constant light or constant dark) are also included in the experimental design of the transcriptomic analysis. No samples were collected during the first day to allow culture acclimation, and samples were collected thereafter for two consecutive days every 4 h starting at CT0, Circadian Time 0, the time point corresponding to subjective dawn.



**Figure 11: Schematic description of the experimental design.** Gradually increased and decreased irradiance is represented during each light phase. Long day condition is represented in blue and short day condition in red. Photoperiods (light periods) correspond to white rectangles and skotoperiods (dark periods) to blue/red filled rectangles. Light blue and red filled rectangles are used to represent subjective photoperiods and skotoperiods under free-running conditions. ZTN, Zeitgeber Time N, marks the time point N hours after dawn (lights on). CTN, Circadian Time N denotes the time point N hours after the subjective dawn.

# Transcriptomic analysis

## Sample collection

From each chemostat, a volume of 50 mL of cell suspension were harvested (4 min centrifugation at 5000 x g and 4 °C) for each time point. Pellets were washed using Phosphate-buffered saline solution (PBS) and flash frozen with liquid Nitrogen before stored at –80 °C.

## Cell disruption

Frozen pellets were resuspended in 400 µL of disruption buffer ([García-Domínguez & Florencio, 1997](#)) and directly added to a 1.5 mL Eppendorf tube (RNase free and phenol-proof) containing 400 µL of phenol:chloroform 1:1 and 100 µL of acid washed glass beads (0.25–0.3 mm diameter). Mechanical disruption was performed by 30 min of repeated cycles of 60 s of vortexing and 60 s of incubating on ice.

## RNA extraction

Extracts were centrifuged (4 °C) for 15 min at 13000 x g, producing three differentiated phases: an upper aqueous phase containing RNA, a white interphase containing DNA and a lower organic phase containing proteins, lipids and glass beads. The upper aqueous phase was collected, mixed with 400 µL of phenol:chloroform 1:1 and centrifuged for 5 min (4 °C at 13000 x g). This process was repeated three more times. In the last wash, only chloroform was used to avoid phenol contamination of the RNA samples. The supernatant was incubated overnight at -20 °C in a solution of 80 µL 10 M LiCl and 550 µL 100 % EtOH for RNA precipitation. Finally, samples were centrifuged 10 min at 13000 x g 4 °C. Pellets were dried to avoid EtOH contamination.

## RNA purification

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 final RNA concentration and integrity were measured using a bioanalyzer 2100 (Agilent RNA 6000 Nano Kit).

## RNA sequencing and processing

Library was prepared in accordance with the manufacturer's instructions and the sequencing was carried out on the Illumina NextSeq500 sequencer. Approximately, 10 million 75 nt long single end reads were generated for each sample. The *Ostreococcus*

*tauri* genome sequence and annotation v3.0 were used as reference genome ([https://phycocosm.jgi.doe.gov/Ostta4221\\_3/Ostta4221\\_3.home.html](https://phycocosm.jgi.doe.gov/Ostta4221_3/Ostta4221_3.home.html)).

Further computational analyses were carried out using the software tools developed in this thesis, ALGAEFUN with MARACAS (microalgae FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq AnalysiS) (Romero-Losada et al., 2022).

## Proteomic analysis

### Sample collection

Sample collection for proteomic analysis followed the same procedure as the one previously described for transcriptomic analysis.

### Cell disruption

Directly onto frozen pellets, 1 mL of TRIsure™ Reagent (Sigma-Aldrich), 100 µL of acid washed glass beads (0.25–0.3 mm diameter) and 40 µL of Protein Inhibitor Cocktail 25x (PIC, Sigma-Aldrich) were applied, followed by 3 disruption cycles (60 s agitation-60 s incubation on ice) using a Mini-Beadbeater (BioSpec Products).

### Proteins extraction

Proteins were extracted using TRIsure™ Reagent (Sigma-Aldrich), according to the manufacturer's instructions. The resulting proteins pellets were resuspended with 2 mL of 0.3 M guanidine solution in 95 % EtOH using 10 sonication cycles (30 s sonication - 30 s of incubating at 4 °C) in a Diagenode Bioruptor Pico Sonicator and then centrifuged at 4 °C during 5 min at 8000 x g. This washing process was repeated twice, followed by two additional washing steps using 90 % EtOH. The final pellets were resuspended in NH<sub>4</sub>HCO<sub>3</sub> 50 mM / 0.2 % Rapigest (Waters) and total proteins were quantified using Qubit system.

### Proteins digestion

First, 50 µg of proteins were incubated with dithiothreitol (DTT 4.5 mM) for 30 min at 60 °C. Then, iodoacetamide was added to a final concentration of 10 mM and incubated for 30 min, under total darkness at room temperature. Treatment with trypsin was done overnight at 37 °C in 1:40 trypsin:protein. Subsequently, formic acid was added and incubated at 37 °C for 1h. Finally, 2 % acetonitrile (v/v) were added to reach a concentration of the digested sample around 0.5 µg of protein/µl of solution.

## **SWATH acquisition**

### ***Equipment and data acquisition method.***

Proteomic data acquisition was performed using a label free quantification platform that employs independent data acquisition (DIA) called SWATH-MS (Sequential Windowed Acquisition of all THeoretical Mass Spectra (SWATH-MS) (Ludwig et al., 2018) using a time-of-flight (TOF) triple quadrupole hybrid mass spectrometer (5600 plus, Sciex) equipped with a nano electrospray source coupled to a nanoHPLC (Eksigent, model 425). The Sciex software Analyst TF 1.7 was used for equipment control and data acquisition. Peptides were first loaded onto a trap column (Acclaim PepMap 100 C18, 5 µm, 100 Å, 100 µm id×20 mm, Thermo Fisher Scientific) under isocratical order in 0.1 % formic acid / 2 % acetonitrile (v/v) at a flow rate of 3 µL/min for 10 min. Subsequently, they were eluted on a reversed-phase analytical column (Acclaim PepMap 100 C18, 3 µm, 100 Å, 75 µm id×250 mm, Thermo Fisher Scientific), coupled to a PicoTip emitter (F360-20-10-N-20\_C12 from New Objective). Formic acid 0.1 % (v/v) was used as solvent A and 2 % acetonitrile with formic acid 0.1 % (v/v) were used as solvent B. Peptides were eluted with a linear gradient of 5-35 % (v/v) of solvent B in 120 min at a flow rate of 300 nL/min. The source voltage was selected at 2600 V and the temperature was maintained at 100 °C. Gas 1 was selected at 20 PSI, gas 2 at zero, and curtain gas at 25 PSI. For proteins identification, Data Independent Acquisition DIA method was used. It consisted of a TOF-MS with a scan window of 400-1250 m/z (accumulation time of 250 ms) followed by 50 MS/MS with a scan window of 230-1500 m/z (accumulation time of 65 ms) and with a cycle time of 2574 s.

## ***Library construction***

Spectral libraries for LD and SD conditions were constructed by making one run with a mixture of the biological replicates corresponding to each time point. ProteinPilot v5.0.1 software (Sciex) was used to identify the proteins in the library. A pooled search of all runs was performed. The parameters of the Paragon method were: trypsin as enzyme and iodoacetamide as cysteine alkylating agent.

The *Ostreococcus tauri* annotated proteome v3.0 file from ORCAE (Sterck et al., 2012) linked to a Sciex Contaminants database were used in library construction. A false positive analysis (FDR) was performed and those with FDR ≤ 0.01 were considered.

## ***SWATH runs***

For each sample, the equivalent of 1 µg of digested protein was injected into each run. Previously, the equipment was self-calibrated using a standard, MS synthetic peptide calibration kit from Sciex, to control sensitivity and chromatographic conditions. Protein identification and quantification were performed using SWATH runs with 60 ms of accumulation time and 3.7 s of cycle time. Three technical replicates for each one of the three biological replicates were analysed resulting in 9 replicates per time point.

## ***Data processing***

The generated libraries (1 % FDR) were analyzed using the Sciex software PeaKView 2.2 with the microapp SWATH 2.0, together with the data obtained from the SWATH runs. Using this software, the chromatographic traces of the ions were extracted and dumped into the Marker view 1.2.1.1 programme where the list of identified proteins with their corresponding areas were generated. The parameters for extraction of ions and areas were: 10 peptides per protein, 7 transitions of each peptide, threshold of confidence of the peptides set at 90 and FDR 1 %. The software NormalizerDE ([Willforss et al., 2019](#)) was used to test several normalization methods in order to probe which one achieved the minimum replicate variation relative to Log2. Quantile normalization was selected as normalization method based on this comparison. Data were imputed with mean imputation method between the nine technical and biological replicates, which means that the missing value on a certain variable is replaced by the mean of the available cases.

## **Cell cycle analysis**

### **Sample collection and cell fixation method**

A volume of 1.5 mL of cell suspension was harvested for each time point. These samples were diluted 1:10 in PBS to ensure a suitable cell concentration for the assay. Two mL of these dilutions were centrifuged and cells in the pellets were fixed with 10 mL of 100 % EtOH before stored at -20 °C for, at least, 24 h.

### **Cell staining method**

After fixation, cell suspensions were centrifuged for 5 min at 3500 x g (room temperature) and resuspended in 1 mL of PBS, washed once with PBS and sonicated for 3 minutes in an Ultrasonic Cleaner (JSP, US21, ultrasonic power 50W), in order to eliminate cell clumps and aggregates before staining. In the staining process, 2 µL of the Vibrant Dye Cycle Green

(V35004, Thermo Fisher) (10 µM final stain concentration) were added to each sample and incubated 30 min (37 °C) for selective DNA labeling. After incubation, cells were washed and transferred to flow cytometry tubes for cell cycle analysis.

## Data acquisition and processing

Flow Cytometry acquisition was performed with a BD FACS Canto II (BD Biosciences) where stained DNA were excited by a 488 nm laser and emission was collected in a 530/30 nm PMT. Flow rate was low and linear amplification were established for the acquisition. Data were analyzed using FlowJo v.10.6.1 (Becton Dickinson & Company BD). Analysis was performed using one of the univariate cell cycle methods provided by FlowJo, specifically the Watson pragmatic algorithm (Watson et al., 1987) to adjust the data to the model.

## Confocal fluorescence imaging

Sample preparation and staining method was executed as described for flow citometry samples. The Leica SP2 Laser Confocal Microscope was utilized to capture confocal fluorescent images of *Ostreococcus* cells. Image acquisition was performed using a 488 nm laser for the excitation of the dye employed for DNA labeling. Uniform adjustments were applied with repetitions to ensure consistency and reproducibility.

## Determination of photosynthetic activity

### Sample collection

Fresh culture was harvested at the different specific times of the day. The samples were diluted 1:1 with growing medium and incubated at 20 °C in total darkness during 10 min.

### Data acquisition

In order to determine photosynthetic parameters, Pulse-Amplitude-Modulation PAM fluorometry measurements were performed using a Waltz DUAL-PAM-100.

After darkness incubation, the non-actinic modulated light (450 nm, 2.8 µE m<sup>-2</sup> s<sup>-1</sup>) was turned on to measure  $F_o$  (fluorescence basal level). Then, to determine  $F_M$  (the maximum fluorescence level), a saturating red light pulse of 655 nm and 5000 µE m<sup>-2</sup> s<sup>-1</sup> was applied to the sample during 400 ms. The  $F_v/F_M$ , that corresponds to the maximum potential quantum efficiency of Photosystem II when all reaction centers were open, was calculated as:

$$\frac{F_v}{F_M} = \frac{(F_M - F_o)}{F_M}$$

# Analytical determinations

## Sample collection

At each time point, 50 mL of fresh culture were harvested and centrifuged at 7000 x g 10 min. Then pellets were washed with 1 % ammonium formate (p/v) to eliminate salts from the growing medium and lyophilized.

## Starch content

### Cell disruption

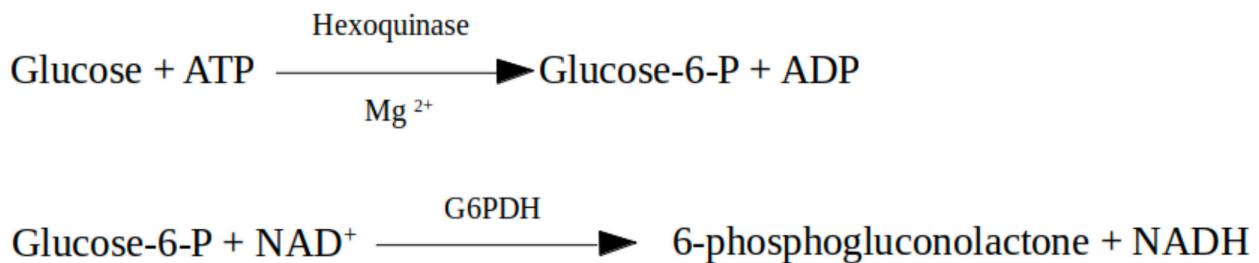
Approximately 2-3 mg of lyophilized biomass were added to hermetic tubes containing 1 mL of glass beads (0.25–0.3 mm diameter) and 2 mL of chloroform:methanol (2:1). Three disrupting cycles (60 s agitation - 60 s incubation on ice) were applied using Mini-Beadbeater (BioSpec Products). Then cellular extracts were separated from the beads and saved in new tubes. Cellular extracts were centrifuged for 4 min at 13000 x g and the supernatant was discarded. The addition of chloroform:methanol (2:1) and centrifugation steps were repeated until the pellets were white in order to ensure the elimination of pigments and lipids that could disturb the determination process. Finally, pigment free pellets were dried.

### Starch solubilization and digestion

The proposed protocol for plants in (Rufty & Huber, 1983) was adapted to *Ostreococcus tauri*. Starch granules in dry pellets were alkaline solubilized with 1 mL of 0.2 M KOH and heated at 100 °C. After 30 min, samples were gradually cooled and pH was adjusted to 5.0 adding 300 µL of 1 M acetic acid. To starch digestion, 7.4 U of  $\alpha$ -amylase were added and incubated 30 min at 37 °C, breaking down starch in small linear and branched oligosaccharides. After that, 5 U of amyloglucosidase were added and incubated 1-2 h at 55 °C releasing glucose residues. Finally, in order to stop enzymatic reaction, the samples were incubated at 100 °C during 2 min and centrifuged at 13000 x g for 10 min discarding pellet. Enzymes were prepared in 0.1 M of sodium acetate pH 4.5.

### Spectrophotometric quantification

The quantification of released glucose residues from starch was achieved following (Rufty & Huber, 1983) protocol. The method combined two enzymatic activities: hexokinase, that phosphorylated glucose residues, and glucose-6-phosphate dehydrogenase (G6PDH) that reduced NAD<sup>+</sup> oxidizing the phosphorylated glucose. NADH generated could be measured at 340 nm and correlated to glucose concentration in a ratio 1:1 (Fig. 12).



**Figure 12: Enzymatic activities that link NADH production to glucose residues released from starch.**

To achieve that measurement, quartz spectrophotometer cuvettes were used containing:

- 100 µL of the sample
- 500 µL of hexoquinase buffer (100 mM HEPES pH 7.7, 10 mM MgCl<sub>2</sub>, 0.04 % BSA p/v, 1 mM DTT)
- 100 µL of ATP mix (containing 10 mM ATP diluted in HEPES 100 mM pH 7.7)
- 100 µL of NAD<sup>+</sup> mix (containing 4 mM NAD<sup>+</sup> diluted in HEPES 100 mM pH 7.7)
- 2 µL of 2.5 U µL<sup>-1</sup> glucose-6-phosphate dehydrogenase
- 200 µL of distilled water.

The absorbance of that mixture was measured at 340 nm followed by the addition of 5 µL U of 1 µL<sup>-1</sup> hexoquinase and the incubation at 27 °C for 20 min. After that, a second measurement allowed to determine the amount of NADH produced during the reaction. NADH absorbance was related to the amount of glucose residues and consequently to the initial amount of starch in each the sample from calibration curves with commercial starch.

## Carotenoid content

### Cell disruption

Four milligrams of lyophilized biomass were added to a hermetic tube containing 1 mL of glass beads (0.25–0.3 mm diameter) and 1 mL of pure acetone. Three disrupting cycles (60 s agitation – 60 s incubation on ice) were applied using Mini-Beadbeater (BioSpec Products).

### Carotenoids extraction

Carotenoids extraction was achieved following the method proposed by (Del Campo et al., 2004). Darkness was maintained during the entire process to avoid pigments degradation. After centrifugation for 4 min at 13000 x g, cellular extracts were collected and saved in

new tubes. Again, 1 mL of pure acetone was added to wash glass beads, centrifuged and the supernatant was collected. This process was repeated until the supernatant turned colorless. Supernatants were pooled together in the same tube and acetone was evaporated using a stream of nitrogen gas. Finally, 350 µL of acetone were added for resuspension.

### ***Carotenoids determination and quantification***

A Hitachi HPLC (Elite LaChrom), equipped with a photodiode-array detector (Hitachi L-2455) was used. Separation was performed on a Waters NovaPak C-18 (3.9×150 mm, 4 µm particle size, 60 Å pore size) column. Following the method proposed by (Böhme et al., 2002), the eluents used to create a gradient through the mobile phase were: eluent A (0.1 M ammonium acetate and 15:85 v/v H<sub>2</sub>O-methanol) and eluent B (44:43:13 v/v methanol-acetonitrile-acetone). Temperature was maintained constant (20 °C) during the whole process and eluents flowed at 800 µL min<sup>-1</sup>.

Different carotenoids were identified based on the retention times and absorption profiles of carotenoids patterns previously analyzed.

### **Enzyme assay**

Enzymatic activities of Nitrate reductase (NR) and Glutamine synthetase (GS) were measured following (Herrero et al., 1981) and (García-Domínguez & Florencio, 1997), respectively. Activity is measured in units of enzymatic activity (U). One unit (1 U) corresponds to the amount of enzyme required to catalyze the production of 1 µmol of nitrite (for NR activity) or glutamine (for GS activity) per minute.

### **Rhythmic patterns analysis**

#### **Rhythmic patterns detection**

The non-parametric methods implemented in the R package RAIN (Thaben & Westermark, 2014) from Bioconductor were used to statistically identify rhythmic patterns in the different collected data. These rhythmic patterns were also independently identified using the cosinusoidal parametrizations developed in the R package CircaCompare (Parsons et al., 2020). Three complete diel cycles from both conditions (LD and SD) were used to detect rhythmic patterns in the different data of this study: expression levels of genes, abundance of proteins, maximum potential quantum efficiency of Photosystem II, cells in the different cell cycle phases, carotenoids and starch contents and enzymatic activities of nitrogen metabolism key enzymes. Rhythmic patterns with a single maximum point over a complete diel cycle were

detected by setting the period parameter from RAIN to 24 hours. Similarly, more complex rhythmic patterns (the ones with two or even three maximum points per day) were detected changing the period parameter from RAIN to 12 and 8 hours respectively. A 0.05 p-value threshold was used in all scenarios. In addition, the last two diel cycles and two consecutive days of continuous light or darkness were considered for the RAIN and CircaCompare analysis described above. In that way, RAIN and CircaCompare could statistically test if a similar pattern was maintained after changing the cycling light regime to free-running conditions consisting in constant light or darkness preventing a bias towards any of the two conditions.

## Rhythmic patterns comparison

The rhythmic patterns were fitted to a cosinusoidal wave in order to be characterized and to enable comparison between them. The statistical significance of the differences in these waves parameters between different of rhythmic patterns was performed using the R package CircaCompare (Parsons et al., 2020) with a p-value threshold of 0.05. The significance of the global differences in the different rhythmic parameters was performed using the Mann-Whitney-Wilcoxon non-parametric test implemented in the R function wilcox.test.

## Data and code availability

The RNA-seq transcriptomic data generated and analysed in this thesis have been deposited in the Gene Expression Omnibus (NCBI) database and are freely available to the research community identified with the accession number GSE155535. The SWATH-MS proteomics data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD046992.

The codes for MARACAS (MicroAlgae RnA-seq and Chip-seq AnalysiS) pipeline and AlgaeFUN (microALGAE FUNctional enrichment tool) are freely available from the following Github repositories:

- <https://github.com/fran-romero-campero/MARACAS>
- <https://github.com/fran-romero-campero/AlgaeFUN>

The actual R code developed in this thesis to perform all the analysis is available as a Rmarkdown report in SANDAL (SeAsonal aNd diel cycles in *Ostreoccocus*) Github repository:

- <https://github.com/fran-romero-campero/SANDAL>

The shiny app MINOTAUR (Multi-omics INtegration in *Ostreococcus TAURI*), developed to facilitate the exploration of the results presented in this thesis can be accessed at:

- <https://greennetwork.us.es/MINOTAUR/>

The code of MINOTAUR is available from the Github repository:

- <https://github.com/ABRomeroLosada/MINOTAUR>



# Hypothesis and Objectives



Seasonal variations in diel cycles play a central role in the physiology and development of living organisms on Earth. Therefore, it is expected that seasonality highly affects microalgae physiology in particular. In fact, seasonal oceanic massive microalgae blooms can be observed from outer space. Nevertheless, there is a lack of research in this field and the responses of microalgae to seasonal variations are not characterized yet using omics techniques. Specifically in *Ostreococcus tauri*, the publications are limited to a single transcriptomic analysis based on microarrays and an independent single proteomic study, both performed over neutral day condition (12 h of light : 12 h of dark). Most of the *Ostreococcus* transcriptome has been reported as rhythmic, but *bona fide* circadian genes maintaining rhythmicity under different seasonal conditions and free-running conditions are not identified yet. Moreover, comparative analysis between rhythmic patterns at the transcriptomic, proteomic and physiological levels have not been performed and are expected to shed light on the responses of living organisms to seasonal variations in diel cycles.

The general objective in this thesis consists on a multiomic characterization of the responses to photoperiodic seasonal variations in diel cycles in *Ostreococcus tauri*. This is achieved pursuing the following specific objectives:

1. To develop software tools, computational pipelines, and models to enhance systems biology studies in microalgae, with a specific focus on exploring the results presented in this thesis.
2. To characterize transcriptomic seasonal variations during diel cycles and free-running conditions.
3. To characterize proteomic seasonal variations during diel cycles and their integration with transcriptomic responses.
4. To characterize seasonal variations in relevant physiological processes during diel cycles and conduct multi-omics integration with transcriptomic and proteomic data.



# Results



# **Chapter 1.**

**ALGAEFUN with MARACAS,  
microALGAE FUNctional  
enrichment tool for  
MicroAlgae RnA-seq and  
Chip-seq AnalysiS**



# **Chapter 1. ALGAEFUN with MARACAS, microALGAE FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq AnalysisS**

In order to characterize the molecular systems regulating microalgae physiology, high throughput sequencing technologies have been recently applied obtaining the genome sequences of a wide range of microalgae (Blanc et al., 2012; Bowler et al., 2008; Cheng et al., 2019; Cortegiani Carpinelli et al., 2014; Hori et al., 2014; Merchant et al., 2007; Moreau et al., 2012; Morimoto et al., 2020; Ottesen et al., 2013; Palenik et al., 2007; Polle et al., 2017; Roth et al., 2017; Worden et al., 2009). This has promoted the emergence of molecular systems biology studies and the use of different omics such as transcriptomics, based on RNA-seq data (Hoys et al., 2021; Serrano-Pérez et al., 2022), and cistromics, based on ChIP-seq data (Ngan et al., 2015; X. Zhao et al., 2021) in microalgae. Nonetheless, the progress of this type of studies on microalgae is limited by the lack of freely available and easy-to-use online tools to analyze and integrate omics data. Processing of the massive amount of high-throughput sequencing data and analysis of the resulting sets of genes and genomic loci obtained from molecular systems biology studies requires computational power, time, effort and expertise that some research groups on microalgae may lack. In addition, researchers must explore different data bases separately, which makes the integration of the results and the generation of biological meaningful information difficult. Therefore, it is imperative the development of frameworks integrating microalgae genome sequences and annotations with tools for high-throughput sequencing data analysis and functional enrichment of gene and genomic loci sets. ALGAEFUN with MARACAS (<https://greennetwork.us.es/AlgaeFUN/>) has been developed to cover these needs of the microalgae research community and to promote studies in molecular systems biology.

This web portal consists of two different tools. First, MARACAS (MicroAlgae RnA-seq and Chip-seq AnalysisS) implements a fully automatic computational pipeline receiving as input RNA-seq or ChIP-seq raw data from microalgae studies and producing set of differentially expressed genes or lists of genomic loci, respectively. These results can be further analyzed using the second tool, ALGAEFUN (microAlgae FUNctional enrichment tool). When receiving the results from an RNA-seq analysis, sets of genes are functionally annotated by performing Gene Ontology (GO) (Ashburner et al., 2000; Carbon et al., 2019) and metabolic pathways enrichment analysis (Kanehisa et al., 2016; Moriya et al., 2007; Ogata et al., 1999).

When genomic loci from a ChIP-seq analysis are inputted, a set of potential target genes is generated together with the analysis of the distribution of the loci over gene features, as well as metagene plots representing the average mapping signal. This set of potential target genes can be further studied using the features for functional enrichment analysis in ALGAEFUN, as described above. The code for ALGAEFUN with MARACAS is publicly available at their respective GitHub repositories from the following links:

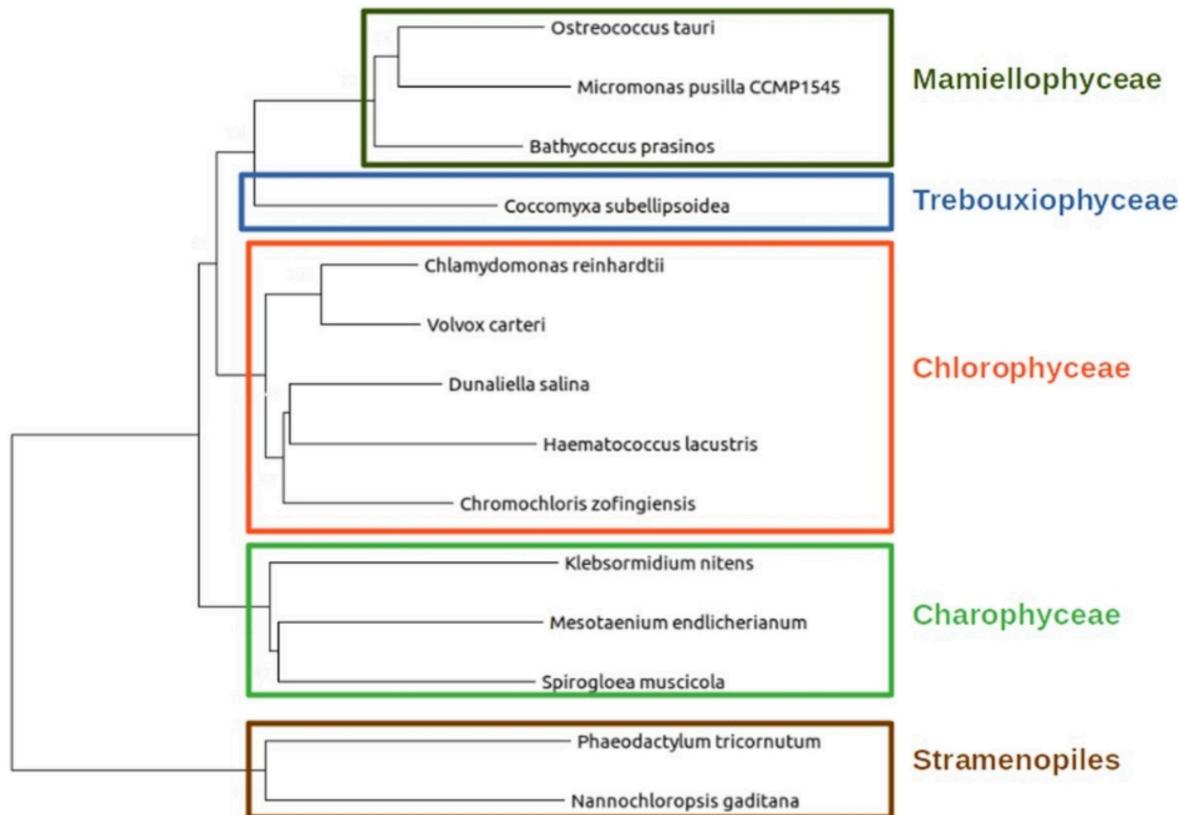
- <https://github.com/fran-romero-campero/ALGAEFUN>
- <https://github.com/fran-romero-campero/MARACAS>

## Implementation

### *Integration of different microalgae databases*

ALGAEFUN with MARACAS supports 14 different microalgae species that cover an ample spectrum of their phylogeny (Fig. 13): *Chlamydomonas reinhardtii* (Blaby et al., 2014; Merchant et al., 2007), *Volvox carteri* (Prochnik et al., 2010), *Chromochloris zofingiensis* (Roth et al., 2017), *Dunaliella salina* (Polle et al., 2017), *Haematococcus lacustris* (Morimoto et al., 2020) (Chlorophyceae), *Coccomyxa subellipsoidea* (Blanc et al., 2012) (Trebouxiophyceae), *Ostreococcus tauri* (Blanc-Mathieu et al., 2014; Palenik et al., 2007), *Bathycoccus prasinos* (Moreau et al., 2012), *Micromonas pusilla CCMP1545* (Worden et al., 2009) (Mamiellophyceae), *Phaeodactylum tricornutum* (Bowler et al., 2008; Yang et al., 2018), *Nannochloropsis gaditana* (Corteggiani Carpinelli et al., 2014; Radakovits et al., 2012) (Stramenopiles), *Klebsormidium nitens* (Hori et al., 2014), *Mesotaenium endlicherianum* (Cheng et al., 2019) and *Spirogloea muscicola* (Cheng et al., 2019) (Charophyceae). Supported species include microalgae used in basic scientific research, as well as those used in biotechnology industry like *H. lacustris* (Hoys et al., 2021), *N. gaditana* (Ajjawi et al., 2017) or *P. tricornutum* (Cui et al., 2019; Pereira et al., 2021).

One of the goals of this tool is to integrate available genome sequences, functional annotations and genomic feature annotation files (Gene transfer files, GTF) for the already sequenced microalgae species in order to generate easily accessible resources. These data are systematically collected from different freely available data bases, depending on the microalgae, Table 3. Specifically, for *N. gaditana* and *P. tricornutum*, Ensembl protist (Howe et al., 2021) is accessed, a web based unicellular species genome browser storing gene annotation; for *C. reinhardtii*, *V. carteri*, *C. zofingiensis*, *D. salina* and *C. subellipsoidea* the JointGenome



**Figure 13: Phylogenetic relationship between the different microalgae species supported in ALGAEFUN with MARACAS.**

Institute (JGI) / Phytozome (Goodstein et al., 2012), a web portal integrating omics for photosynthetic organisms is queried; for *M. endlicherianum* and *S. muscicola* a figshare associated to the corresponding publication is accessed; *M. pusilla*, *O. tauri*, *B. prasinos* and *K. nitens* genome sequence and annotation are downloaded from the JGI / PhycoCosm (Grigoriev et al., 2021), a comparative algal genomic resource; finally the genome of *H. lacustris* is available at NCBI genome database ([https://www.ncbi.nlm.nih.gov/genome/67110?genome\\_assembly\\_id=839686](https://www.ncbi.nlm.nih.gov/genome/67110?genome_assembly_id=839686)).

**Table 3: Resources used to collect genome sequences, functional and gene feature annotations for each supported microalga.**

Ensembl protists	PhycoCosm	Phytozome	Genomes NCBI datasets	Figshare associated to publication
<i>N. gaditana</i>	<i>B. prasinos</i>	<i>C. reinhardtii</i>	<i>H. lacustris</i>	<i>M. endlicherianum</i>
<i>P. tricornutum</i>	<i>K. nitens</i>	<i>V. carteri</i>		<i>S. muscicola</i>
	<i>M. pusilla</i>	<i>C. zofingiensis</i>		
	<i>O. tauri</i>	<i>D. salinas</i>		
		<i>C. subellipsoidea</i>		

Genome sequence and gene feature annotation files are downloaded in fasta and GTF format, respectively. When necessary, different chromosome and/or scaffold files are colligated programmatically to produce a single genome file. The GTF format in the gene feature annotation files consists of a data frame with nine columns. Each line corresponds to a specific gene feature. The first eight columns must contain information related to the gene feature (3' UTR, 5' UTR, gene, CDS or mRNA), chromosome start and end positions, the strand where it is positioned, and some other attributes in a well-defined format. The ninth column is not restricted to a specific format and can contain any type of information. Nonetheless, the mappers used in MARACAS assume that this last column follows the format present in GTF files from the data base Ensembl. In order to be able to use GTF files from other databases besides Ensembl, an R script to translate any GTF file into the format followed by Ensembl is developed.

Systematic functional annotation files consisting of Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genome) Orthology (KO) terms are also downloaded for each microalga from the previously mentioned databases. Gene Ontology terms seeks the development of a human-readable and machine-readable hierarchical vocabulary to relate genes with their molecular functions, biological processes in which they are involved and the cellular components where they perform their function (Ashburner et al., 2000; Carbon et al., 2019). Complementary, KEGG Orthology terms associate genes to metabolic pathways and modules based on their orthologous relationships in sequenced genomes (Kanehisa et al., 2016; Moriya et al., 2007; Ogata et al., 1999). However, for microalgae species lacking these annotation systems, HMMER (biological sequence analysis using profile hidden Markov models) (Potter et al., 2018) is used to identify protein domains according to the PFAM (Protein Family) nomenclature (Mistry et al., 2021). PFAM terms are subsequently converted into GO terms using pfam2go (<http://current.geneontology.org/ontology/external2go/pfam2go>). KO terms are associated to genes applying KAAS (KEGG Automatic Annotation Server) (Moriya et al., 2007). Whenever possible, other systematic functional annotation format are also included: Protein Analysis Through Evolutionary Relationships (PANTHER) terms (Mi et al., 2021), that classifies, based of evolutionary families and subfamilies, gene products into classes capturing molecular function, biological process and metabolic pathways; Enzyme Commission numbers (EC numbers), that consists of a numerical classification identifier for enzymes, related to the biochemical reactions they perform; and Eukaryotic Orthologous Groups (KOG) terms, used to identify orthologue and parologue groups of proteins (Galperin et al., 2021).

## ***Development of functional annotation and genomic packages***

In order to use all these annotation systems in ALGAEFUN two different types of R annotation packages are developed. The systematic sources of functional annotation mentioned previously are gather together using the function makeOrgPackage from the Bioconductor R package AnnotationForge (Carlson & Pagès, 2019) which generate annotation packages. These packages are instrumental when performing functional enrichment analysis over gene sets obtained from RNA-seq data analysis. In addition, gene featuring annotation of each microalga from the previously downloaded and processed GTF files is stored in different packages applying the function makeTxDbFromGFF from the Bioconductor R package GenomicFeatures (Lawrence et al., 2013). These packages are central to carry out analysis over genomic loci obtained in ChIP-seq analysis. Both functional annotation and gene feature packages of each microalga are freely available on our Github repository, in order to enable the research community in microalgae to use them and perform omics analysis independently from the tools available from ALGAEFUN with MARACAS:

<https://github.com/fran-romero-campero/AlgaeFUN/tree/master/packages>

## ***MARACAS implementation: high-throughput sequencing data processing***

The computational core of this tool consists of a parallel fully automatic computational pipeline or workflow synchronized through blackboards. This workflow is managed by the job scheduling system SLURM (Simple Linux Utility for Resource Management) and bash scripting. The inputs to our pipeline comprise the pre-computed index of the corresponding microalga reference genome, the previously processed gene feature annotation file in GTF format (both already included in MARACAS) and the raw sequencing data in FASTQ format from RNA-seq and ChIP-seq microalgae studies provided by the user. In turn, this workflow produces, as outputs, two lists of differentially expressed genes or DEGs (activated and repressed genes) when the FASTQ files correspond to an RNA-seq study; or a list of genomic loci or regions significantly occupied by the transcription factor or histone modification of interest in the case of a ChIP-seq study.

MARACAS requires the user to input specifications such as the microalga of interest, the names for control and experimental conditions, number of replicates and location of raw high-throughput sequencing files.

**Table 4: Description of parameters file used as input in MARACAS for RNA-seq analyses.**

<b>Parameters file for RNA-seq analyses</b>	
<b>data_source</b>	This parameter indicates the source of the data to be analyzed. It can take the value FILES when the FASTQ files are already located in a folder in the computer where MARACAS is installed or the value DB when the data has been already deposited in the GEO data base.
<b>cluster</b>	Parameter specifying the execution mode. SERVER mode executes MARACAS with a sequential analysis of the different samples. Whereas SLURM mode executes MARACAS in a parallel manner, processing samples simultaneously in different computational nodes. In this last case, SLURM needs to be installed in user's computer cluster.
<b>number_processors</b>	Number of processors that can be used by MARACAS.
<b>working_directory</b>	It can take the values FALSE, when your data is single end, and TRUE, when your data is paired end.
<b>microalgae</b>	It indicates the location where the analysis folder will be generated.
<b>read_mapper</b>	Name of the microalgae of interest.
<b>main_folder</b>	This parameter specifies the software tool to perform read mapping. Two different options are provided: HISAT2 and Kallisto.
<b>number_of_samples</b>	Name of the folder that will be created at the working directory to contain the outputs from the analysis.
<b>control_condition_name</b>	Total number of samples to be analyzed.
<b>experimental_condition_name</b>	Name of control condition.
<b>loc_sampleN</b>	Name of experimental condition.
<b>acc_sampleN</b>	When paired-end: FALSE and data-source: FILES, this parameter indicates the path and file name of sampleN. N will take values from 1 to number-of-samples.
<b>loc_sample_leftN</b> <b>loc_sample_rightN</b>	When data-source: DB, this parameter specifies accession number of FASTQ files in GEO.
<b>condition_sampleN</b>	When paired-end: TRUE and data-source: FILES, this parameters indicates the path and file name of the FASTQ samples containing the left and right reads, respectively.
<b>fold_change</b> <b>q_value</b>	This parameter specifies which condition name of the ones chosen in control-condition-name or experimental-condition-name correspond to each sample.
	These parameters specify the fold-change and q-value used to determine differential expressed genes in the experimental condition when compared to the control condition.

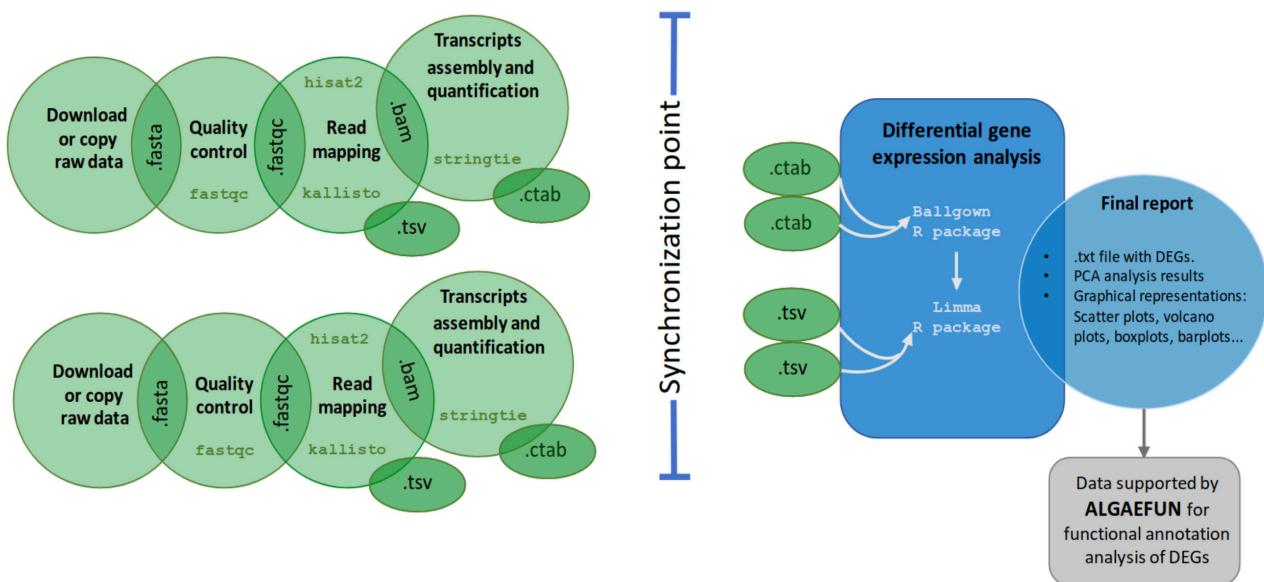
**Table 5: Description of parameters file used as input in MARACAS for ChIP-seq analyses.**

<b>Parameters file for ChIP-seq analyses</b>	
<b>data_source</b>	FILES or DB as it is explained in Table 4.
<b>cluster</b>	SERVER or SLURM as it is explained in Table 4.
<b>number_processors</b>	Number of processors that can be used by MARACAS.
<b>paired_end</b>	FALSE or TRUE as it is explained in Table 4.
<b>working_directory</b>	It indicates the location where the analysis folder will be generated.
<b>microalgae</b>	Name of the microalgae of interest.
<b>main_folder</b>	Name of the folder that will be created at the working directory to contain the outputs from the analysis.
<b>number_of_samples</b>	Total number of samples to be analyzed.
<b>included_control</b>	It can receive yes if your experimental design includes a control condition such as input, mock or similar. Use no in negative cases.
<b>mode</b>	Use the values transcription-factor or histone-modification to specify if your ChIP-seq data was generated for a transcription factor or a histone modification study.
<b>transcription_factor</b>	When mode: transcription-factor this parameter specifies the name of the chosen transcription factor.
<b>histone_modification</b>	When mode: histone-modification this parameter specifies the name of the chosen histone modification.
<b>loc_chip_replicate_N</b>	When paired-end: FALSE and data-source: FILES, this parameter indicates the path and file name of ChIP sample and, in case of included-control: yes, the control sample. N will take values from 1 to number-of-replicates.
<b>loc_control_replicate_N</b>	When data-source: DB, this parameter specifies accession number of ChIP sample in GEO. In case of included-control: yes, control-replicate-N will specify the one for control sample in GEO.
<b>loc_chip_replicate_left_N</b> <b>loc_chip_replicate_right_N</b> <b>loc_control_replicate_left_N</b> <b>loc_control_replicate_right_N</b>	When paired-end: TRUE and data-source: FILES, this parameter indicates the path and file name of the ChIP samples containing the left and right reads, respectively. If included-control: yes, they also indicate this information for control samples.

In case of analyzing already published omic data, its accession number can be specified instead. Additionally, the user can set the statistical parameters to perform the corresponding analysis. Specifically, the fold-change and significance level cutoff thresholds for the identification of differentially expressed genes can be selected. In addition, user can choose as read mapper software either the short read mapper HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2) or the pseudoalignment method implemented in kallisto. Whereas HISAT2 is an exact method requiring several hours for processing a typical sample (Kim et al., 2015; Pertea et al., 2016), kallisto produces near-optimal gene expression quantification in only a few minutes (Bray et al., 2016). Meanwhile, in the MARACAS ChIP-seq data analysis pipeline is used the ultra-fast and memory-efficient short read mapper bowtie2 (Langmead & Salzberg, 2012). All this information is collected into a parameters file (Table 4 and Table 5), which is the main input received by the pipeline. The first step in this workflow loads the parameters file and generates the working directory, including one folder where the samples are downloaded or copied, another one where the results are generated and a third one, called “logs”. In this last folder, several text files are generated containing messages written by the different processes in the pipeline that are useful to keep track of the important events. More interestingly, a different type of text file, called blackboard, is created and the computational processes in this workflow are given read and write permissions for it. Files functioning as blackboards are used for indirect communication between the computational processes in the parallel pipeline in order to synchronize them. In the next step, the pipeline forks into two different modes to process either RNA-seq or ChIP-seq data.

The fork dedicated to RNA-seq analysis is represented in Fig. 14. Depending on the raw data source, the first step consists of either downloading and extracting the corresponding FASTQ files from a database or copying them to the specified working directory. Next, a quality control of these raw data is performed and short reads are mapped to the reference genome using HISAT2 or kallisto. When HISAT2 is used, this step generates BAM (Binary Alignment Maps) files that contain the mapped reads. The following steps take this mapping as input and perform transcripts assembly and expression quantification using STRINGTIE (Pertea et al., 2015, 2016). In this step, GTF files with the assembled transcripts and CTAB files (chemical table files) with the expression quantification values are generated for all samples. However, when kallisto is used, transcripts are mapped and quantified in the same step, so a TSV (tabular separated values) file containing the result of the quantification is generated (Bray et al., 2016). These processing steps are carried out in parallel, simultaneously for each

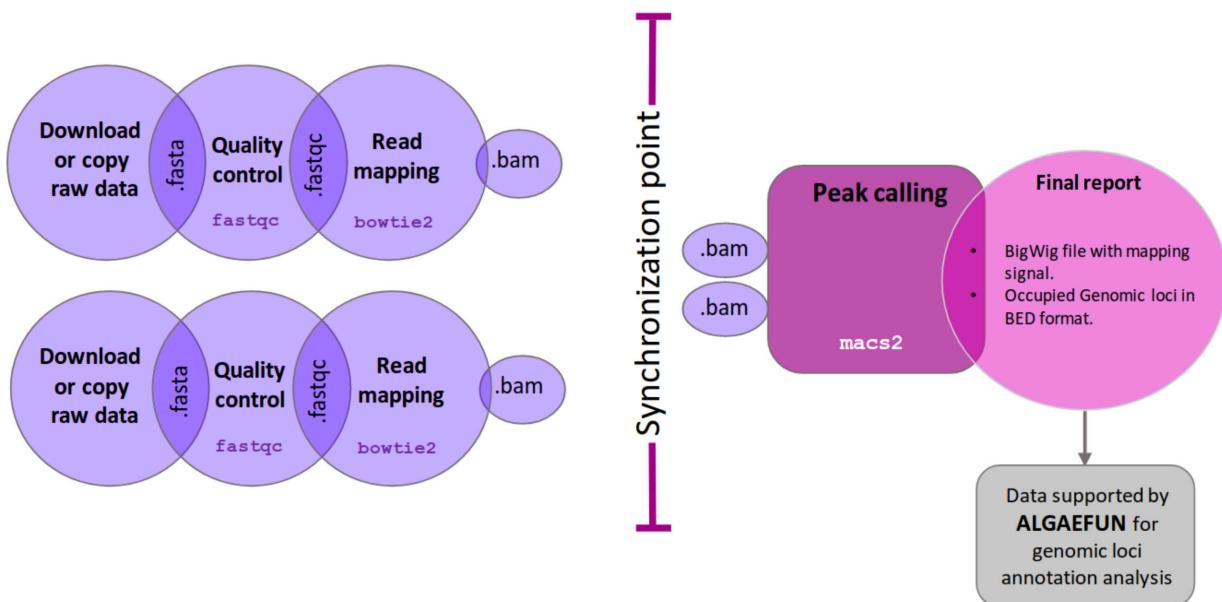
sample, being synchronized using a blackboard file. These parallel processes write on the blackboard when a goal is reached (for example, BAM files are generated) in order to keep track of their progress. When all parallel processes reach their goals, the following sequential steps of the pipeline are executed. A differential expression analysis is carried out using the Bioconductor R packages Ballgown ([Frazee et al., 2015](#); [Pertea et al., 2016](#)) and Limma ([Ritchie et al., 2015](#)). Specifically, DEGs are selected using the statistics based on a moderated t-student. Finally, reports in HTML and PDF format are generated with details on sequence quality analysis, mapping process and normalization. Graphics for exploratory analysis such as principal components, box plots, scatter plots, volcano plots and bar plots of individual genes are also included. These reports include links to download gene expression estimates as well as lists of activated and repressed DEGs. The outputs of this pipeline are compatible with the input formats for ALGAEFUN in order to facilitate further functional enrichment analysis and visualization.



**Figure 14: Workflow of the automatic pipeline for the analysis of RNA-seq data in MARACAS.** Green is used for steps occurring in parallel and blue for sequential steps. Each bubble represents a step of the workflow. The intersection between two bubbles highlight the file format resulting as output from the first process and serving as input of the following one. Outputs expected in the synchronization point before moving to the differential gene expression analysis are represented in oval bubbles.

The second fork of our pipeline dedicated to ChIP-seq analysis (Fig. 15) shares with the above RNA-seq fork the first steps consisting of downloading or copying the FASTQ raw files for the ChIP samples (data corresponding to the chromatin immunoprecipitation condition)

and control samples into the generated working directory and performing a quality control. Specifically, in this part of the workflow, short read mapping to the reference genome is executed using bowtie2 (Langmead & Salzberg, 2012) and the synchronization point requires that BAM files for every ChIP and control sample is generated. This synchronization point is also achieved using a blackboard. After synchronization, the last step takes as input all mappings in BAM format and performs peak calling using the software tool macs2 (Gaspar, 2018). This final step consists of the identification of genomic loci or regions significantly enriched with mapped reads, indicating the genomic occupation of the transcription factor or histone modification of interest. Reports in HTML and PDF format are also generated with details on sequence quality analysis and mapping process. These reports include links to download the identified peaks in BED (Browser Extensible Data) format and the genome wide mapping signal in BigWIG format. These outputs are compatible with the input formats for ALGAEFUN as well.



**Figure 15: Workflow of the automatic pipeline for the analysis of ChIP-seq data in MARACAS.** Violet is used for steps occurring in parallel and pink for sequential steps. Each bubble represents a step of the workflow. The intersection between two bubbles highlight the file format resulting as output from the first process and serving as input of the following one. Outputs expected in the synchronization point before moving to the peak calling are represented in oval bubbles.

## ALGAEFUN implementation: functional annotation analysis

The generated output from MARACAS, either sets of genes or genomic loci, can be functionally annotated using the next software tool in the portal, ALGAEFUN. Although this tool can also be used to functionally annotate genes sets or genomic loci generated independently from other tools. The user interface of ALGAEFUN is shown in Fig. 16 and Fig. 17. Two different modes of operation, depending on the kind of input data to be analyzed, are implemented in ALGAEFUN.

**(A)** **ALGAEFUN with MARACAS**  
microALGAE FUNCTIONal enrichment tool for MicroAlgae  
RnA-seq and Chip-seq Analysis

**(B)** Choose your favourite microalgae  
Ostreococcus tauri

**(C)** Choose your desirable analysis  
 GO terms enrichment  
 KEGG pathways enrichment analysis  
 Both

Choose gene ontology:  
 Biological process  
 Cellular Component  
 Molecular Function

Which will be your chosen p-value?  
0,05

**(D)** Insert a set of genes  
Insert set of genes  
Example Clear  
Choose File with Gene Set to Upload  
Browse... No file selected  
Would you rather use your own background set?  
 Yes  
 No  
Have fun!

**Figure 16: Microalgae Functional Annotation tool (ALGAEFUN) user interface for gene sets functional annotation.** (A) Navigation bar for selecting the tool to use between the ones included in our web app; (B) Drop-down menu to select the microalga of interest; (C) Sidebar panel to select parameters for GO and/or KEGG pathway enrichment; (D) Main panel to input set of genes to analyze and select gene background.

In one of the modes, input data consists of sets of genes obtained, for instance, from an RNA-seq study. Here after selecting this analysis mode (Fig. 16-A), the user chooses the microalga of interest (Fig. 16-B) and the analysis to carry out, a GO term and/or KEGG

pathway enrichment analysis at the selected significance level ([Fig. 16-C](#)). The set of genes to study can be inputted through a text box or uploaded from a file. Users can also choose between using their own background gene set or the entire microalga genome ([Fig. 16-D](#)). In order to allow users to explore the functionalities of ALGAEFUN and to check the required gene id format, an example gene set for each microalga is included. This example can be accessed and inputted in the corresponding text box by clicking on the example button ([Fig. 16-D](#)). These examples have been generated during the testing of MARACAS using previously published RNA-seq data sets and have in turn been used in the testing and validation of ALGAEFUN. The GO term and KEGG pathway enrichment analysis are carried out using the Bioconductor R package ClusterProfiler ([Wu et al., 2021](#); [Yu et al., 2012](#)). This package implements statistical analysis and visualization of functional profiles for gene clusters or sets using the annotation packages developed for each microalga, integrating the systematic sources of functional annotation previously discussed.

The outputs for the GO term and KEGG pathway enrichment analysis are presented in two different tabs. The first output in the GO enrichment tab consists of a table that summarizes the results from the GO enrichment analysis carried out over the input gene set. The user can find one row for each GO term and six columns that represent some relevant information about the enrichment. The first column shows the GO term identifier, followed by the second column where the user can find its description. Users can access more information about the GO term represented in a specific row by clicking on its identifier to be redirected to the web portal AmiGO ([Carbon et al., 2009](#)) where GO terms are described in detail. The third and fourth column represents the p-value and q-value, and the fifth one shows the enrichment value:  $E = (m/n) / (M/N)$ . The value “m” is the number of genes from the inputted gene set annotated with the corresponding GO term whereas “M” is the number of genes from the background annotated with the mentioned GO term. In a similar way, “n” is the number of genes with annotation from the gene set whereas “N” is the number of genes with annotation from the gene background. Finally, the last column shows the genes from the input set of genes annotated with each GO term. The user can click on them to get more information from the gene entry on the corresponding database from which the annotation was retrieved for the specific microalga under study. Furthermore, ALGAEFUN also generates several graphs that represent the GO term enrichment. Five visualization methods are used to illustrate the results:

- Acyclic graph: Each node stands for a GO term and their color indicates the level of significance (from grey, non-significant, to intense red, highly significant). An arrow is drawn from GO term A to GO term B to represent that A is a more general GO term than B.
- Bar-plot: Each bar represents an enriched GO term whose length corresponds to the number of genes in the gene set annotated with the given GO term. Once again, the bar color shows the level of significance (from blue, less significant, to red, more significant).
- Dot-plot: Each dot represents an enriched GO term. The x-position of the dot corresponds to the ratio between the number of genes annotated with the corresponding GO term and the total number of annotated genes in the gene set. The dot color captures the level of significance (from blue, less significant, to red, more significant).
- Enrichment Map (emap-plot): Each node represents an enriched GO term and the size of each node is proportional to the number of genes annotated with the corresponding GO term in the inputted gene set. The node colors represent the level of significance (from less significant in blue to more significant in red). These nodes are connected by edges when the corresponding GO terms are semantically related.
- Gene-concept network (cnet-plot): The beige nodes represent GO terms and the grey nodes genes. An edge is drawn from a gene to a GO term when the gene is annotated with the corresponding GO term. The size of nodes representing GO terms is proportional to the number of genes annotated with the corresponding GO term.

The outputs shown on the KEGG pathway enrichment tab consist of:

- Table summarizing the result of the KEGG pathway enrichment analysis: Each row represents a pathway significantly enriched in the inputted gene set with respect to the selected gene background. The first column represents the KEGG pathway identifier and the user can click on it to read more information about the pathway. The second column contains its description. The third and fourth column present the p-value and q-value, and the fifth column displays the corresponding enrichment value E (m/n; M/N) as previously described. Finally, the last column shows the list of genes from the inputted gene set assigned to the corresponding enriched pathway. KEGG pathways can be more informative than GO term since they are not general but specific to the corresponding organism of interest.

- KEGG pathway map: Users can choose a specific enriched pathway using a drop-down menu to generate the corresponding KEGG pathway map where genes from the inputted gene set are highlighted in red.
- Table summarizing the result of the KEGG module enrichment analysis: Each row represents a module significantly enriched in the gene set with respect to the selected gene universe or background. The columns in this table are organized in the same manner as the previously described. KEGG modules are distinct recurrent components of KEGG pathways in this respect they are more specific and can be more informative.

In the other mode, input data consists of genomic loci or regions obtained, for instance, from a ChIP-seq study. This analysis mode is selected from the side bar panel in Fig. 17-A. The microalga of interest can be selected using the drop-down menu from Fig. 17-B. Next, the distance around the transcriptional start site (TSS) that will be considered defining gene promoters must be specified. Gene features that will be considered when assigning gene targets to genomic loci or regions also need to be selected (Fig. 17-C). Genomic loci or regions to analyze can be inputted through a text box or uploading a file. Additionally, a BW file containing the number of mapped reads or signal in each position of the genome can be uploaded (Fig. 17-D). Similar to the previous mode, it is possible to explore the functionalities of this tool and check the required genomic loci or regions format, using an example included for each microalga by clicking on the example button (Fig. 17-D). These examples have been generated during the testing of MARACAS, using previously published ChIP-seq data sets and analysis, and have in turn been used in the testing and validation of ALGAEFUN. The functional annotation of the inputted genomic loci or regions is performed using the Bioconductor R packages ChIPseeker (Yu et al., 2015) and ChIPpeakAnno (Zhu, 2013). These package implements statistical analysis and visualization of genomic loci and regions using the gene feature annotation packages generated for our tool. The outputs generated in this type of analysis consist of:

- A pie chart representing the distribution of the genomic loci or regions over the different type of gene features selected by the user such as the promoter, 3' UTR, 5'UTR, intron or exon.
- A table enumerating the target genes associated one generated when the data comes from an RNA-seq study. It represents each gene located in the enriched genomic loci

and its different annotation terms. This set of genes can be downloaded and subsequently annotated functionally using ALGAEFUN.

- A visualization of the average level of signal around TSS and transcriptional end site (TES). For each individual gene, it generates a visualization of the signal and identification of DNA motifs recognized by transcription factors and regulators in microalgae.

**(A)**  **ALGAEFUN with MARACAS**

microALGAE FUNctional enrichment tool for MicroAlgae  
RnA-seq and Chip-seq Analysis

**IBV**   
Instituto Biológico Valenciano  
y su Fundación

**UNIVERSIDAD DE SEVILLA** 

**CSIC** 

**(B)** Choose your favourite microalgae  
Ostreococcus tauri

Choose your desirable analysis  
 GO terms enrichment  
 KEGG pathways enrichment analysis  
 Both

Choose gene ontology:  
 Biological process  
 Cellular Component  
 Molecular Function

Which will be your chosen p-value?  
0.05

**(C)** Insert a set of genes  
Insert set of genes  
Example Clear  
Choose File with Gene Set to Upload  
Browse... No file selected

Would you rather use your own background set?  
 Yes  
 No

**(D)** Have fun!

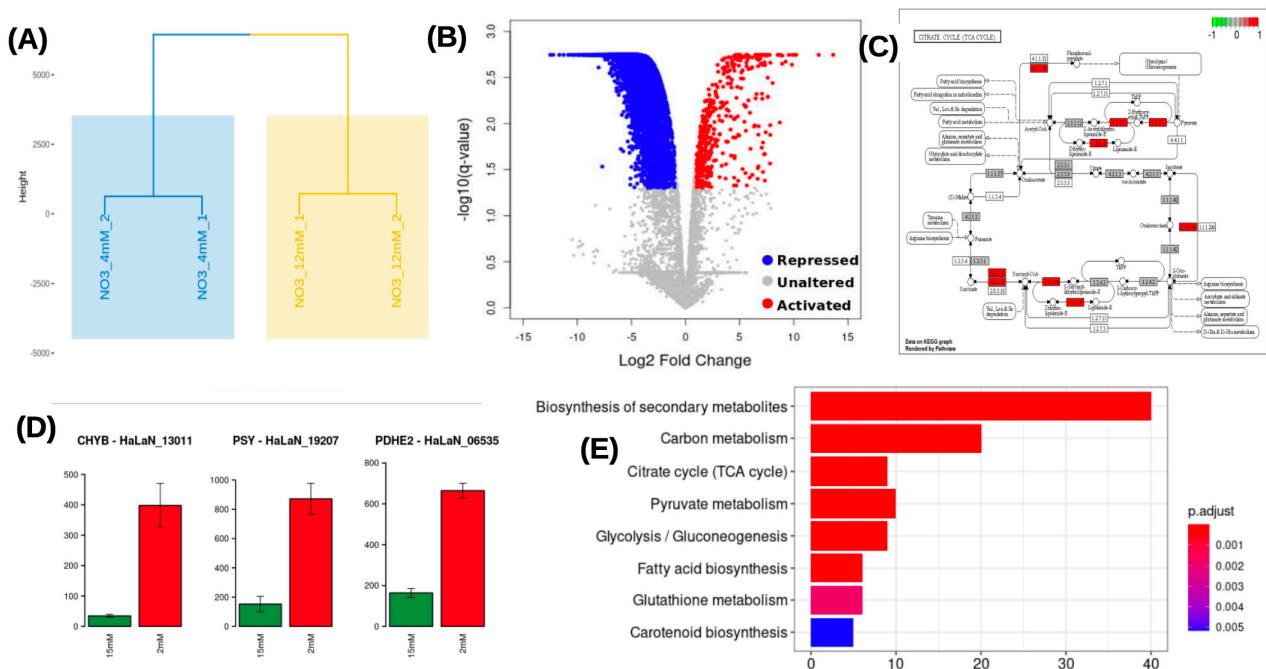
**Figure 17: Microalgae Functional Annotation tool (ALGAEFUN) user interface for genomic loci annotation.** (A) Navigation bar for selecting the tool to use; (B) Drop-down menu to select the microalga of interest; (C) Sidebar panel to select parameters for the identification of the promoter regions and the gene features or parts that will be considered to assign genes to genomic loci; (D) Main panel to input set of genomic loci and signal in BigWIG format obtained from a ChIP-seq analysis.

## **Case study 1: from RNA-seq raw sequencing data to biological processes and pathways**

This case study is based on RNA-seq data generated in a project different to the one presented in this thesis. It consists of an RNA-seq study carried out using *Haematococcus*, a microalga of industrial interest for the bio-production of astaxanthin (Hoys et al., 2021), where the tool ALGAEFUN with MARACAS was instrumental for publication. The analysis has been performed for vegetative *Haematococcus pluvialis* cells, grown both under N sufficiency and under moderate N limitation in order to unveil the transcriptomic programme enhancing astaxanthin biosynthesis under N deprivation. This illustrates the type of information that ALGAEFUN with MARACAS is able to reveal from raw sequencing data.

The high-throughput sequencing raw data in FASTQ format is processed using MARACAS in order to obtain the set of differentially expressed genes using a criterion based on a fold-change value of 2 and a q-value of 0.05. The report produced by MARACAS confirms that all samples are of high quality and notifies no problem during read mapping to the reference genome with mapping rates greater than 86 %. Scatter plots comparing gene expression between samples are also produced in the MARACAS report. High Pearson correlations greater than 98 % are identified between replicates of the same condition. Accordingly, automatically generated hierarchical clustering combined with Principal Components Analysis (PCA) lead to the identification of well defined clusters, due to the noise reduction that has been achieved with PCA (Fig. 18-A). It identifies two clearly separated clusters constituted by the control (N sufficiency) and experimental condition samples (moderate N deprivation). Volcano plots comparing the transcriptomes under moderate N limitation and N sufficiency are used in the report to represent the repressed genes and activated genes. Moderate N limitation result in a strong repressing effect over the transcriptome with respect to N sufficient condition, identifying 414 activated and 5348 repressed genes (Fig. 18-B). These lists of genes can be then inputted into ALGAEFUN to determine significantly over-represented biological processes or pathways affected in the studied experimental conditions. As described previously, ALGAEFUN can perform two different types of functional enrichment analysis when a set of genes from an RNA-seq analysis is used as input: GO terms and KEGG pathways enrichment. The GO terms analysis can identify, for example, specific biological processes enriched in the set of genes. Barplots are one of the available graphical representations of the GO enrichment results in ALGAEFUN (Figure 18-E), presenting the biological processes

enriched in the activated set of genes obtained from the *Haematococcus* RNA-seq study under moderate N limitation. Reduced availability of nitrogen activates processes like biosynthesis of secondary metabolites, TCA cycle or pyruvate metabolism at a transcriptomic level. GO term analysis provide a general overview of the functional annotation of gene sets since they constitute universal functional terms not specific to a particular organism. Complementary, KEGG pathway enrichment is specific to the corresponding organism and can be more informative in some cases. It allows to identify the transcriptomic activated enzymes in enriched pathways (Fig. 18-C). The genes involved in these processes can be further studied individually and their expression level in both conditions can be compared (Fig. 18-D).



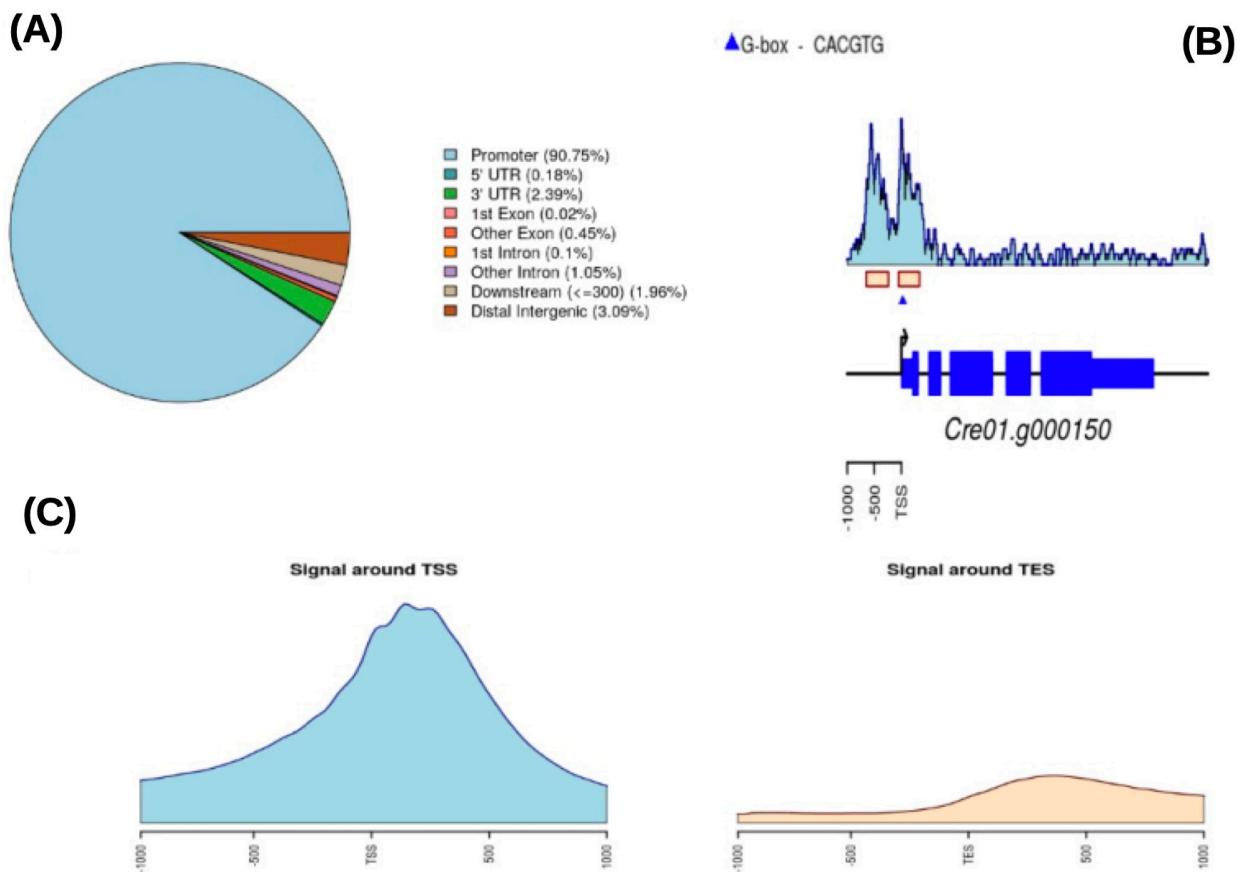
**Figure 18: Results obtained from MARACAS final report and from ALGAEFUN analysis.** (A) Hierarchical clustering combined with PCA generated by MARACAS; (B) Volcano-plot generated by MARACAS showing the global effect over the transcriptome as well as the activated and repressed genes detected (in red and blue color, respectively); (C) KEGG representation of one of the enriched pathways generated by ALGAEFUN, genes present in the list used as input are colored in red color; (D) Barplots representing individual gene expression levels of key enzymes, figure generated by MARACAS. Expression level of the selected gene under N sufficiency is represented in green color. Red color is used to represent moderate N deprivation condition; (E) Horizontal barplot representing the enriched biological process in the set of genes used as input, figure generated by ALGAEFUN. Gradient from red to blue color is used to represent p-value.

ALGAEFUN with MARACAS analysis describes differential expression of hundreds of genes affecting key pathways converging into astaxanthin biosynthesis and storage. The affected pathways are further studied at metabolic level and a massive cell reprogramming is verified for cells growing under moderate N limitation. Furthermore, a major advance is made by discerning the underlying control mechanisms by finding differentially activated enzymes in the astaxanthin biosynthesis pathway. Moreover, the identification of significantly present DNA sequences in the promoters of these key enzymes is carried out and, for the first time, the family of transcription factors bHLH, is proposed as candidates for transcriptional regulators of these key enzymes in astaxanthin biosynthesis (Hoys et al., 2021).

## Case study 2: From ChIP-seq raw sequencing data to marked genes

Histone modifications play a central role in gene expression control. The genome-wide distribution of the mark H3K4me3 associated to gene activation has been determined in *Chlamydomonas* (Ngan et al., 2015). ChIP-seq raw data from that publication is re-analyzed using MARACAS, as previously described, and 12814 genomic loci are identified as significantly occupied by H3K4me3 in the *Chlamydomonas* genome under standard growth conditions. The target genomic loci and the corresponding genome wide mapping signal file in BigWIG format are uploaded in ALGAEFUN. The region of two kilobases around the Transcriptional Start Site (TSS) is considered as gene promoter and all the gene features are selected to determine the H3K4me3 marked genes. The outputs are presented in the graphical interface in different tabs. A downloadable table with the marked genes and their available annotation is generated. This gene list can, in turn, be analysed by ALGAEFUN to perform a GO term and/or pathways enrichment analysis. A total of 11558 H3K4me3 marked genes are identified. Graphs representing the distribution of the genomic loci overlapping different gene features (Fig. 19-A) and the distribution of upstream and downstream signal around TSS and Transcriptional End Site (TES) are also represented. In agreement with the previously published results, the 90.75 % of the genomic loci occupied by H3K4me3 are located at gene promoters in *Chlamydomonas*. As in this case study, when a BigWIG file with the genome wide mapping signal is provided, specific marked genes can be selected to visualize the signal profile over their gene bodies and promoters. A gene example presenting two H3K4me3 peaks on its promoter is depicted to illustrate this functionality (Fig. 19-B). Moreover, DNA motifs recognized by specific transcription factors and regulators in photosynthetic organisms can be identified in the promoter of the selected gene. Finally, a visualization of the average level of signal around TSS and TES across all marked genes is generated (Fig. 19-C). For the case of

H3K4me3 in *Chlamydomonas* further evidence is obtained, showing that this epigenetic mark specifically and exclusively locates at the TSS of marked genes and not at the TES.



**Figure 19: Summary of the outputs generated by ALGAEFUN when a genomic loci list is used as input.** (A) Pie chart representing the distribution of peaks or genomic loci over the different type of gene features such as the promoter, 3' UTR, 5' UTR, intron or exon; (B) Visualization of the signal and identification of DNA motifs recognized by transcription factors and regulators in photosynthetic organisms for an individual gene; (C) Visualization of the average level of signal around TSS and TES over the target genes.

## Contribution of ALGAEFUN with MARACAS to the field

ALGAEFUN with MARACAS constitutes one of the first steps that has been taken for the development of tools that would enable the microalgae research community to exploit high throughput next generation sequencing data by applying systems biology techniques. The first difference between ALGAEFUN with MARACAS with respect to already existing tools consists in the wide range of supported microalgae species (Fig. 13). For the model microalgae *Chlamydomonas reinhardtii*, researchers can find several online tools to functionally annotate set of genes, such as Algal Functional Annotation Tool (López et al., 2011) and

ChlamyNET (Romero-Campero et al., 2016). Only the online tool AgriGO (Tian et al., 2017) offers the possibility of analysing a restrictive number of different microalgae species beyond *Chlamydomonas*. The second biggest difference between ALGAEFUN and other tools is the annotation system they use. Most available functional enrichment tools can only perform functional annotation of gene sets based exclusively on GO enrichment analysis. The identification of significantly enriched KEGG pathways in the inputted sets of genes is only supported by ALGAEFUN and Algal Functional Annotation Tool. A fundamental difference between ALGAEFUN and other tools consists of the statistical tests. Whereas AgriGO and ChlamyNET are based on Fisher's exact test, ALGAEFUN and Algal Functional Annotation Tool compute statistical significance according to Hypergeometric tests. It has been shown that, in general, the hypergeometric test has more statistical power than Fisher's exact and  $\chi^2$  (Masseroli et al., 2004).

**Table 6: Comparison between ALGAEFUN with MARACAS and other functional enrichment analysis tools.**

Algal functional annotation tool	AgriGo	ChlamyNET	ALGAEFUN with MARACAS
Gene sets as input	YES	YES	YES
Genomic loci as input	NO	NO	YES
GO enrichment	YES	YES	YES
KEGG pathways enrichment	YES	NO	YES
Several microalgae	NO	YES	NO
Statistical test	Hypergeometric tests	Fisher's exact test	Fisher's exact test
			Hypergeometric tests

Moreover, none of these tools can be used as a complete and integrated tool to process high-throughput sequencing raw data from RNA-seq or ChIP-seq experiments, or functionally annotate genomic loci obtained from a ChIP-seq analysis. In this respect, ALGAEFUN with MARACAS improves and implements several novel functionalities of similar already existing software tools (Table 6).

ALGAEFUN with MARACAS is a constantly growing tool that will include new packages and new microalgae whenever their sequenced genomes are available. Also, it has settled the bases to build numerous tools by other members of our laboratory aiming at generating an enabling technology for the microalgae research community doing systems biology studies.

# **Chapter 2.**

## **Transcriptomic analysis of diel and seasonal cycles in**

### ***Ostreococcus tauri***

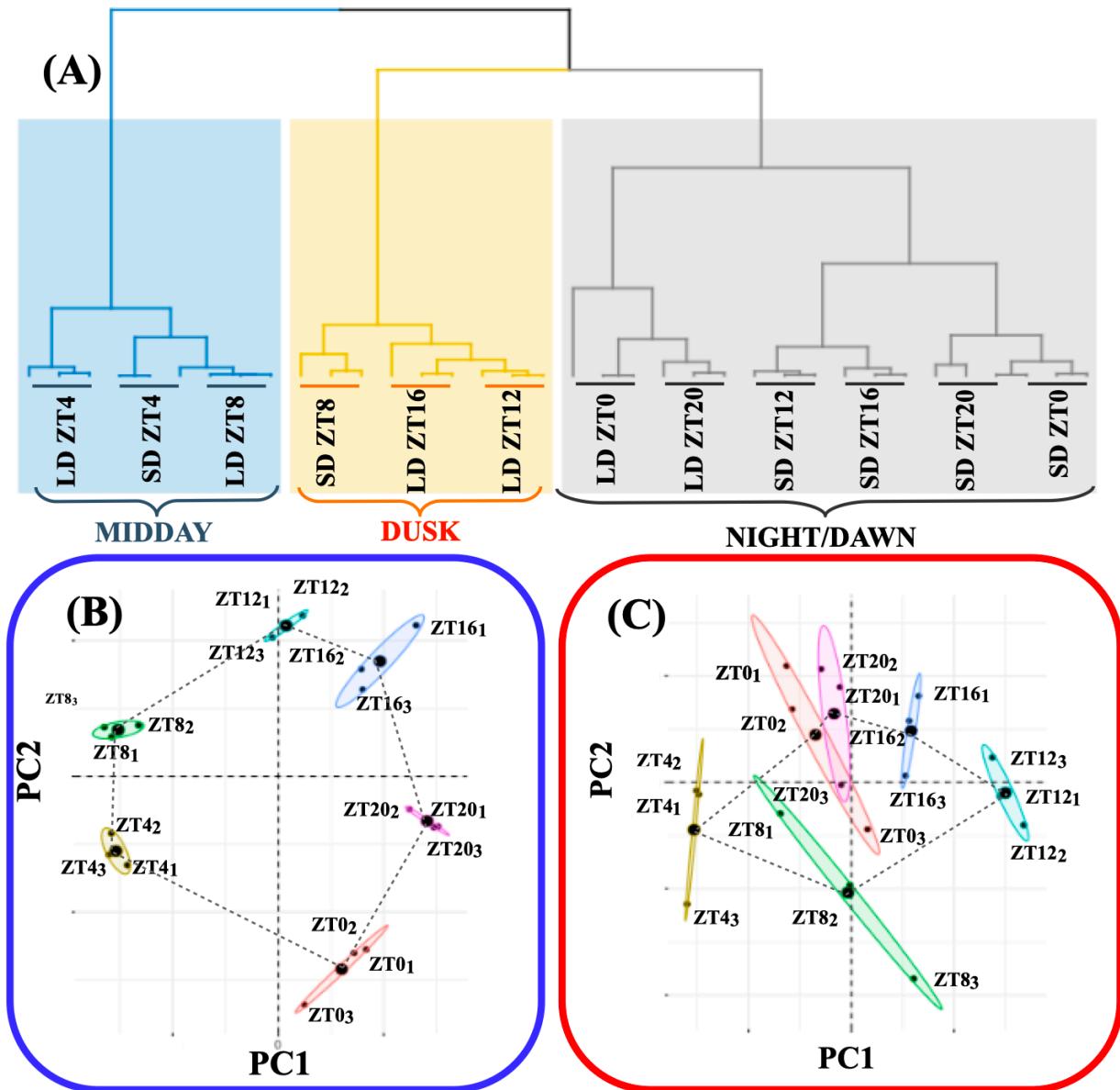


## Chapter 2. Transcriptomic analysis of diel and seasonal cycles in *Ostreococcus tauri*

This chapter aims to characterize *Ostreococcus* transcriptome rhythmicity under different photoperiodic variations in diel cycles as well as free-running conditions consisting in constant light and dark. Accordingly, culture samples are collected from summer long day condition (LD, 16 h light : 8 h dark) and winter short day condition (SD, 8 h light : 16 h dark) entrained cultures for three consecutive days every 4 h. Subsequently, to determine *bona fide* circadian genes, cultures are transferred to free-running conditions consisting of constant light (LL) and dark (DD) and samples are collected for two consecutive days every 4 h (Fig. 11).

High-throughput transcriptome sequencing produced approximately 10 million short reads per sample (Appendix 1), allowing to accurately estimate gene expression levels measured as FPKM (Fragments Per Kilobase of exon per Million reads mapped). Indeed, out of the 7668 genes currently annotated in the *Ostreococcus tauri* genome (Blanc-Mathieu et al., 2014; Palenik et al., 2007), only 3 genes are never expressed and 260 genes never exceed an expression level of 10 FPKM. This shows that practically the entire *Ostreococcus tauri* genome is expressed under seasonal and diel cycles. A hierarchical clustering analysis performed over the 36 transcriptomes corresponding to the time points taken during three days under light/dark cycles shows a clear structure. The transcriptomes corresponding to the same time points during the three different days cluster together (Fig. 20-A). A cyclic circular organization of the transcriptomes is revealed over diel cycles constituted by three distinct clusters corresponding to midday (LD ZT4, SD ZT4 and LD ZT8), dusk (SD ZT8, LD ZT16 and LD ZT12) and night/dawn constituted by LD ZT0 and LD ZT20 on one hand and SD ZT12, SD ZT16, SD ZT20 and ZT0 on the other hand (Fig. 20-A). The transcriptomes at time points in the LD and SD nights 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 and night/dawn transcriptomes when compared to the midday one (Fig. 20-A). In order to obtain a deeper understanding of the underlying structure in these data, a principal components analysis is performed over the LD (Fig. 20-B) and SD (Fig. 20-C) transcriptomes. Under LD condition, the transcriptomes corresponding to the same time point in the three different days tightly cluster together, constituting a circular structure. Nonetheless, under SD condition, more variability is observed and transcriptomes form a structure resembling an ellipse. This

could indicate that whereas in LD condition gene expression in globally cycling precisely with a similar period, a more complex behavior is expected under SD condition.



**Figure 20: RNA reliability analysis.** (A) Hierarchical clustering of the RNA-seq data corresponding to the 36 time points collected under alternating dark/light cycles simulating long and short day conditions. Three distinct clusters are observed corresponding to midday (blue rectangle), dusk (yellow rectangle) and night/dawn (grey rectangle); (B) Principal Component Analysis of the time point global transcriptomes under long day condition. Small dots correspond to the 2D projection of each time point global transcriptome. Big dots correspond to the average of the three replicates 2D projections for each time point. Ellipses mark the 95 % confidence regions corresponding to each time point global transcriptome; (C) Principal Component Analysis of the time point global transcriptomes under short day condition. Points and ellipses are used as described before.

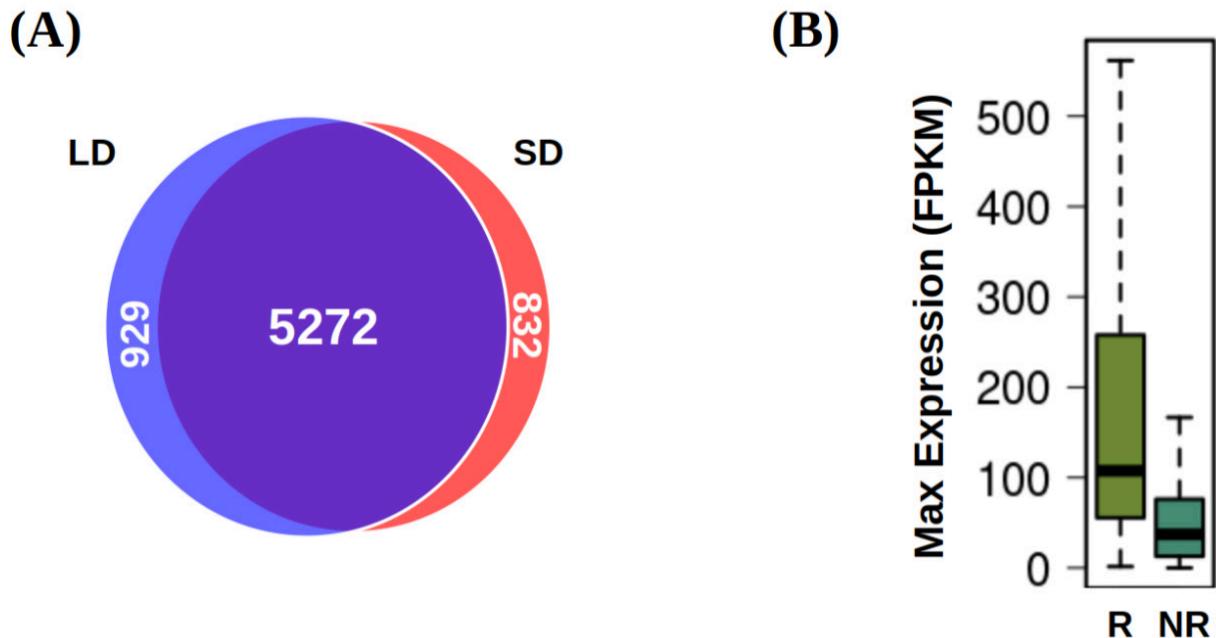
In addition, it is remarkable the high similarity between the transcriptomes corresponding to ZT0 and ZT20 under SD condition that is not that obvious under LD condition. This suggests that the transcriptomic response at the end of a SD night is already preparing all molecular systems for the incoming light availability at dawn, whereas this anticipation is not so evidently observed under LD condition. These results support that the experimental design grants a high level of synchronization in the data, allowing to proceed to the identification and comparison of genes exhibiting rhythmic expression patterns under LD and SD conditions, which can be also explored individually accessing the MINOTAUR tool.

## Transcriptomic characterization of diel expression profiles

### ***Global rhythmicity analysis of long day and short day transcriptomes***

Applying bioconductor R package RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods) (Thaben & Westermark, 2014), genes exhibiting diel (rhythmic profile with a 24 h period) expression patterns under both seasonal conditions have been identified. Specifically, more than 6000 genes comprising approximately 80 % of the entire *Ostreococcus* genome present diel expression patterns. This result agrees with previous studies in *Ostreococcus tauri* (Monnier et al., 2010) and *Chlamydomonas reinhardtii* (Zones et al., 2015) under neutral day conditions (12 h of light and 12 h or dark). The transcriptome rhythmicity in chlorophyte phytoplankton seems to be much higher than the ones found in other organisms such as *Arabidopsis thaliana*, with 30-50 % of its genes presenting diel expression patterns (Bläsing et al., 2005), *Solanum tuberosum* 18-45 % (Hoopes et al., 2022) *Drosophila melanogaster* 24 % (Ma et al., 2021) or *Mus musculus* 3-10 % (Miller et al., 2007).

Seasonal variations in photoperiod do not affect transcriptome rhythmicity, as the sets of rhythmic genes under LD and SD entrainment are almost coincident (Fig. 21-A). The 20 % of the genome of *Ostreococcus* that is apparently identified as non-rhythmic, present three time lower expression levels than rhythmic genes. This difference is significant according to a p-value of  $1.45 \times 10^{-4}$ , computed using Mann-Whitney-Wilcoxon test (Fig. 21-B). Since the rhythmicity analysis method used requires high levels of expression to perform optimally (Laloum & Robinson-Rechavi, 2020), it is possible part of these 20 % of the genome is also rhythmic. In addition, some of these non-rhythmic genes are associated with stress responses (list of genes available at MINOTAUR), so they could potentially become rhythmic once the stress signals are present and their expression level is increased, possibly resulting in a fully rhythmic transcriptome.



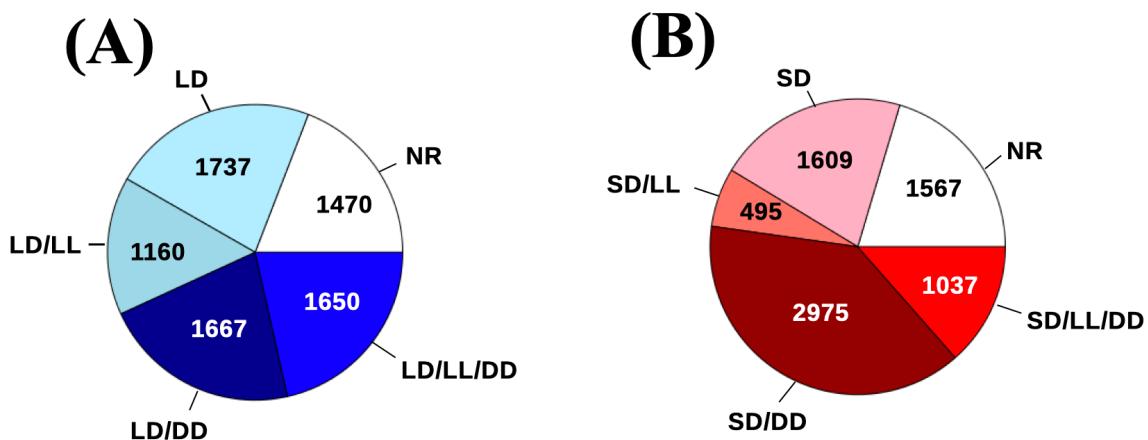
**Figure 21: Diel rhythmic expression patterns detected in the transcriptome of *Ostreococcus tauri* under both photoperiods.** (A) Venn diagram comparing rhythmic genes under LD condition (light blue circle) and SD condition (light red circle); (B) Boxplot representing the maximum expression level of rhythmic genes (R) and non-rhythmic genes (NR). Gene expression levels are measured as FPKM (Fragments Per Kilobase of transcript per Million fragments mapped).

### ***Identification of bona fide circadian genes***

As it was described previously in the introduction, circadian processes are self-sustained and maintain their rhythmicity even when the specific zeitgeber, or synchronizing environmental signal, becomes constant. Following this definition, circadian genes can be identified and distinguished from light or dark responding ones from data generated under free-running conditions consisting in constant light (LL) and / or constant dark (DD).

Specifically, approximately 21 % of the transcriptome (1470 and 1567 genes in LD and SD, respectively) is completely reliant on light/dark cycles to maintain rhythmicity since these genes lose their rhythms under both LL and DD, independently from the photoperiod of entrainment (Fig. 22, lists of genes available at MINOTAUR). In contrast, the transcriptome proportion maintaining its oscillations exclusively under LL or DD is found to be dependent on the previous photoperiod of entrainment. Whereas 1160 genes (15 % of the entire genome) with a previous LD entrainment maintain their rhythmicity only under constant light (LL) condition (Fig. 22-A, list of genes available

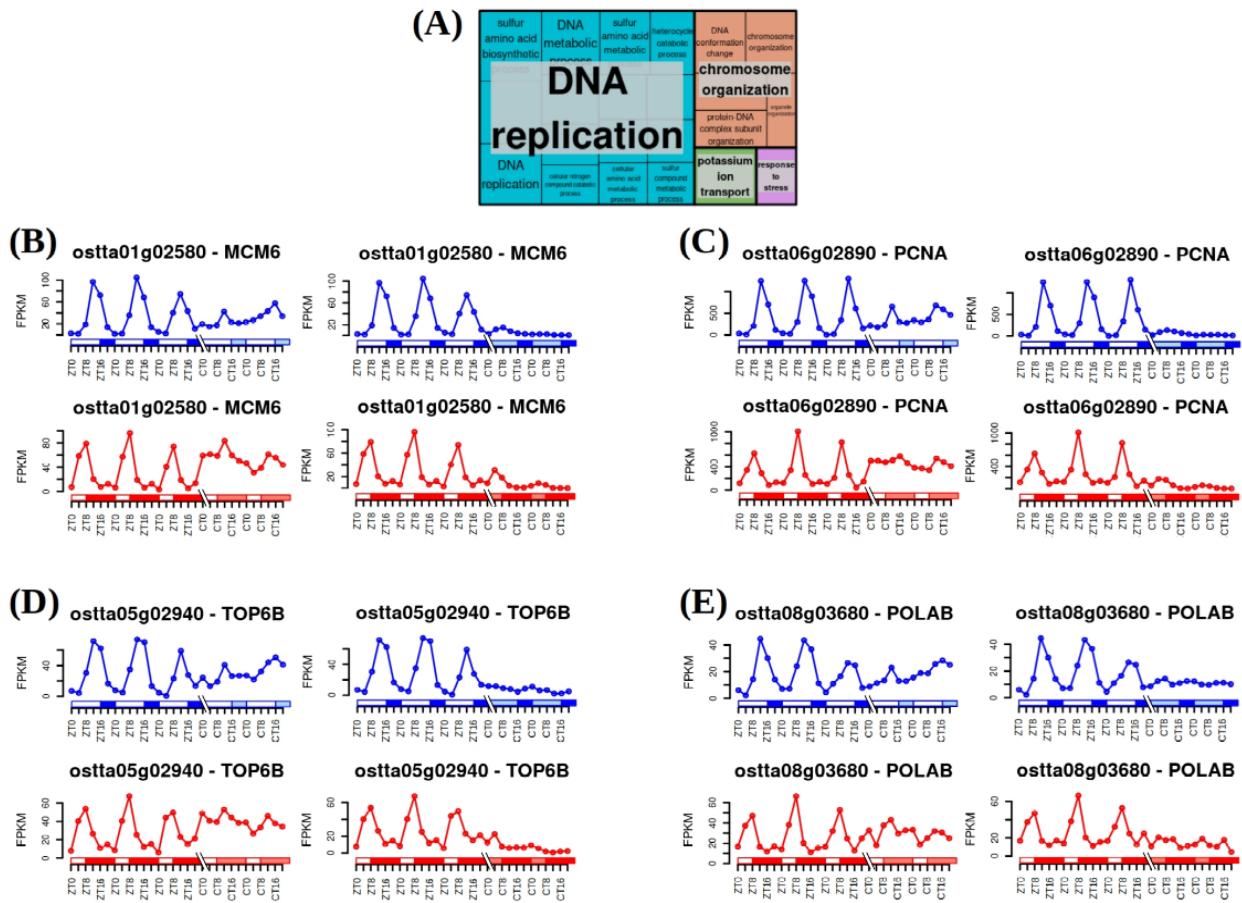
at MINOTAUR), only 495 (6,4 % of the entire genome) are rhythmic under LL after SD entrainment (Fig. 22-B, list of genes available at MINOTAUR).



**Figure 22: Rhythmicity maintenance under free-running conditions:** (A) Pie plot representing with blue colors different rhythmic gene sets under LD entrainment. NR section correspond to the number of non-rhythmic genes, LD for genes that become non-rhythmic under any free running condition after long day entrainment, LD/LL for genes that only maintaining their rhythmicity under LL, LD/DD for the ones maintaining their rhythmicity under DD and LD/LL/DD for genes maintaining their rhythmicity under LL and DD; (B) Pie plot representing with red colors different rhythmic gene sets under SD entrainment. NR section correspond to the number of non-rhythmic genes, SD for genes that become non-rhythmic under any free running condition after short day entrainment, SD/LL for genes that only maintaining their rhythmicity under LL, SD/DD for the ones maintaining their rhythmicity under DD and SD/LL/DD for genes maintaining their rhythmicity under LL and DD. The different sets of genes and their annotation can be downloaded from MINOTAUR.

The detrimental effect of DD over transcriptome rhythmicity is smaller than LL, with almost 22 % (1667 genes) maintaining oscillations after LD entrainment (Fig. 22-A, list of genes available at MINOTAUR) and 39 % (2975 genes) after SD entrainment (Fig. 22-B, list of genes available at MINOTAUR). These results suggest the presence of a dark period is crucial as a zeitgeber for the rhythmic expression of a substantial number of genes. Specially evident for rhythmic gene expression under SD condition which is found to be very dependent on the presence of a dark period. Regulatory mechanisms are often composed of large networks influenced by a wide range of inputs. Circadian clocks are strongly influenced by external environmental signals but there exists a complex interplay between the clock and cell

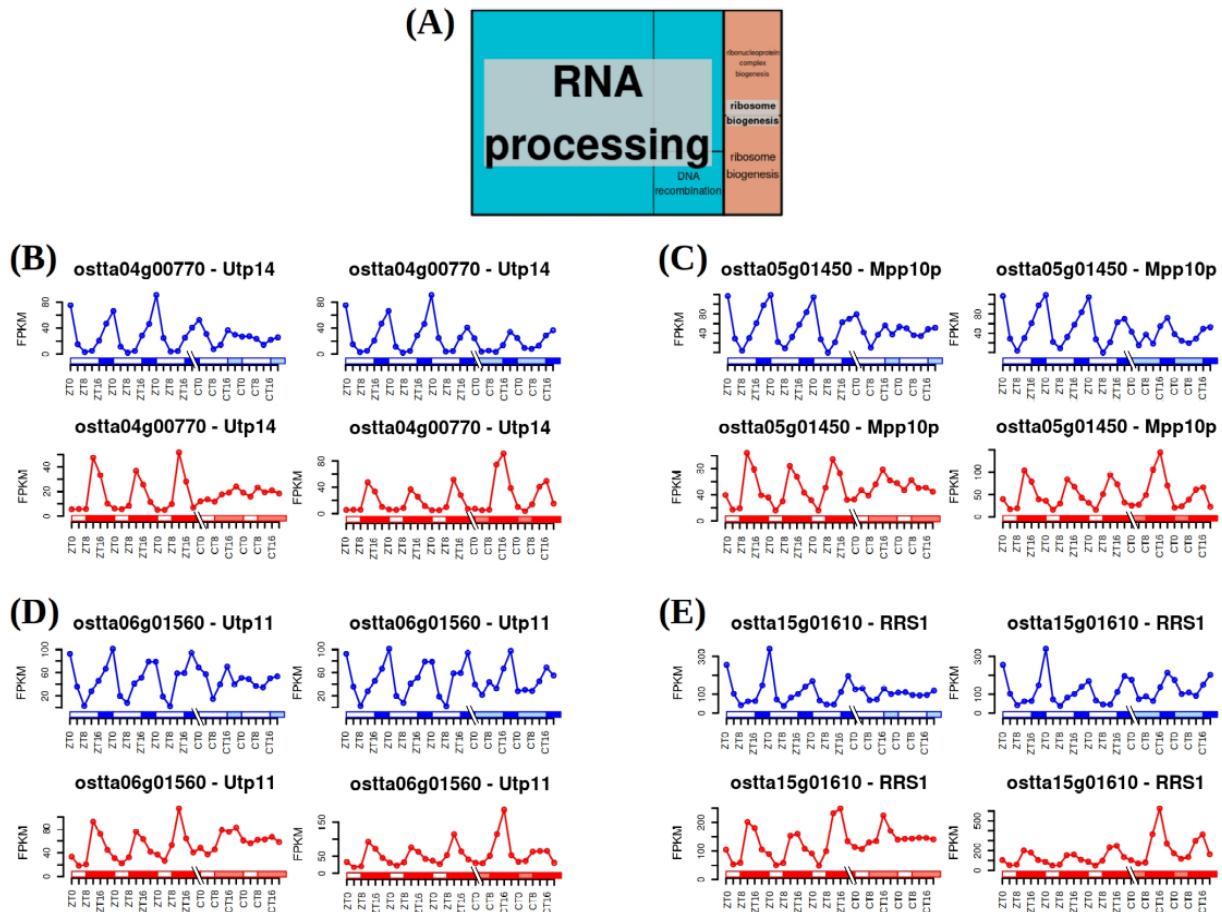
physiology as well (Mazzoccoli et al., 2020; Morris et al., 2020). In line with that observation, genes maintaining their rhythmic expression profiles only under LL or DD could be regulated by the circadian clock, but also influenced by other regulatory mechanisms. The oscillating genes under LL, both after LD and SD entrainment, are found to be significantly involved in DNA replication and chromosome organization (Fig. 23).



**Figure 23: Biological processes significantly enriched in the genes exhibiting rhythmicity under alternating light/dark cycles and constant light.** (A) Treemap summarizing the significantly enriched biological processes. Semantically similar biological processes are grouped into the same colored rectangle. The most representative biological process is shown for each rectangle; (B), (C), (D) and (E) Gene expression profiles under LD/LL (top left), LD / DD (top right), SD / LL (bottom left) and SD / DD (bottom right) condition for Minichromosome Maintenance 6 (MCM6), Proliferating Cell Nuclear Antigen (PCNA), Topoisomerase 6 subunit B (TOP6B) and DNA Polymerase Alpha subunit B (POLAB).

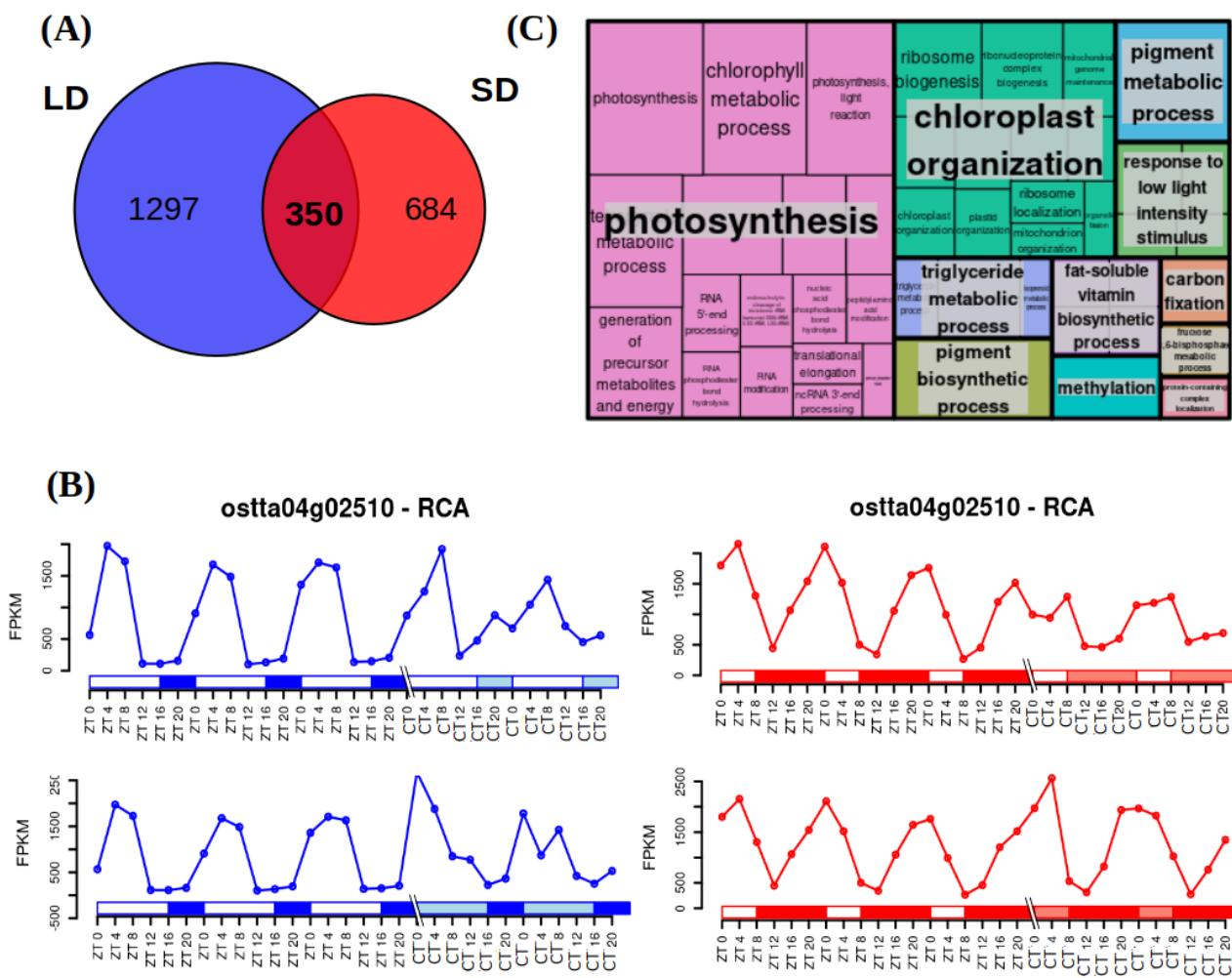
In addition, RNA processing and ribosome biogenesis are found to be significantly enriched biological processes among the rhythmic genes under DD, both after LD and SD entrainment (Fig. 24). These processes are commonly programmed at transcriptomic level to take place

during the night, so translation of proteins can be achieved during the day (Merchant et al., 2017). It could suggest that a dark input is needed in order to respond to a synchronized circadian regulation of those genes, since their activation take place during skotoperiods (dark periods). However, in contrast to the genes involved in DNA replication and chromosome organization (Fig. 23-B, C, D, E) whose expression levels are strongly reduced without the light input, RNA processing and ribosome biogenesis genes (Fig. 24-B, C, D, E) do not seem to be inactivated by the absence of dark but simply desynchronized.



**Figure 24: Biological processes significantly enriched in the genes exhibiting rhythmicity under alternating light/dark cycles and constant dark.** (A) Treemap summarizing the significantly enriched biological processes. Semantically similar biological processes are grouped into the same colored rectangle. The most representative biological process is shown for each rectangle; (B), (C), (D) and (E) Gene expression profiles under LD / LL (top left), LD / DD (top right), SD / LL (bottom left) and SD / DD (bottom right) condition for U3 small nucleolar RNA-associated protein 14 (Utp14), M-phase phosphoprotein 10 (Mpp10p), U3 small nucleolar RNA-associated protein 11 (Utp11) and ribosome biogenesis regulator 1 (RRS1).

*Bona fide* circadian genes are considered those presenting rhythmicity under both LD and SD as well as maintaining their rhythmic expression profiles under both LL and DD free-running conditions. A clear dependence on the previous entrainment regime is observed with 1647 and 1034 genes keeping rhythmicity under both free-running conditions after LD and SD entrainment respectively (Fig. 22, list of genes available at MINOTAUR). These two sets overlapped partially identifying 350 *bona fide* circadian genes corresponding to 4.6 % of the *Ostreococcus* transcriptome (Fig. 25-A, list of genes available at MINOTAUR).

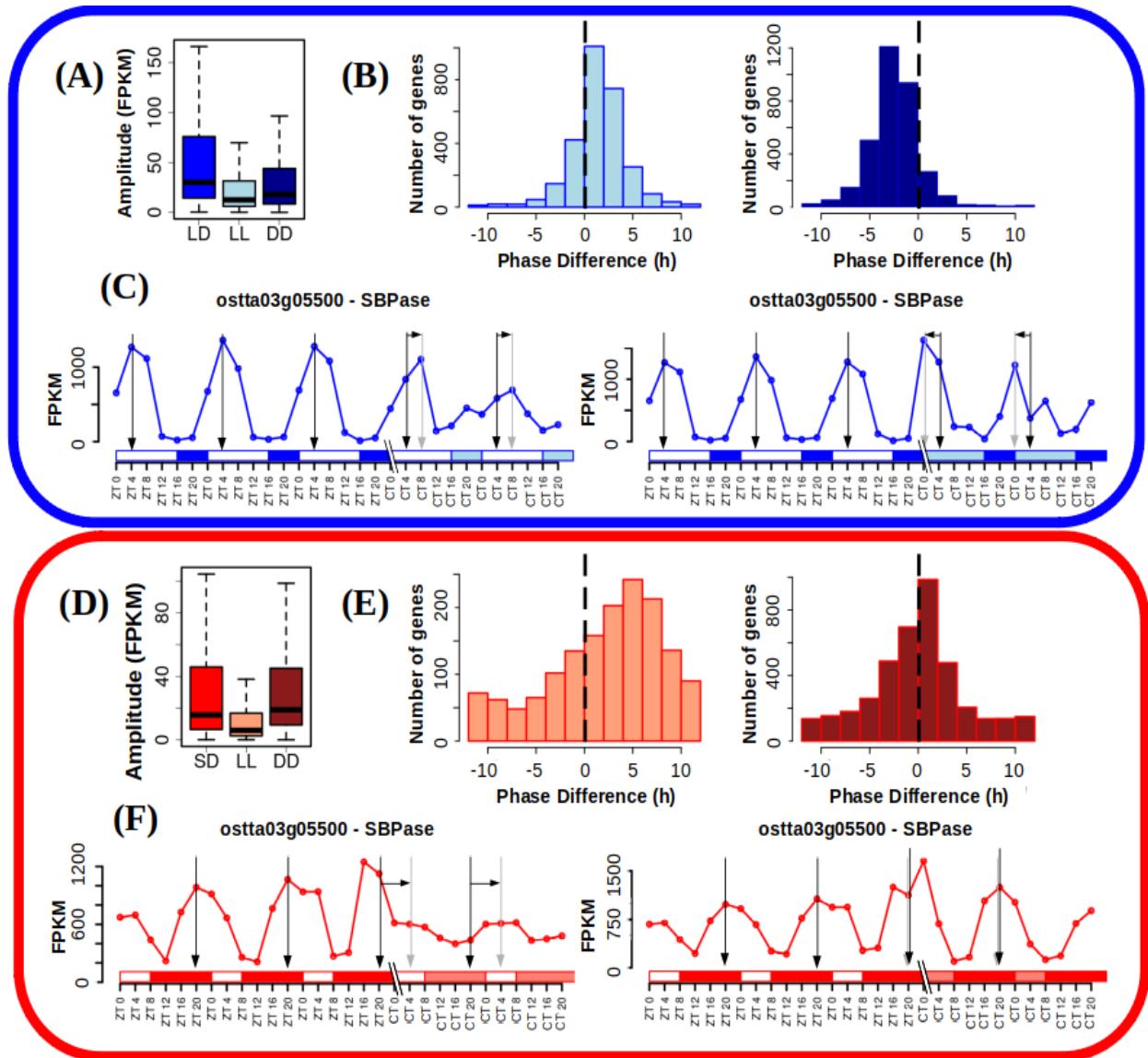


**Figure 25: Identification of bona fide circadian genes and their functional enrichment analysis:** (A) Venn diagram comparing circadian genes identified after LD entrainment (blue circle) and after SD entrainment (red circle); (B) Gene expression profiles under LD (blue), SD (red), LL (top) and DD (bottom) of RuBisCO Activase (RCA); (C) Treemap summarizing the biological processes significantly enriched over the bona fide circadian genes. Semantically similar biological processes are grouped together into the same colored rectangles. The most representative biological process is shown for each rectangle.

In previous work, RuBisCO activase (RCA) and RuBisCO small subunit (RBCS) mRNA were quantified in *Arabidopsis* under a light dark cycle, constant light and constant darkness conditions (Pilgrim & McClung, 1993). Their results are in agreement with our transcriptomic data in *Ostreococcus*, where the gene expression profile of RuBisCO activase (ostta04g02510, RCA) under the different conditions exemplifies how the rhythmicity of *bona fide* circadian genes is self-sustained (Fig. 25-B). In fact, functional enrichment analysis revealed that these genes are significantly involved in photosynthesis, chloroplast organization and processes related to pigment metabolism (Fig. 25-C). Some of those processes are known to present a circadian physiological activity in plants and microalgae like *Euglena*, but there is a lack of confirmation at the transcriptomic level in most cases like the one provided in this work for *Ostreococcus* (Cumming & Wagner, 1968; Noordally & Millar, 2015; Panter et al., 2019).

### ***Rhythmicity analysis of transcriptomes corresponding to free-running conditions***

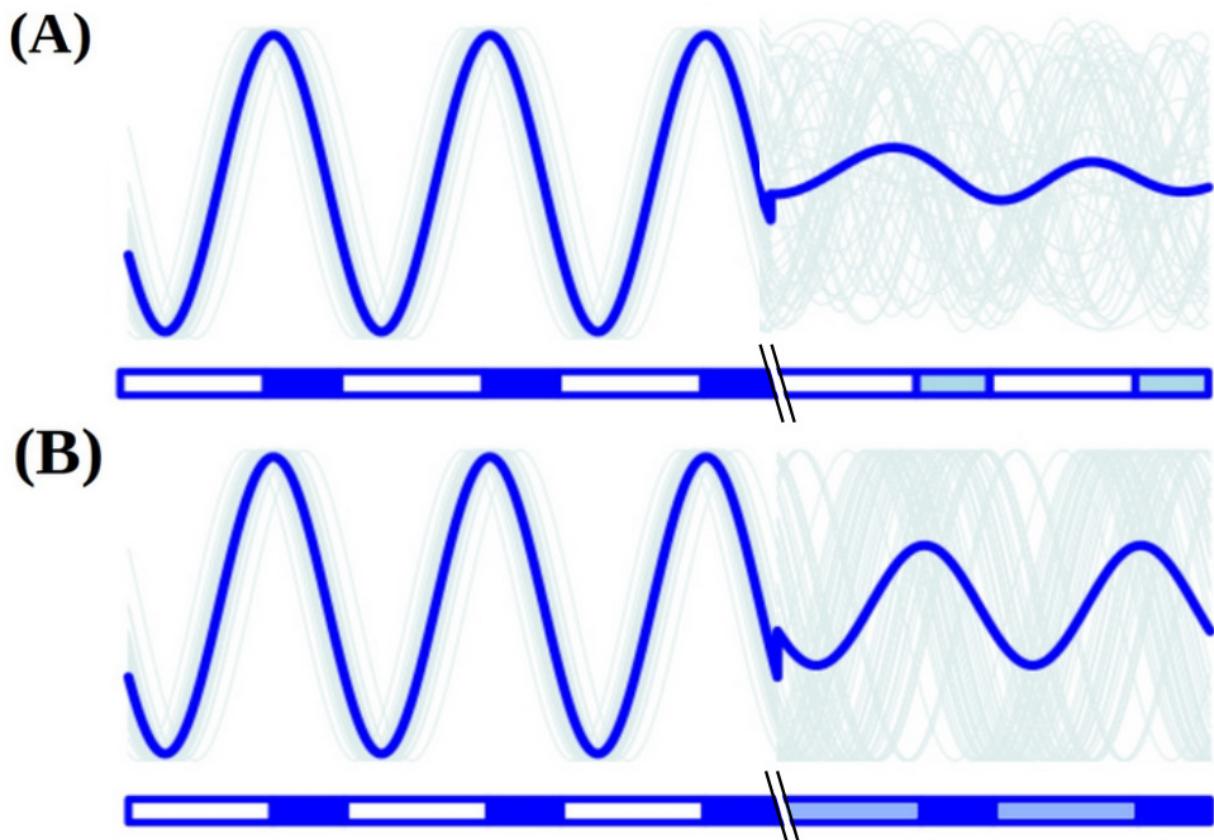
The effect of the transition to constant light (LL) or dark (DD) on rhythmic gene expression patterns is studied using the cosinusoidal parametric method implemented in the R package circacompare. In the LD entrained cultures, significant reductions in amplitude are detected when transferred to both LL and DD, being more drastic under LL according to p-values of  $4.23 \times 10^{-140}$  and  $1.09 \times 10^{-60}$  respectively (Fig. 26-A). Specifically, 97.68 % of the genes maintaining their rhythmicity under free-running conditions present a decrease in amplitude when transferred to LL and 79.43 % when transferred to DD after LD entrainment. Moreover, the reduction in amplitude is significantly stronger under LL than under DD with a p-value of  $6.71 \times 10^{-28}$ . In contrast, in SD entrained cultures, amplitude reduction is only present when transferred to LL with a p-value of  $5.94 \times 10^{-65}$ . This effect is more drastic than the one observed after LD entrainment and it is present on 87.90 % of rhythmic genes. Nonetheless, a slight but significant amplitude increase is found when transfer to DD after SD entrainment, according to a p-value of  $2.50 \times 10^{-8}$  (Fig. 26-D). Free-running conditions also affected the phase (time points when the maximum expression level is reached) of gene expression patterns. Independently from the previous entrainment regime, forward positive phase shifts or delays are detected when cultures are transferred to LL (Fig. 26-B, C, E and F left). Whereas backward negative phase shifts or advances are found when cultures are transferred to DD (Fig. 26-B, C, E and F right). In addition, the phase delays in LL are more drastic for SD than LD entrained cultures although the advances in DD are more evident for LD than SD entrained cultures.



**Figure 26: Free-running conditions effects over gene expression profiles.** (A,D) Boxplot representing rhythmic genes amplitude reached under LD or SD condition (blue top and red bottom, respectively), when cultures were kept under free-running condition consisting of constant light (light blue/red) and when cultures were kept under free-running condition consisting of constant dark (dark blue/red); (B,E) Histograms showing the distribution of the number of genes exhibiting positive and negative phase shifts under LL (light blue/red, left) and DD (dark blue/red, right) free-running condition when compared to LD (top) or SD (bottom). Vertical dashed lines mark no shift; (C-F) Gene expression profiles under LD (top) and SD (bottom) followed by LL (left) and DD (right) of Sedoheptulose-bisphosphatase (SBPase). Vertical black arrows mark LD/SD phases, vertical grey arrows mark LL and DD phases and horizontal black arrows represent phase shifts.

Constant dark or light as free-running conditions have been widely studied in nocturnal mammals (Bartoszewicz et al., 2010; Hundahl et al., 2012; Imai et al., 2020), plants (Edwards et al., 2010; Ohara et al., 2015) and other organisms (Biebach et al., 1991; Vatakis et al.,

2018). The effect over the amplitude of biological rhythms has been previously observed at different molecular levels and is commonly associated with a loss of synchrony (Paajanen et al., 2021; Vatakis et al., 2018). Therefore, although *Ostreococcus* growth is dependent on photosynthesis and, thus, on light, the drastic reduction in amplitude detected under LL is possibly due to a decreased culture synchrony at the transcriptomic level. Individual cell transcriptional programmes could be desynchronized by constant light, leading to largely out of phase individual gene expression profiles, resulting in damped average gene expression oscillations at the entire cell culture level (Fig. 27). Similar desynchronization effects at the single cell level have been reported in *Arabidopsis* leaves under constant light (Wenden et al., 2012; Yakir et al., 2011).



**Figure 27: Reductions in amplitude under free-running conditions can be explained by a decline in culture synchrony.** Figurative culture average gene expression profile under LD followed by LL (A) and DD (B) after a day of acclimation is represented by a thick blue line. Figurative examples of individual cell gene expression profiles are represented by thin grey lines.

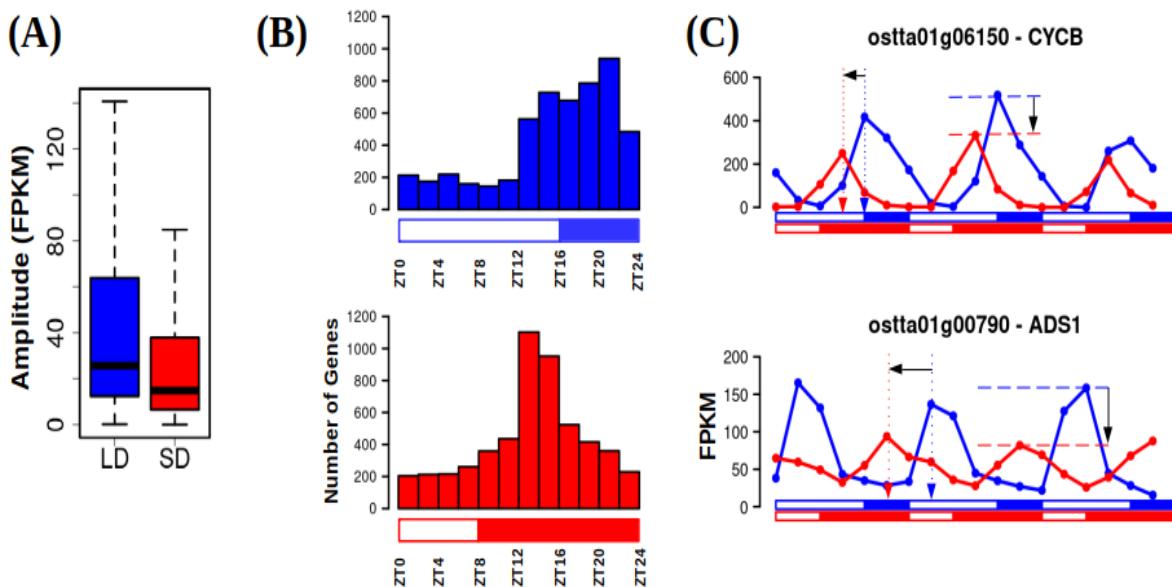
The free-running rhythms observed in *Ostreococcus* could be possibly found in other photosynthetic organisms as well, nonetheless there is considerable lack of research on this topic, since light is currently considered the primary transcriptional zeitgeber or synchronizing

signal in plants (Wang et al., 2022; Wenden et al., 2012), disregarding the relevance of dark periods in regulating diel rhythms at the transcriptomic level. Moreover, analysis of the rhythmic patterns under free-running conditions unveiled typical responses of nocturnal character, such as backward and forward phase shifts under DD and LL, respectively, when compared to LD and SD. This has never been characterized in photosynthetic organisms before, although positive and negative phase shifts under LL and DD have also been observed related to photosynthetic pigments content in algae-coral symbiosis (Sorek et al., 2013). However, similar responses have been observed in behavioral traits in nocturnal animals, where DD condition have been extensively studied (Bartoszewicz et al., 2010; Hundahl et al., 2012; Imai et al., 2020; Vatakis et al., 2018). In fact, constant light is commonly used as a circadian activity disruption model in those organisms. The results presented in this work provides evidence of the importance of dark periods in synchronizing diel cycles in *Ostreococcus tauri*, possibly due to a predominant nocturnal activity of its transcriptome.

## **Transcriptomic characterization of seasonal effects over gene expression profiles**

Specific differences in phase and amplitude between LD and SD rhythmic gene expression patterns have been identified. A significant reduction in amplitude is observed in SD compared to LD entrained cultures, with a p-value of  $1.16 \times 10^{-111}$  (Fig. 28-A). Specifically, more than 2000 genes presented this effect. The reduction in amplitude could be attributed to an increased desynchronization of individual cells under SD when compared to LD condition, as discussed earlier (Fig. 28).

Negative phase shift or phase anticipation is also globally observed over the transcriptome of *Ostreococcus*. Whereas under LD entrainment, gene phases accumulate uniformly from the end of the day, ZT12, to the end of the night, ZT20, under SD entrainment gene phases are concentrated during the first half of the night, from ZT8 to ZT16 (Fig. 28-B). Specifically, 3424 genes comprising 64.95 % of the rhythmic genes exhibit a significantly anticipated phase under SD condition. In addition, only a low number of rhythmic genes exhibit their phase or maximum level of expression during the light period in LD and SD. This indicates that the main transcriptomic activity takes place during the night in *Ostreococcus*, which supports the nocturnal character of its transcriptome as described previously when analyzing the effects of free-running conditions.



**Figure 28: Comparison of amplitude and phase of gene expression profiles under both photoperiods of entrainment.** (A) Boxplot representing rhythmic genes amplitude or maximum expression level reached under LD and SD. SD amplitudes are significantly reduced with respect to LD according to a  $p$ -value of  $1.16 \times 10^{-11}$  computed using Mann-Whitney-Wilcoxon test; (B) Histograms showing the distribution of the number of genes with phase or maximum expression level at specific time points during the day under LD condition (blue, top) and SD condition (red, bottom); (C) Gene expression profiles under LD (blue line) and SD (red line) of Cyclin B (*ostta01g06150*, *CYCB*, top) and Delta-9 acyl-lipid desaturase 1 (*ostta01g00790*, *ADS1*, bottom). Blue and red vertical dotted arrows mark LD and SD phases. Horizontal black arrow represents backward phase shifts under SD when compared to LD. Blue and red horizontal dashed lines mark LD and SD amplitudes. Vertical black arrows represent the reductions in amplitude under SD with respect to LD.

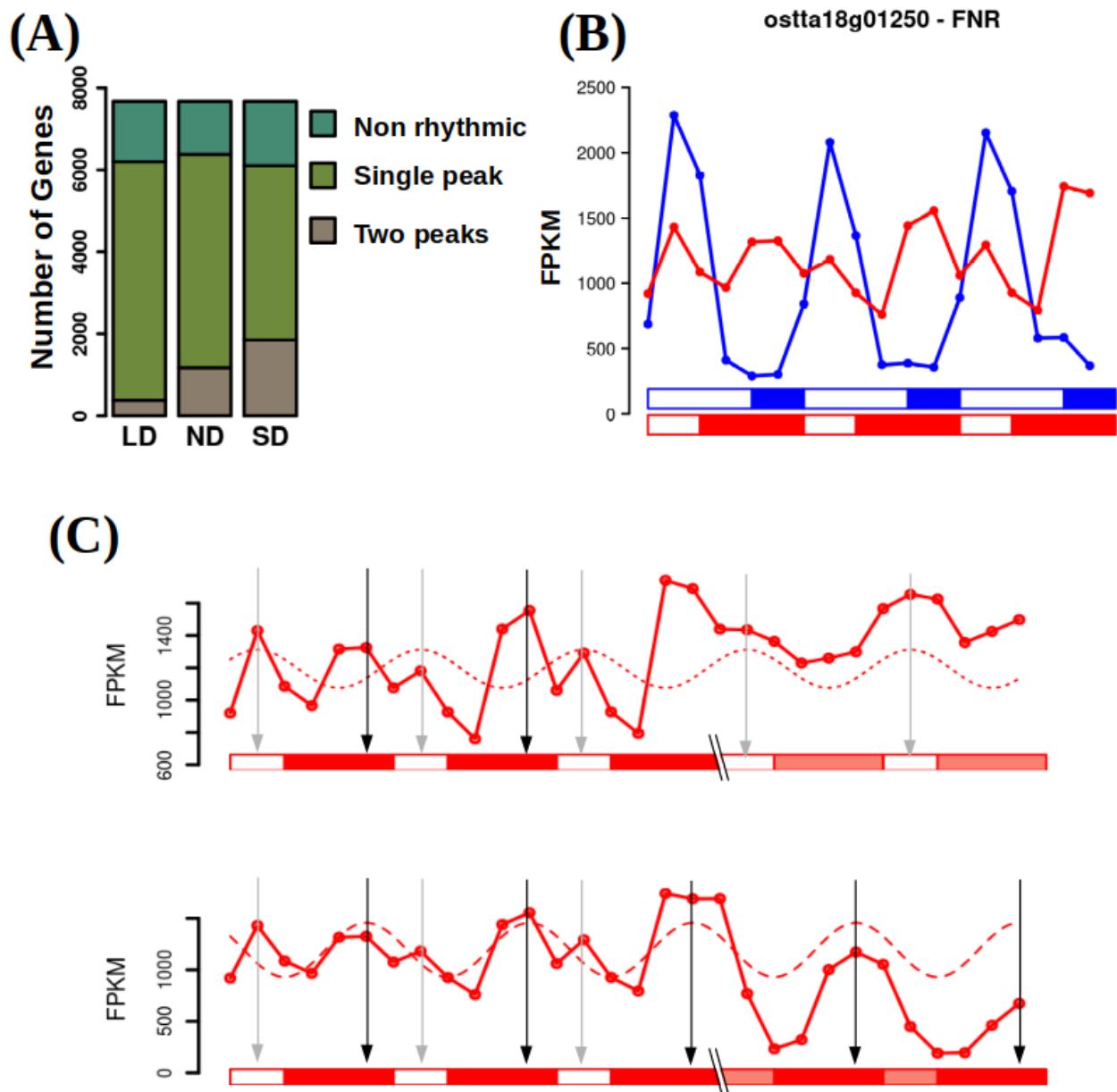
These transcriptomic seasonal responses to changes in photoperiod length are captured in a predictive model based on linear interpolations able to estimate the phase and amplitude of any gene expression profile over the year (Appendix 2). Previously published microarray data from *Ostreococcus*, generated under neutral day condition (12 h light : 12 h dark) (Monnier et al., 2010) have been used to evaluate its predictive power. Model simulations show that approximately two thirds of the rhythmic genes under both LD and SD conditions respond to seasonal changes in photoperiod length by gradually adjusting their phases in accordance with the increasing or decreasing length of the day over the year.

Photoperiodic dependent changes in amplitude and phase shifts in biological rhythms have been already described in mammals (Messenger et al., 2000; Sumová et al., 2003; Van Dongen

et al., 1997; Wucher et al., 2022), as well as in plants and microalgae (Flis et al., 2016; Panter et al., 2019; Serrano et al., 2009). Depending on the organisms under study, increments or decrements in amplitude as a response to shortening or lengthening photoperiods have been reported. Global phase shifts over the transcriptome of *Arabidopsis thaliana* as an adaptation to changes in photoperiod has been described, and changes in amplitude were observed in individual genes in both *Arabidopsis* and *Chlamydomonas*. For example, the rhythmic expression profile of the gene CrCO potential ortholog of the CONSTANS gene in *Arabidopsis thaliana* shows an increase in amplitude under SD condition as well as a negative (backward) phase shift (Flis et al., 2016; Serrano et al., 2009). Same effects linked to seasonal changes can be observed at the gene level in *Ostreococcus*. For example, Cyclin B (ostta01g06150, CYCB) and Delta-9 acyl-lipid desaturase 1 (ostta01g00790, ADS1) exhibit the described phase anticipation and amplitude reduction under SD condition when compared to LD condition (Fig. 28-C).

### ***Analysis of rhythmic profiles with 12 h period***

When comparing SD and LD conditions, another phenomenon has been identified over the transcriptome of *Ostreococcus*, emerging as a response to photoperiod shortening. Under LD condition, almost every rhythmic gene (5825 genes covering 75.97 % of the entire genome) reach its maximum level of expression once a day, presenting one single peak every 24 h in its expression profile (Fig. 29-A, B). However, more complex rhythmic expression profiles with two expression peaks per day (every 12 h) are found in a small number of genes, specifically 376 genes (Fig. 29-A, list of genes available at MINOTAUR). Under SD condition the number of genes presenting one single expression peak per day decreases to 4249 (55.41 % of the entire genome). This is coupled to an increasing number of genes, namely 1855 genes, presenting two expression peaks per day (Fig. 29-A, B, list of genes available at MINOTAUR). When re-analyzing a published microarray data set generated from cultures of *Ostreococcus tauri* under neutral day condition (12 h light:12 h dark) (Monnier et al., 2010), bimodal gene expression patterns have been also found. Most of the rhythmic genes present rhythmic expression profiles with a 24 hours period and, thus, a single expression peak. However, 1171 genes present rhythmic expression pattern with an apparent period of 12 hours (two peaks of expression per day). This photoperiod can be considered an intermediate step between the two extreme photoperiods studied in this thesis (LD and SD). Indeed, the data show an intermediate number of genes with two peaks of expression per day between the ones identified in this thesis for LD and SD conditions (Fig. 29-A).



**Figure 29: Emergence of 12 h period rhythmic expression profiles under SD.** (A) Barplot representing in different green colors the number of non rhythmic, single peak rhythmic and two peaks rhythmic genes under LD, ND and SD conditions; (B) Gene expression profiles under LD (blue line) and SD (red line) of Ferredoxin-NADP<sup>+</sup> reductase (*ostta18g01250*, FNR); (C) Gene expression profiles under SD and free-running conditions consisting of constant light (LL) or constant dark (DD) of Ferredoxin-NADP<sup>+</sup> reductase (*ostta18g01250*, FNR). The two unimodal profiles extracted from the real bimodal profile detected are represented: the one depending on the photoperiod as dotted line with phase marked with a grey vertical arrow maintaining its rhythmicity only under LL (top); and the one depending on the skotoperiod (dashed line) with phase marked with a black vertical arrow maintaining its rhythmicity only under DD (bottom).

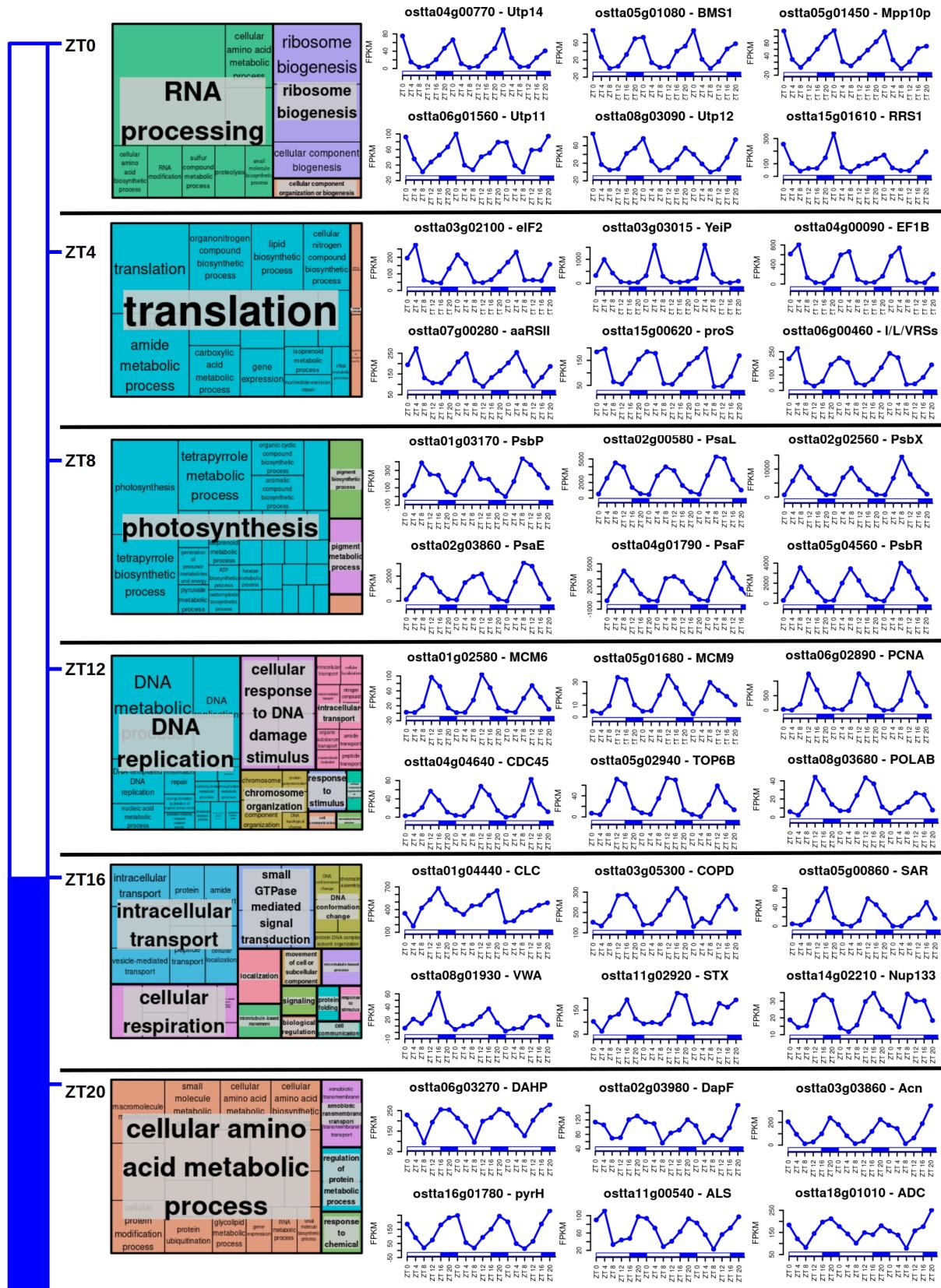
Also, in agreement with our results, this bimodal rhythmic pattern can be found in data generated from other organisms like *Chlamydomonas* under neutral day (Zones et al. 2015). This suggests that, at least in *Ostreococcus* (and possibly other microalgae), there is an increasing number of genes that present high levels of expression twice a day (every 12 h) as the photoperiod gets shorter, although its biological role remains to be determined.

Tides take place following rhythms of approximately 12 h and can give rise to circatidal gene expression patterns which persist under free-running conditions (Fauré-Fremiet, 1951; Rock et al., 2022). Biological rhythms with specifically 12 h periods are hypothesized to be a reminiscent of the ~ 12 h circatidal rhythms of coastal and estuarine organisms (Ballance & Zhu, 2021). The maintenance of these rhythms after evolving to live on land could provide an advantage in the adaptation to metabolic stress that peak at transition periods during diel cycles (Pan et al., 2020; B. Zhu et al., 2017). Indeed, bimodal rhythms presenting two peaks per day that are maintained under free-running conditions have been observed in animals (Binkley & Mosher, 1985; Foà & Bertolucci, 2001; Kyorku & Brady, 1994; Prabhakaran & Sheeba, 2012; Watanabe et al., 2007), plants (Hayes et al., 2010; Van Gelderen, 2020) and some microalgae like *Euglena* (Mohabir & Edmunds, 1999). They have been described only for specific processes, genes or compounds, not as a global response in the transcriptome to external signals. However, although tides play a central role in the dynamics of the natural environment of *Ostreococcus*, this hypothesis could be discarded in this study since the observed 12 h period rhythmic gene expression profiles in *Ostreococcus* transcriptome are not maintained under free-running conditions. Instead, bimodality disappears and only one of the expression peaks is maintained under constant light whereas the other one persists only under constant darkness (Fig. 29-C). This observation suggests that the seasonal effect observed over the transcriptome of *Ostreococcus* is not a self-sustained 12 h rhythm, but a combination of two distinct rhythmic profiles: one dependent on the photoperiod or light period (maintaining its rhythmicity only under LL) and another one sustained by the skotoperiod or dark period (maintaining its rhythmicity only under DD). To support this hypothesis, a model capturing this response is developed. First, a decomposition is performed over the bimodal gene expression profiles present under SD conditions (based on Non Linear Squares tools (Baty et al., 2015)) into two cosinusoidal single peak profiles, one peaking during the photoperiod and another one peaking during the skotoperiod (Fig. 29-C). Then, linear interpolation is used to simulate phase shifts of these two unimodal profiles as a response to photoperiod shortening. As a validation of its predictive accuracy, this model correctly predicted the emergence of

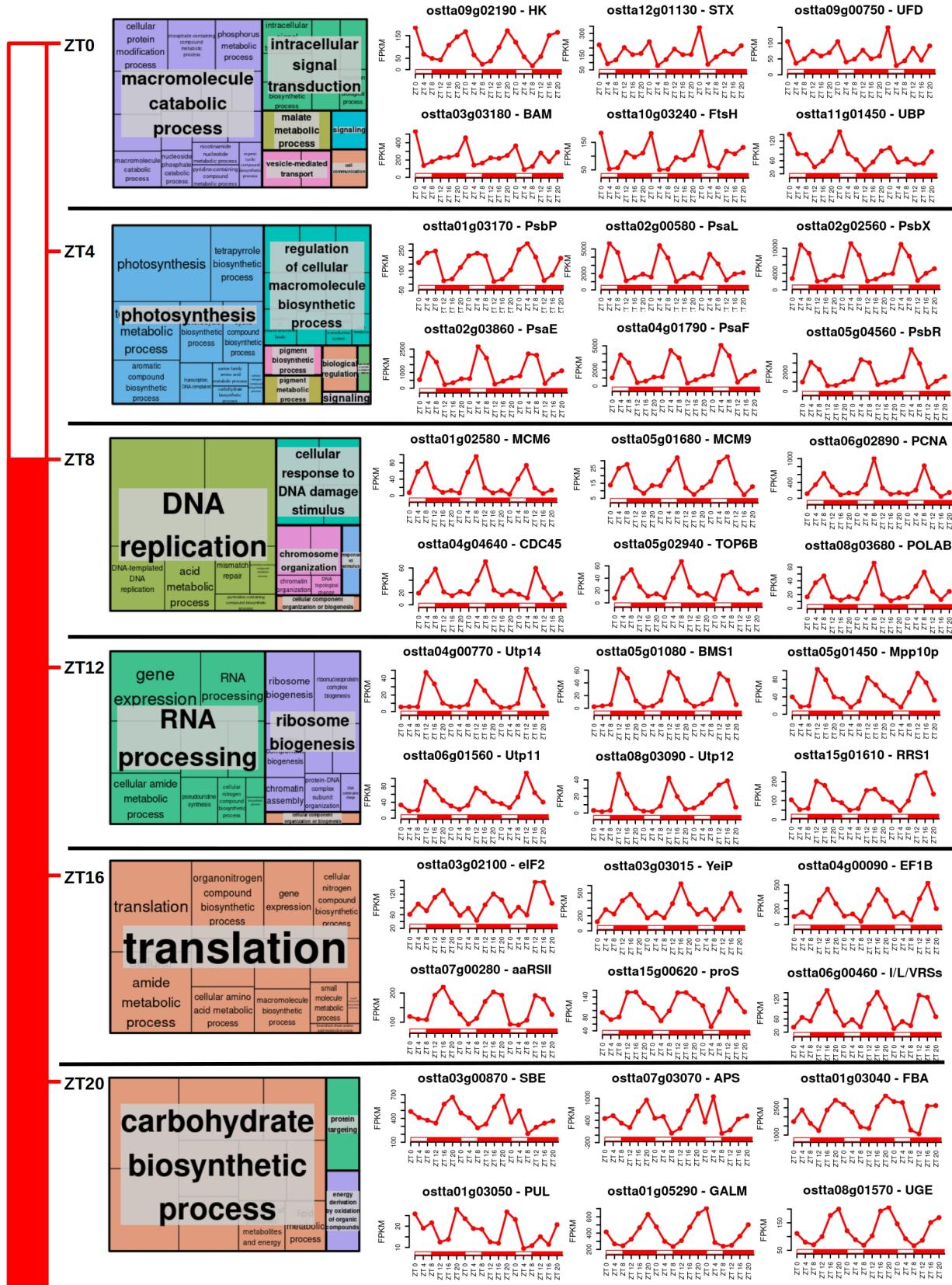
bimodal rhythmic profiles under neutral conditions. As a result, model simulations show how, under LD condition, the two distinct profiles regulated independently by the photoperiod and skotoperiod overlap in time, producing a single peak, while as the photoperiod shortens and the skotoperiod lengthens they become out of phase resulting in a bimodal profile under SD condition. Suggesting a joint but independent regulation exerted by the photoperiod and skotoperiod in the expression of this set of genes (Fig. 29-C, Appendix 3).

### ***Characterization of temporal transcriptional programmes during diel cycles***

In order to determine the temporal organization of the transcriptional programme in *Ostreococcus* under summer LD and winter SD conditions, gene clusters (available at MINOTAUR) have been defined based on their phases (when their maximum level of expression is reached). Functional enrichment analyses of the resulting gene clusters reveal the cellular processes activated at the transcriptomic level in each temporal point. Under long day condition (Fig. 30), *Ostreococcus* activates genes involved in RNA processing and ribosome biogenesis at dawn (ZT0). Examples for such genes are U3 small nucleolar RNA-associated protein 14 (ostta04g00770, Utp14) and the Ribosome Biogenesis Factor BMS1 (ostta05g01080, BMS1). During the first part of the morning (ZT4) *Ostreococcus* transcriptome is almost completely focused on genes involved in translation, such as eukaryotic Initiation Factor 2 (ostta03g02100, eIF2) and translation elongation factor P (ostta03g03015, YeiP). The genes mainly involved in photosynthesis but also in carbohydrate metabolism reach their maximum expression level during midday, when irradiance is maximum (ZT8). Some of those genes are subunits of both photosystems: ostta01g03170 (PsbP), ostta02g00580 (PsaL), ostta02g02560 (PsbX), ostta02g03860 (PsaE), ostta04g01790 (PsaF), ostta05g04560 (PsbR), etc. During the afternoon (ZT12), the activation of genes involved in DNA replication takes place. Some minichromosome maintenance proteins ostta01g02580 (MCM6) and ostta05g01680 (MCM9)) are found, as well as the proliferating cell nuclear antigen ostta06g02890 (PCNA) which is central to the DNA replication process. Intracellular transport and cellular respiration are the two most prominent biological processes whose genes reach their maximum expression level at dusk (ZT16) under LD condition. Genes like lysophospholipases ostta01g04440 (CLC) or nucleoporins ostta14g02210 (Nup133) are some examples for this time point. Finally, during midnight in LD, *Ostreococcus* focuses on expressing genes involved mainly in cellular amino acid metabolic process, like 3-Deoxy-D-arabinoheptulosonate 7-phosphate synthase ostta06g03270 (DAHP), which encodes the first enzyme in the biosynthesis of the amino acids phenylalanine, tyrosine and tryptophan.



**Figure 30: Cellular processes transcriptionally activated in each time point under LD condition.** For each cluster of genes with phase in the different time points, there are: a treemap summarizing the biological processes significantly enriched and expression profiles representation of relevant genes.

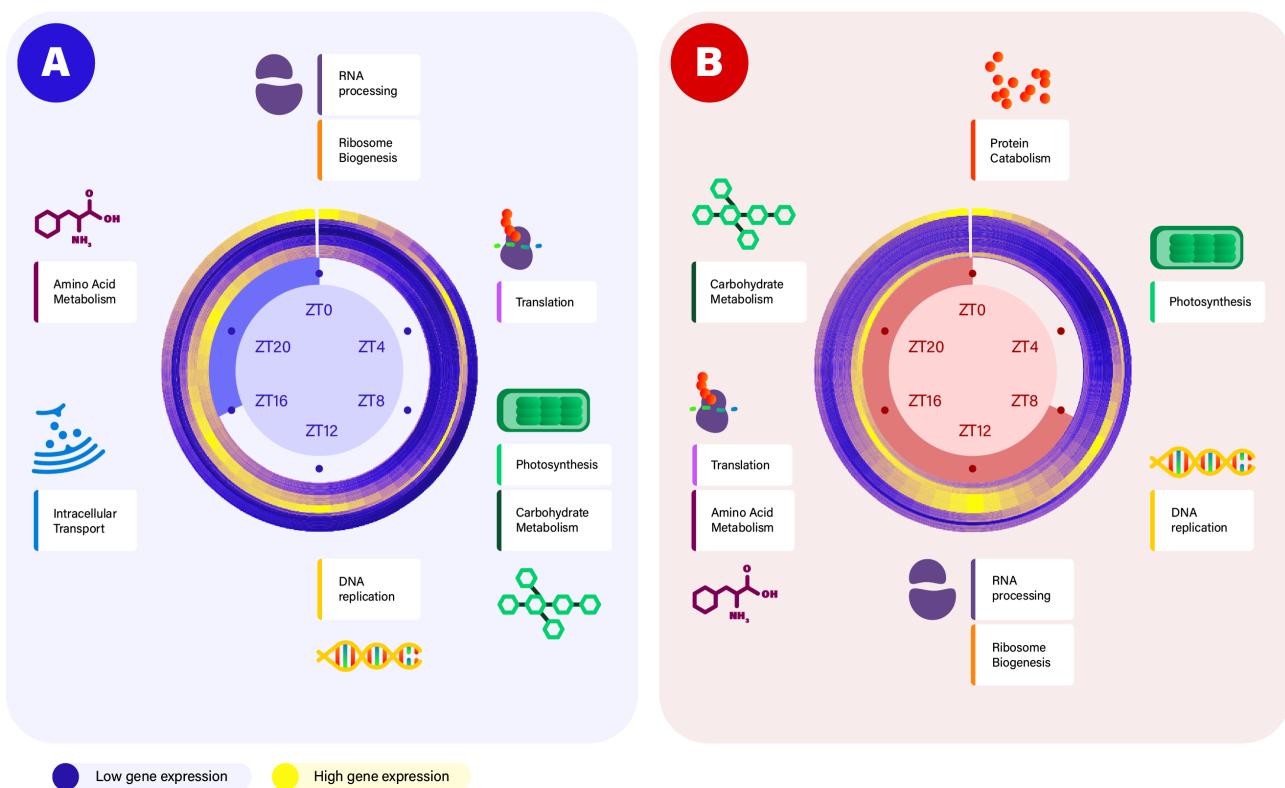


**Figure 31: Cellular processes transcriptionally activated in each time point under SD condition.** For each cluster of genes with phase in the different time points, there are: a treemap summarizing the biological processes significantly enriched and expression profiles representation of relevant genes.

Under short day condition ([Fig. 31](#)), the most prominent biological process whose genes reach their maximum expression level at dawn (ZT0) is protein catabolism. Examples for such genes are Signal transduction Histidine Kinase ostta09g02190 (HK), or Ubiquitin Fusion Degradation protein ostta09g00750 (UFD). In agreement with what was observed in long day condition, genes having their peak of expression during midday (which in short day takes place at ZT4, maximum irradiance time) are involved in photosynthesis. Once again, genes encoding photosystems subunits are clustered in this time point. When the dusk comes at ZT8, genes involved in DNA replication show their maximum level of expression during this time point. The same genes used as an example in LD are found also in SD but anticipating their peak of expression 4h. Then, RNA processing and ribosome biogenesis are the two most prominent biological processes early during the night (ZT12), showing an anticipation of 12 h compared with LD temporal programme. In addition, genes involved in translation also show a 12 h anticipation reaching its maximum expression level during midnight (ZT16) and genes involved in carbohydrate metabolism reach their maximum expression level after midnight at ZT20.

Focusing on the most enriched processes in each time point, a transcriptional temporal programme of *Ostreococcus* under each photoperiod is illustrated in [Fig. 32](#). These programmes consists of distinct times for the activation of biological processes, produced by the previously mentioned gene phase backward shifts under SD with respect to LD ([Fig. 28-B](#)). It is observed as simple anticipation in time for some processes, resulting in the same sequential order of activation but 4-12 h earlier in SD. For example, genes involved in photosynthesis peak at ZT8 under LD ([Fig. 32-A](#)) and at ZT4 under SD ([Fig. 32-B](#)), coinciding with the moment of maximal light irradiance in both cases. Similarly, genes involved in DNA replication reach their maximum expression level at the end of the day under LD, ZT12 ([Fig. 32-A](#)), and at dusk under SD, ZT8 ([Fig. 32-B](#)). These are clear examples of set of genes that present phase shifts caused by seasonal changes that simply adjust their phases in order to align with specific times of the day, like midday and evening times, in any photoperiod and they take place always in the same sequential order (photosynthesis always before DNA replication). However, rearrangements in the sequential order of activation of specific biological processes are also found. For instance, under LD conditions, genes involved in amino acid biosynthesis, ribosome biogenesis, and translation reach their full activation sequentially at ZT20, ZT0 and ZT4, respectively ([Fig. 32-A](#)). However, in SD entrained cultures, first ribosome biogenesis genes are fully activated at ZT12 and subsequently genes involved in amino acid biosynthesis

and translation reach their maximum expression level at the same time (ZT16) (Fig. 32-B). A re-arrangement in time is also observed for carbohydrate metabolism genes, which align with the activation times of photosynthesis genes in LD (both take place at midday) but reach their full activation 8 h before photosynthesis genes do in SD entrained cultures (Fig. 32). This suggests that phase shifts globally observed over the transcriptome of *Ostreococcus tauri* as an adaptation to photoperiods, not only consist in anticipating processes but also in a re-arrangement of the complete diel gene activation programme.



**Figure 32: Summary of the transcriptional temporal programme of *Ostreococcus tauri* under LD and SD conditions.** Circular heatmap representing the temporal organization of gene expression profiles under LD condition (A) and SD condition (B). Dark blue stands for low expression whereas yellow represents high expression. Genes are clustered depending on their phase or time of maximum expression where genes with phase at ZT0 are located in the outer circle and genes with phase at subsequent ZTs are placed sequentially into inner circles. The most enriched processes in each gene cluster are illustrated capturing the diel gene activation programme under LD and SD condition.



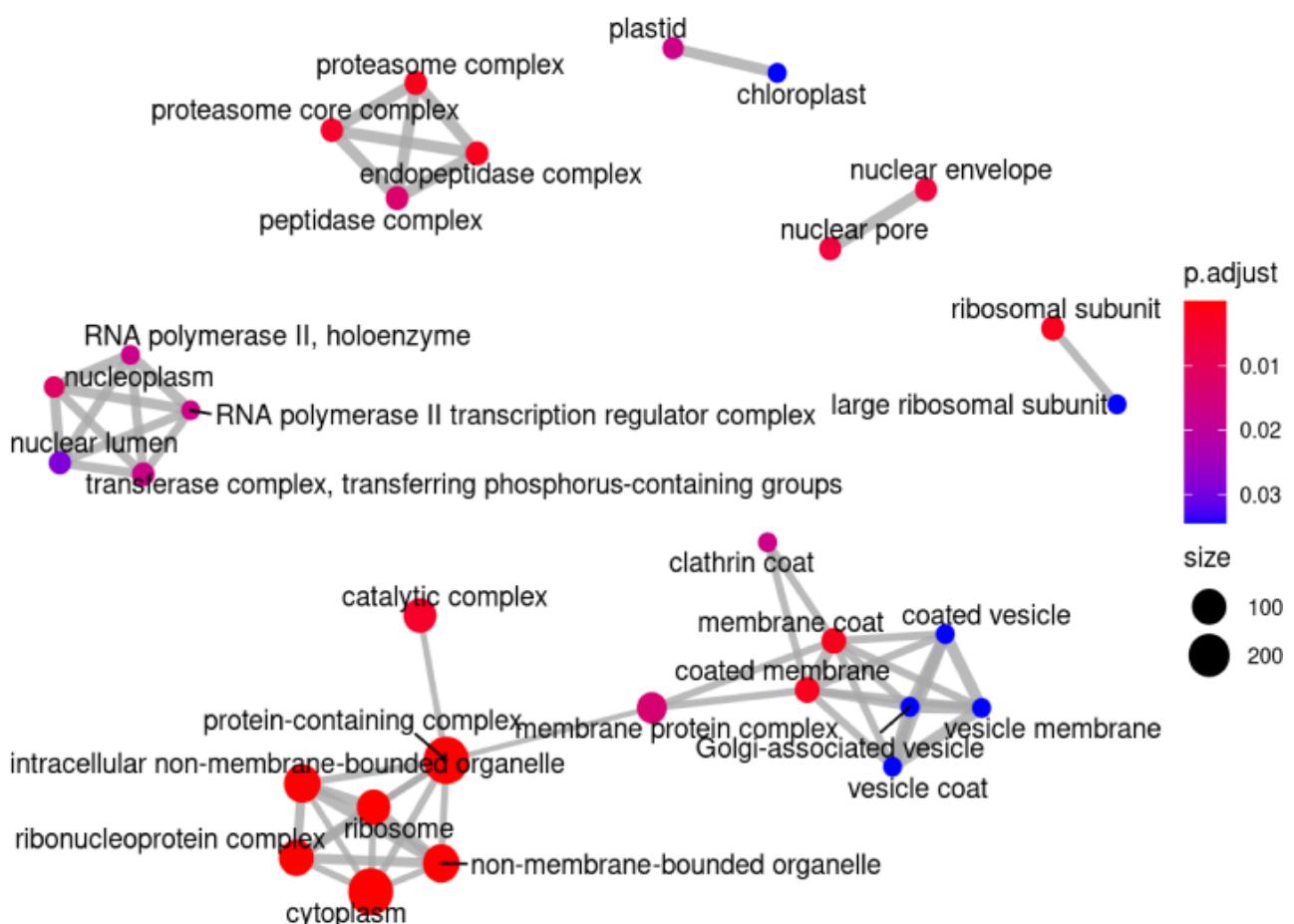
# **Chapter 3.**

## **Proteomic analysis of seasonal variations in diel cycles in *Ostreococcus tauri***



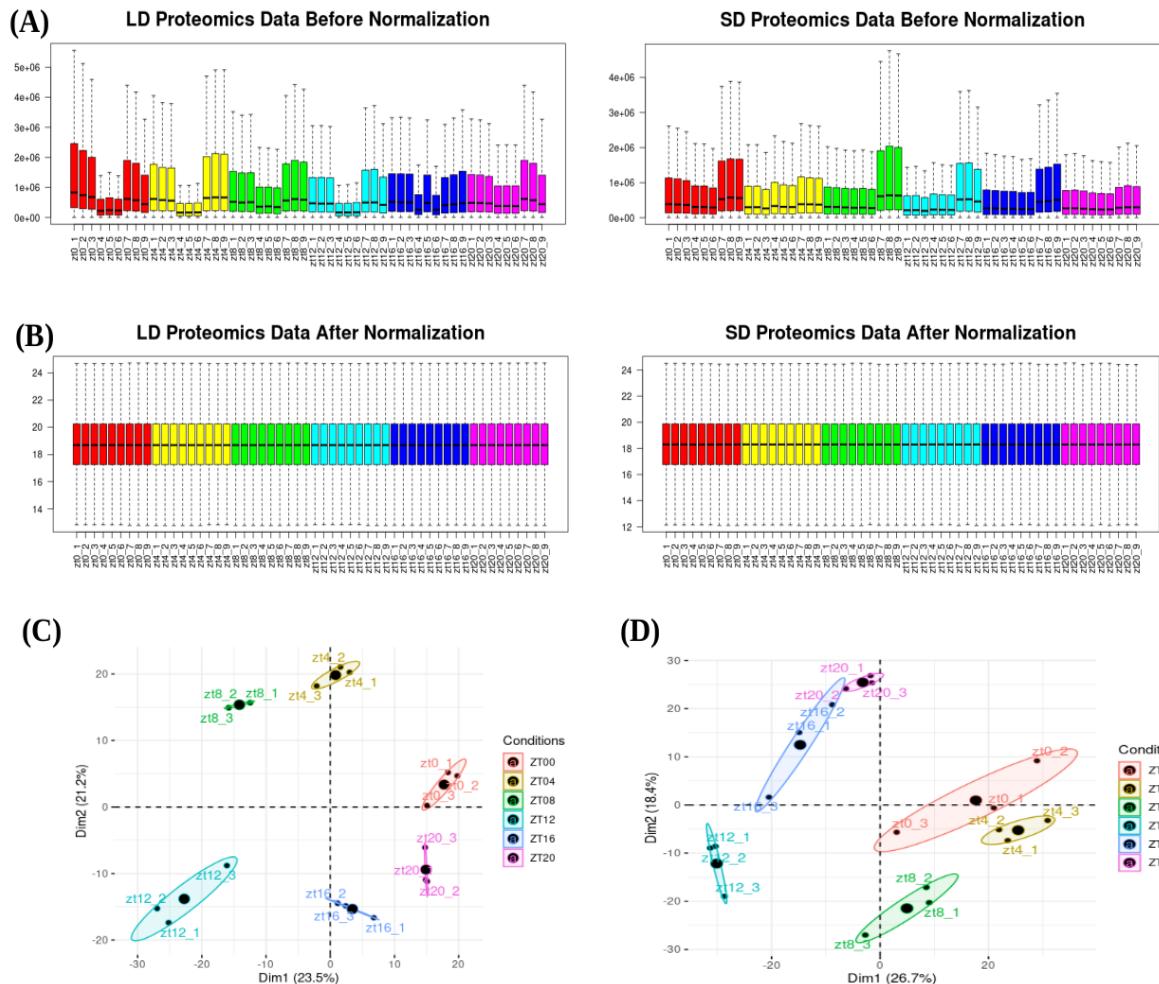
## Chapter 3. Proteomic analysis of seasonal variations in diel cycles in *Ostreococcus tauri*

Proteins are the primary molecular entities that carry out the majority of biological processes in living cells. Therefore, proteomic analysis, which involves the identification and quantification of proteins, play a crucial role in characterizing the responses of living organisms to diel cycles (Kay et al., 2021; Le Bihan et al., 2011). In this study, a total of 3672 proteins have been successfully identified and quantified under LD or SD conditions, accounting for approximately 48 % of the *Ostreococcus* annotated proteome. In addition, the subcellular location of the detected proteins cover most cellular components or organelle (Fig. 33).



**Figure 33: Enrichment map representing the cellular components or organelle significantly covered by the proteins detected.** Dots or nodes sizes represent the number of proteins identified located in the corresponding organelle. A blue to red gradient is used to represent the level of significance. Lines or edges link dots or nodes representing related organelle.

Since our data is obtained from 3 technical replicates of each biological replicate (3 light/dark cycles), it can be observed that the global distribution of each group of 3 technical replicates is similar in the raw data (Fig. 34-A). The differences observed over the global distribution of different groups of technical replicates are solved with log2 quantile normalization of the data, ensuring data comparability (Fig. 34-B).



**Figure 34: Proteomic data normalization and reliability.** (A) Boxplots representing global distributions (before normalization) of protein abundances over the technical and biological replicates of long day condition on the left and short day condition on the right. Boxes corresponding to the replicates of the same time point are represented in the same color, as the legend in the bottom shows; (B) Boxplot representing global distributions after normalization; (C-D) Principal Component Analysis of the time point global rhythmic proteomes under LD and SD condition, respectively. Small dots correspond to the 2D projection of each time point global rhythmic proteome. Big dots correspond to the average of the three replicates 2D projections for each time point. Ellipses mark the 95 % confidence regions corresponding to each time point global rhythmic proteome.

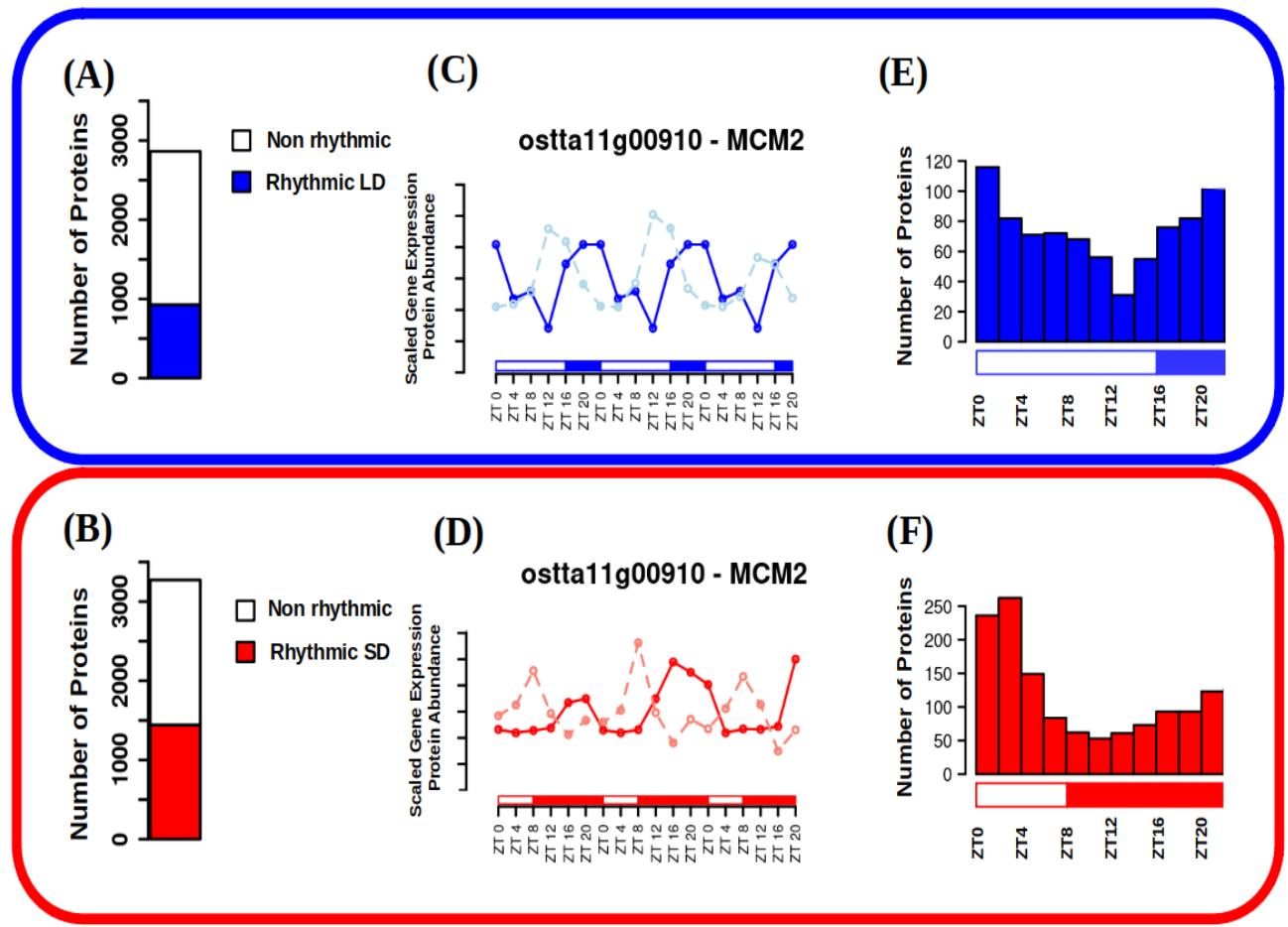
In order to obtain a deeper understanding of the underlying structure in the proteomic data, principal components analysis (PCA) is performed separately over the LD ([Fig. 34-C](#)) and SD ([Fig. 34-D](#)) rhythmic proteomes. Under LD condition, the biological replicates tightly cluster constituting a circular structure ([Fig. 34-C](#)) like the one observed over the transcriptomic data ([Fig. 20-B](#)). Under SD condition more variability is observed, as it was also observed over the SD transcriptomic data ([Fig. 20-C](#)), but the circular structure is still maintained in this case ([Fig. 34-D](#)). Also, the high similarity between the transcriptomes corresponding to late-night and sunrise in SD condition ([Fig. 20-A](#)), ZT20 and ZT0, is not present in proteomic data. This suggests that the anticipation to the light hours is present only at the transcriptomic level, activating during the late-night hours the transcription of genes that will be translated into proteins during the morning. In fact, a higher differentiation between the night samples and the light samples is observed in the proteomic data than in the transcriptomic data ([Fig. 34-C, D](#)). In overall, these results support that the experimental design grants good quality data and a high level of synchronization in this rhythmic proteomic data allowing to proceed to the identification and comparison of rhythmic proteins under LD and SD conditions, which can be also explored individually accessing the MINOTAUR tool.

## Proteomic characterization of diel rhythmic abundance profiles

Similarly to transcriptomic data analysis, the R packages RAIN and circacompare were used for proteomic data rhythmicity analysis. In LD entrained cultures, 928 proteins, comprising approximately 32 % of the total proteins detected, present diel rhythmic abundance profiles ([Fig. 35-A](#), list of proteins available at MINOTAUR). An increase in rhythmicity is observed under SD condition, detecting 1442 rhythmic proteins comprising 44 % of the detected proteome ([Fig. 35-B](#), list of proteins available at MINOTAUR). In contrast to the results on transcriptome rhythmicity, seasonality has a major effect over the identity of the rhythmic proteins: only 9 % of the detected proteome present rhythmicity under both conditions. In addition, the rhythmicity levels in the proteome are much lower than the ones observed in the transcriptome of *Ostreococcus tauri*, indicating a more dynamically cycling character in the transcriptome than in the proteome under diel cycles.

The decrease in proteome rhythmicity has also been reported under ND conditions where 67 % of the detected proteome present rhythmic abundance profiles ([Kay et al., 2021](#)) while more than 80 % of the transcriptome is known to be rhythmic under these light conditions ([Monnier et al., 2010](#)). Therefore, transcripts rhythmic expression profiles do not seem to always lead to rhythmic

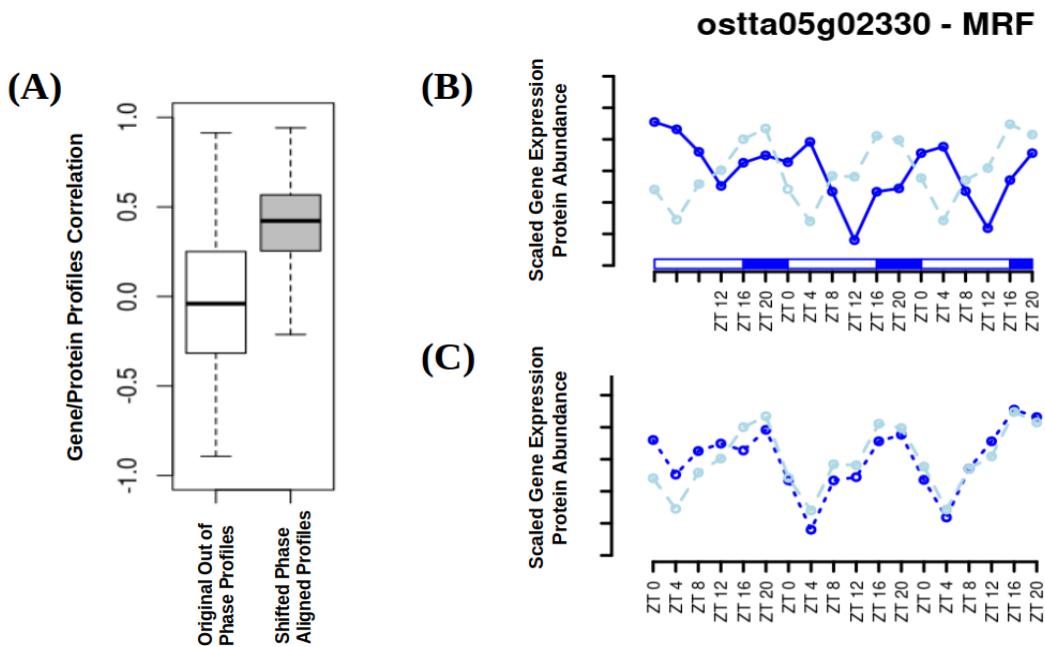
protein abundance profiles. This phenomenon highlights the relevance of multi-omics integration, since transcripts behavior cannot always be extrapolated to proteins or cellular processes, neither vice versa.



**Figure 35: Proteome rhythmicity under alternating light/dark cycles.** (A-B) Barplots representing the number of identified proteins under long day (A) and short day (B) conditions; (C-D) Protein abundance profiles under LD (C) and SD (D) conditions represented together with gene expression profiles under LD (light blue, dashed line) and SD (light red, dashed line) conditions for Minichromosome Maintenance 2 (*ostta11g00910*, MCM2); (E-F) Histograms showing the distribution of the number of proteins with phase or maximum abundance at specific time points under LD (E) and SD (F) conditions

Rhythmic protein abundance profiles are compared to the corresponding rhythmic gene expression profiles detecting phase offsets between them (Fig. 35-C, D). It suggests that translation may take place few hours after gene expression. In fact, while the majority of the transcripts have their maximum expression level during night hours in both photoperiods (Fig. 28-B), the majority of the rhythmic proteins present their maximum abundance level during

light hours. Rhythmic transcripts which encode for proteins that maintain their abundance levels constant over diel changes have been a paradox in chronobiology research. Protein degradation may account for differences between transcriptomic and proteomic data. This hypothesis suggests that rhythmic transcripts are needed to maintain a constant abundance level of the encoding protein when its degradation process is rhythmic (Lück et al., 2014). In that way, cells have to rhythmically supply new transcripts to balance protein synthesis/degradation ratio. In addition, the phase shifts dependent on photoperiods that are found in the transcriptomic data do not seem to be apparent in the proteome. Under both conditions, rhythmic proteins reach their maximum abundance around ZT0-ZT4 (Fig. 35-E, F). Indeed, no coincidences between the corresponding rhythmic protein and transcript abundance profiles are observed, indicating a lack of global correlation between proteome and transcriptome dynamics (Fig. 36-A, B).

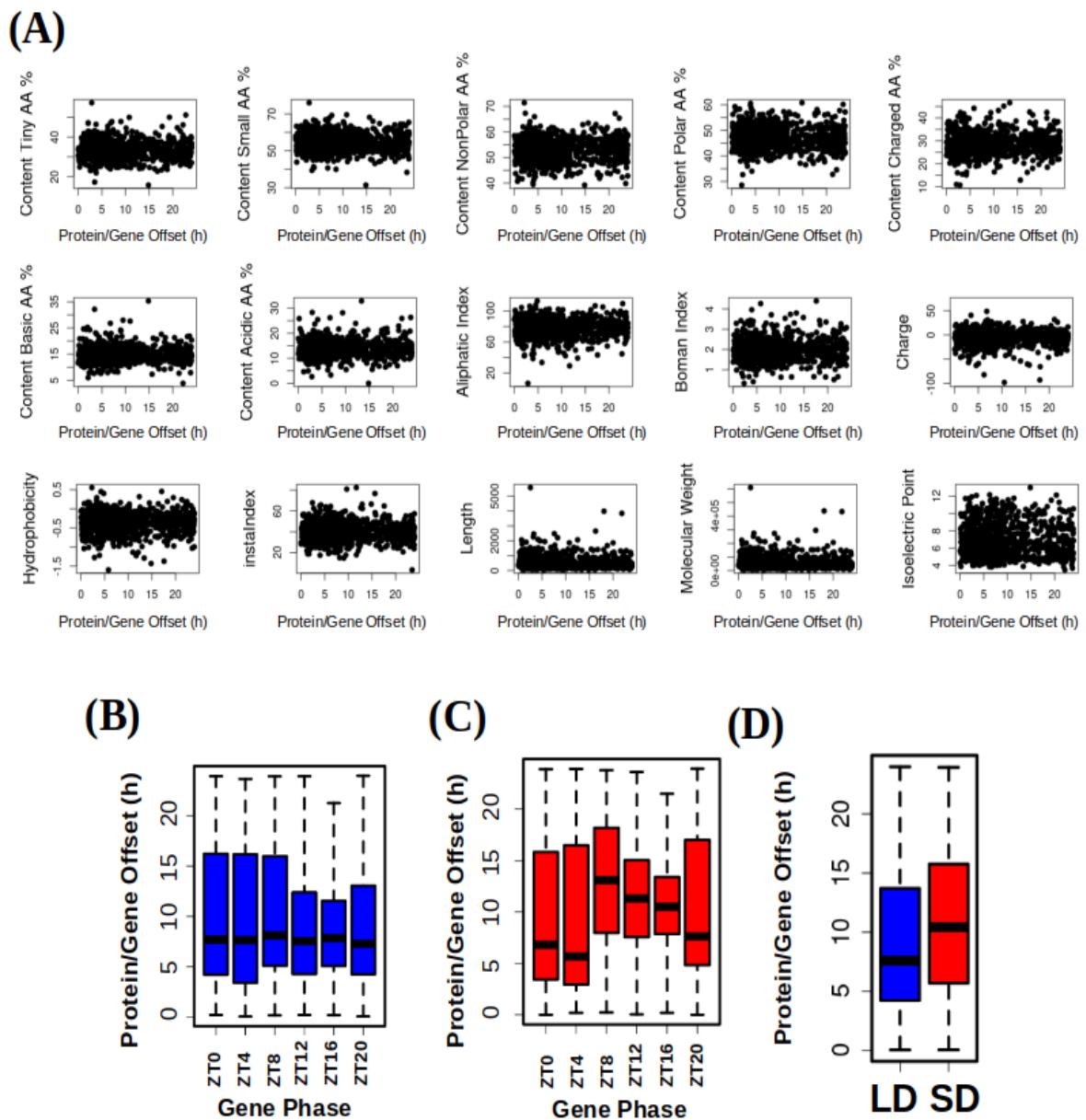


**Figure 36: Transcript-protein correlation analysis.** (A) Boxplots representing the global distribution of the correlations between protein abundance and gene expression profiles (white box) and shifted aligned profiles with coincident phases (grey); (B) Protein abundance (continuous blue line) and gene expression (dashed light blue line) profiles under LD condition for MA3 domain-containing translation Regulatory Factor (*ostta05g02330*, MRF); (C) Phase aligned protein abundance (dotted blue line) and gene expression (dashed light blue line) profiles.

These observations point to the decoupling of transcription and translation and the existence of a significant regulation over translation initiation. However, when protein and transcript abundance profiles are aligned omitting transcript-protein offsets, a positive high average correlation value of 0.84 is found (Fig. 36-A, C). This indicates that transcriptome and proteome dynamics are indeed linked but separated temporally. Similar temporal offsets between transcripts and proteins have been reported in other organisms, such as *Mus musculus* (Robles et al., 2014), and over 50 % of transcript phases were found to precede their corresponding protein phases by more than 6 hours.

### ***Correlation analysis of transcript-protein phase offsets***

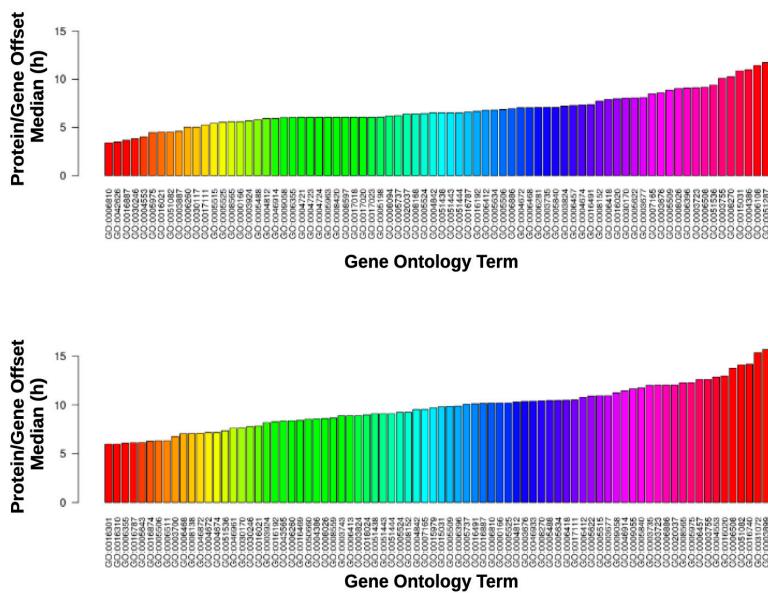
In order to obtain a deeper understanding of transcript-protein phase offsets, an extensive analysis of its correlation with different factors and protein properties was carried out. One of the possible hypothesis is that transcript-protein offsets depended on common indexes or properties computed from protein sequences such as amino acid composition, charge or hydrophobicity. However, no relation is detected between any protein index and transcript-protein offsets (Fig. 37-A). Another of the considered hypothesis is that transcript-protein offsets depend on the time when the maximum gene expression level is reached (gene phase). For example, a gene that reaches its maximum level of expression during the morning could take less time to translate than a gene which peak of expression takes place in the middle of the night. However, under LD condition, no significant differences are found when comparing transcript-protein offsets of different genes sets with specific phases (Fig. 37-B). Nonetheless, under SD condition transcript-protein offsets are significantly longer for transcripts peaking during the skotoperiod, ZT8, ZT12, ZT16 and ZT20, when compared to those genes with transcript phases during the photoperiod, ZT0 and ZT4 (Fig. 37-C). In addition, our analysis reveals that seasonal variations in photoperiod length affect transcript-protein offsets. In fact, global transcript-protein offsets are significantly longer under SD condition with respect to LD condition according to a p-value of  $1.2 \times 10^{-9}$  computed using Mann-Whitney-Wilcoxon test (Fig. 37-D). Finally, the key factor associated to differences in transcript-protein offsets identified in this analysis seems to be the specific biological process in which the corresponding protein is involved according to Gene Ontology annotation. In order to explore this, the average transcript-protein offset of each GO term is computed for both photoperiods of entrainment (Fig. 38-A).



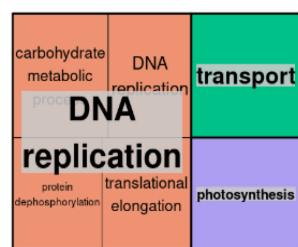
**Figure 37: Analysis of transcript-protein offsets correlation with protein properties, gene phase and photoperiod of entrainment.** (A) Scatter plots where each dot stands for a protein, x-coordinates represent transcript-protein offsets and y-coordinates different protein indexes or properties computed from their sequences such as amino acid composition, charge and hydrophobicity; (B-C) Boxplot representing transcript-protein offsets under LD (B) and SD (C) for different gene sets with specific phases or maximum expression time points; (D) Boxplot representing the offset in hours between protein abundance and gene expression phases under LD (blue) and SD (red) conditions.

Specifically, the most representative biological processes with short offsets are DNA replication and photosynthesis both under LD and SD conditions (Fig. 38-B). Whereas, processes involved in translation are the most representative ones exhibiting long transcript-protein offsets (Fig. 38-C). This finding could indicate the existence of a differential post-transcriptional regulation under diel cycles for each specific biological process. One of the processes with the shortest transcript-protein offsets is photosynthesis (Fig. 38-B). Some proteins involved in photosynthesis present a high turnover rate (Critchley & Russell, 1994; Nelson et al., 2014; Nelson & Millar, 2015), which can be preserved with short times between gene expression and translation. Genes involved in photosynthesis reach their maximum of expression during high irradiance hours, presenting a simple anticipation in time or phase shift between seasons (Fig. 32).

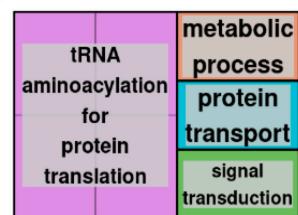
(A)



(B)



(C)



**Figure 38: Analysis of transcript-protein offsets correlation with their biological function.** (A) Median transcript-protein offset for gene sets annotated with the same Gene Ontology (GO) term under LD condition (top) and SD condition (bottom). (B-C) Treemaps summarizing the biological processes with shortest transcript-protein offsets: (B) and with longest transcript-protein offsets (C)

The activation of this set of genes is strategically placed at that time of the day taking in count their short transcript-protein offset, so no further anticipation to the high irradiance hours is needed. In addition, translation is identified as one of the biological processes with larger

transcript-protein offsets (Fig. 38-C). This process has been highlighted in the previous chapter for being one of the processes that are re-arranged in time over the diel gene activation programme between seasons (Fig. 32), which could be linked to the long transcript-protein offset that it presents.

These two observations suggest that *Ostreococcus tauri* severely reorganize its transcriptional temporal programme considering not only the current photoperiod but also the different transcript-proteins offsets that each biological function presents. In that way, every biological process is re-arranged in a certain time of the day ensuring that proteins will be ready in the exact time when they are needed, even compensating the long transcript-proteins offset.



# **Chapter 4.**

## *Diel and seasonal multi-omic integration with physiological data*



## **Chapter 4: diel and seasonal multiomic integration with physiological data**

In order to attain a comprehensive understanding of how some biological processes are regulated by seasonal variations in diel cycles, it is imperative to integrate the multiomic data generated with the dynamics of physiological processes. Such integration serves not only as a biological validation of computational omics analysis but also as a crucial component in unraveling the intricate organization present in biological systems, in general, and the response to diel cycles, in particular.

### **Cell Division Cycle (CDC) of *Ostreococcus tauri* under seasonal variations in diel cycles**

The cell division cycle (CDC) is a highly regulated sequence of processes that govern cell proliferation being conserved across eukaryotes. The impact of diel cycles on cell division has been studied in various phyla, including plants and microalgae such as *Chlamydomonas*, *Euglena* and *Gonyaulax* (Bruce, 1970; Edmunds & Laval-Martin, 2019; Fung-Uceda et al., 2018; Homma & Hastings, 1989) as well as in mice and humans (Fu et al., 2005; Matsuo et al., 2003). Nevertheless, the confirmation that circadian regulation controls cell division has been a topic of controversy in some organisms, such as the common microalgae model organism, *Chlamydomonas reinhardtii*. While some studies have concluded that the cell division cycle of this microalgae is subject to circadian regulation (Bruce, 1970), others have proposed that the observed periodicity is linked to cyclic changes in energy status, resulting from circadian regulation of photosynthesis (Spudich & Sager, 1980). Currently, it is widely accepted the existence of patent evidence regarding direct regulation exerted by the circadian clock over cell cycle progression in nearly all organisms, including photosynthetic ones. This regulation has been shown to persist under free-running conditions and present the ability to synchronize with different photoperiods (Roenneberg & Merrow, 2005), regardless of photosynthetic capacity. The cell division cycle, therefore, possesses a complex regulatory mechanism comprising robust circadian clock regulation as well as light-dependence in photosynthetic organisms, since light serves as their primary energy source (Goto & Johnson, 1995; Hagiwara et al., 2002; Moulager et al., 2007, 2010).

In line with the results presented in Chapter 2, the expression patterns of DNA replication genes remain rhythmic under constant light, while their rhythmicity is disrupted being strongly repressed under constant darkness (Fig. 23), similar to previously reported results (Roenneberg

& Merrow, 2005). This observation supports the notion that a light stimulus is necessary to sustain rhythmicity under free-running conditions, which agrees with the complex regulatory mechanism that the cell division cycle (CDC) present in photosynthetic organisms. The CDC of *Ostreococcus* follows the known phases of a simple binary fission. Initially, there is a Gap 1 (G1) phase that is dependent on the light-energy status, during which the cell undergoes growth and commitment occurs (Moulager et al., 2007). In cell division cycle studies, the term commitment refers to the moment when the cell, taking into consideration its energy status, decides whether is ready or not to continue with the progression of the cell division cycle. Once cells are committed, cell division is not impaired by darkness. Consequently, if commitment is achieved, G1 phase is followed by the DNA Synthesis or S phase where DNA replication takes place. The initiation of the S phase is typically timed several hours after sunrise (Moulager et al., 2007, 2010). After DNA replication is completed, cells enter the final Gap 2 and Mitotic (G2|M) phase, where they prepare for cell division (G2) and undergo mitosis (M). These two phases are often considered together since they are the shortest and most challenging to distinguish using common techniques. In all eukaryotes, the progression of cells through the cell division cycle is controlled by cyclins and cyclin-dependent kinases (CDKs). *Ostreococcus tauri* possesses an extremely limited set of cyclins and CDKs, with only a single copy of most genes (Robbens et al., 2005). Additionally, the genome of *Ostreococcus* contains a canonical cell division control protein 25 (CDC25), which is not present in plants (Khadaroo et al., 2004), and a plant-specific CDKB (Corellou et al., 2005).

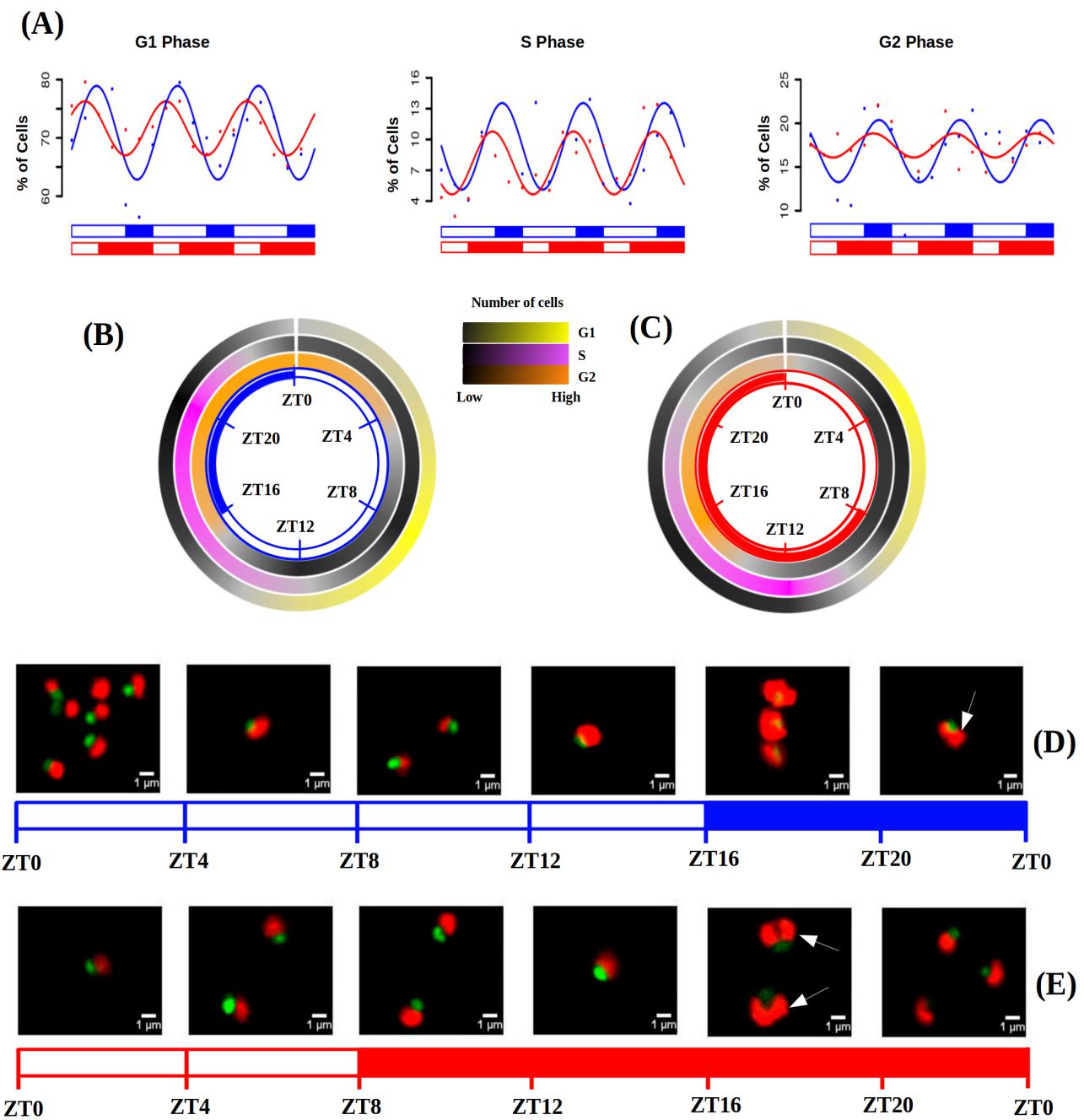
In the preceding chapters, genes and proteins involved in DNA replication (S phase) have been highlighted several times. Now, in order to validate these observations, an estimation of the distribution of cells in each phase over diel cycles is conducted. This integration could unveil the adaptation of the cell division cycle in *Ostreococcus* to different seasons and contributes to unraveling the molecular mechanisms of circadian regulation of cell division in microalgae.

### ***Temporal programme of cell division cycle under summer and winter photoperiod***

The phases of the cell division cycle have been determined by estimating the DNA content of cells through flow cytometry, and the division of chloroplasts have been observed using fluorescence microscopy, as outlined in Materials and Methods section. The rhythmicity analysis have been conducted using data obtained from three consecutive days under light-dark cycles.

During LD photoperiod, the G1, S, and G2|M phases exhibit noteworthy rhythmic profiles, with p-values of  $9.14 \times 10^{-6}$ ,  $4.45 \times 10^{-6}$  and  $1.53 \times 10^{-3}$ , respectively (Appendix 4). Conversely, under the SD photoperiod, only the G1 and S phases demonstrate significant rhythmicity, with p-values of  $9.07 \times 10^{-6}$  and  $8.69 \times 10^{-4}$ , respectively (Appendix 5). Consistent with the findings from transcriptomic and proteomic analyses, a decline in synchronization, characterized by a reduction in amplitude, is evident under the short photoperiod (Fig. 39-A). The reduction in amplitude is so substantial in the G2|M phase that the RAIN package used in the rhythmicity analysis fails to detect a rhythmic profile. These observations suggest that the cell division cycle of individual cells within the culture is more synchronized during LD, which would align with the higher synchronization under this condition that has been also found at the transcriptomic level and presented in Chapter 2 (Fig. 28-A, C). Few daylight hours also appear to have a direct effect on cell cycle progression, since a significant reduction of approximately 24 % in the number of committed cells is observed under short photoperiod. Additionally, significant backward shifts of approximately 4 hours were observed in the cell cycle phases under short photoperiod (Fig. 39-A). These observations are consistent with the observed backward shifts in the time points when transcript abundances reach their maximum levels in response to photoperiod shortening (Fig. 28-C).

The mean percentage of cells involved in G1, S and G2|M phases was calculated for each time point, allowing the characterization of the temporal progression of the cell division cycle in *Ostreococcus* under different photoperiods (Fig. 39-B, C). In LD entrained cultures, G1 phase predominantly occurs during the light hours, with the maximum percentage of cells in this phase coinciding with the maximum irradiance hours (around ZT8). After commitment, the percentage of cells in G1 phase gradually decreases, while the percentage of cells in the S phase increases. The highest number of cells in the S phase is reached around sunset and the first part of the night (ZT16-ZT20). Subsequently, the percentage of cells in the G2|M phase gradually rises as DNA duplication is successfully accomplished. The transition from the G2|M to G1 phase takes place during the first part of the morning (ZT4). This indicates that cell division in *Ostreococcus* mainly takes place after sunrise in summer long photoperiods (Fig. 39-B).



**Figure 39: Cell division cycle (CDC) of *Ostreococcus tauri* under long and short photoperiod.** (A) Percentage of cells in G1, S and G2|M phases during the three days of sampling. Points correspond to real data and lines represent waves approximations made by Circacompare during the rhythmicity analysis; (B-C) Circular heatmap representing mean percentage of cells in G1, S and G2|M phases during long and short photoperiod, respectively; (D-E) Photographs of cells under the fluorescence microscope. Each photograph corresponds to a different time point of long and short photoperiod, respectively. Nucleus dyed in green, chloroplast in red. White arrows point cells that possess two chloroplasts.

In SD entrained cultures, the G1 phase coincides with maximum irradiance hours (ZT4) and the S phase occurs 4 hours after sunset (around ZT8), consistent with the observations made in LD entrained cultures. However, the G2|M phase exclusively occurs by night in SD. As soon as the sun rises, cell division is completed, allowing cells to undergo growth in the morning. These findings suggest that the cell division cycle is strongly influenced by the circadian clock and can anticipate cyclic changes, such as the reduced light period during winter photoperiods. To ensure this anticipation, the circadian clock in *Ostreococcus* ensures that all cells enter the G1 phase precisely at sunrise, maximizing the utilization of available daylight hours (Fig. 39-C).

This anticipation is also observed in chloroplast division. In LD entrained cultures, chloroplast duplication is achieved during the latter part of the night (ZT20) (Fig. 39-D), while it takes place at ZT16 in SD entrained cultures. Before sunrise, a substantial number of cells already possess only one chloroplast during the short photoperiod (Fig. 39-E).

### ***Integration of CDC programme with transcriptomic and proteomic data.***

*Ostreococcus tauri* annotated genes involved in cell division cycle were organized in three different groups in order to mark in which phase of the cell cycle they are present, according to the cell cycle model in plants (Carneiro et al., 2021; Gutierrez, 2009).

Cyclin D and CDKA are present specifically during the first part of the G1 phase, facilitating the progression of the cell cycle to the S phase. Transcription factors like E2F and Dp as well as other proteins (Rb, cell division control protein 6 and ORCs) are also involved in G1 phase, playing a crucial role in the activation of genes associated with the S phase (Table 7). Cyclin A (CYCA) is involved in the orchestration of the DNA replication process during the S phase. The expression of polymerases and replication-related proteins (such as MCM complexes, replication factors, PCNA, primases, helicases, ligases, and others) are achieved during this phase (Table 8). Cyclin B is a key component in the molecular machinery that controls entry and exit from mitosis in the cell cycle. Therefore, this protein is present during the G2|M phase along with subunits of the anaphase-promoting complex (APC) and cell division control proteins (CDC20 and CDC25) (Table 9) (Carneiro et al., 2021; Corellou et al., 2005; Gutierrez, 2009; Moulager et al., 2007, 2010; Robbens et al., 2005).

**Table 7: Genes and their encoded proteins that are typically present during the G1 phase of the cell division cycle in Ostreococcus.**

Gene Ids	Protein names
ostta18g01570	CYCD
ostta04g00110	CDKA
ostta04g01050	CDC6
ostta16g02100	Rb
ostta02g01950	E2F
ostta10g01190	Dp
ostta08g02940	DEL
ostta04g05220	ORC1
ostta15g02820	ORC2
ostta16g01070	ORC3
ostta11g01820	ORC4
ostta03g02760	ORC5

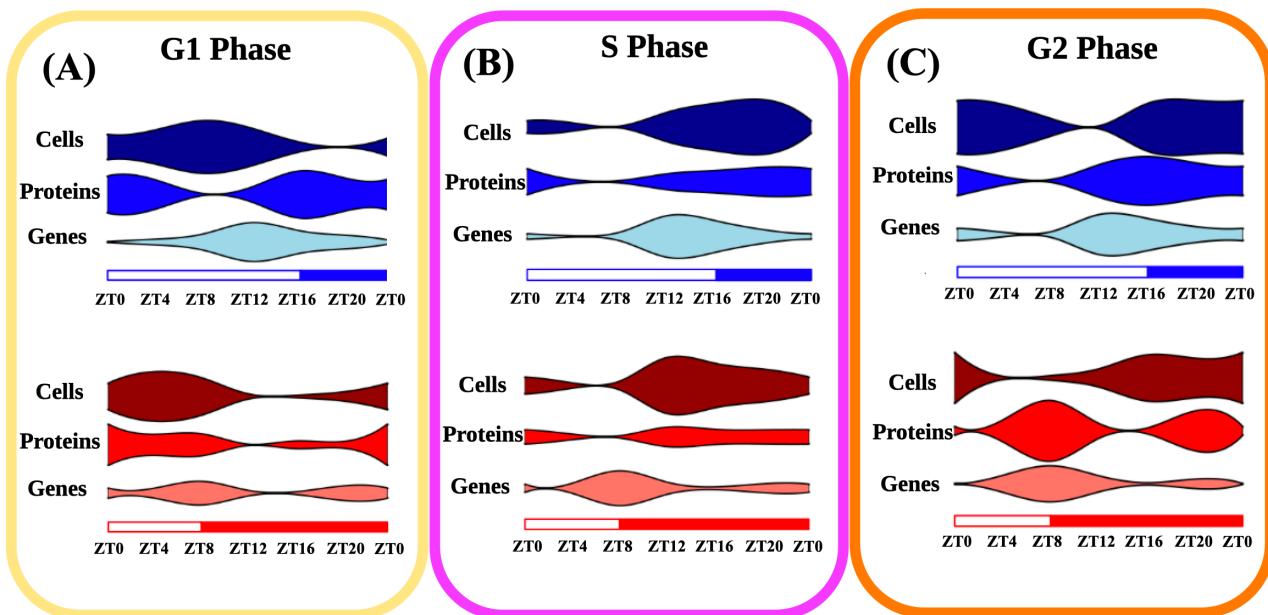
**Table 8: Genes and their encoded proteins that are typically present during the G2|M phase of the cell division cycle in Ostreococcus.**

Gene Ids	Protein names
ostta01g06150	CYCB
ostta15g00670	CDKB
ostta06g02700	APC1
ostta11g00730	APC2
ostta06g00360	APC3
ostta02g03470	APC4
ostta06g04290	APC5
ostta01g01000	APC6
ostta07g02520	APC8
ostta10g02910	APC10
ostta11g03040	APC11
ostta04g04580	Cdc20
ostta02g03080	Cdc25
ostta13g02370	CDH1
ostta08g02230	FTSZ1
ostta07g01610	FTSZ2

**Table 9: Genes and their encoded proteins that are typically present during the S phase of the cell division cycle in Ostreococcus.**

Gene Ids	Protein names
ostta02g00150	CYCA
ostta11g00910	MCM2
ostta02g00690	MCM3
ostta14g01050	MCM4
ostta11g01760	MCM4-2
ostta04g00450	MCM5
ostta01g02580	MCM6
ostta02g02010	MCM7
ostta12g01020	MCM8
ostta14g01690	GINS1
ostta02g00490	PSF2
ostta16g02410	GINS3
ostta04g00390	SLD5
ostta02g00500	RFC1
ostta02g00555	RFC2
ostta09g00990	RFC4
ostta07g02710	RFA
ostta02g03430	DNA polymerase II
ostta05g00150	DNA polymerase III γ
ostta08g00710	DNA polymerase α
ostta11g01400	DNA polymerase α
ostta11g01400	DNA polymerase α
ostta07g02610	DNA polymerase δ
ostta08g03680	DNA polymerase α/ε
ostta12g00110	DNA polymerase α/ε
ostta11g01840	DNA polymerase ε
ostta06g02890	PCNA
ostta08g01110	DNA primase
ostta11g00940	DNA primase
ostta13g02040	DNA primase
ostta08g02290	XPG
ostta10g00340	DNA helicase
ostta10g00640	DNA ligase
ostta13g01140	RPA replication protein A
ostta10g00670	Wee1
ostta07g01870	Cks

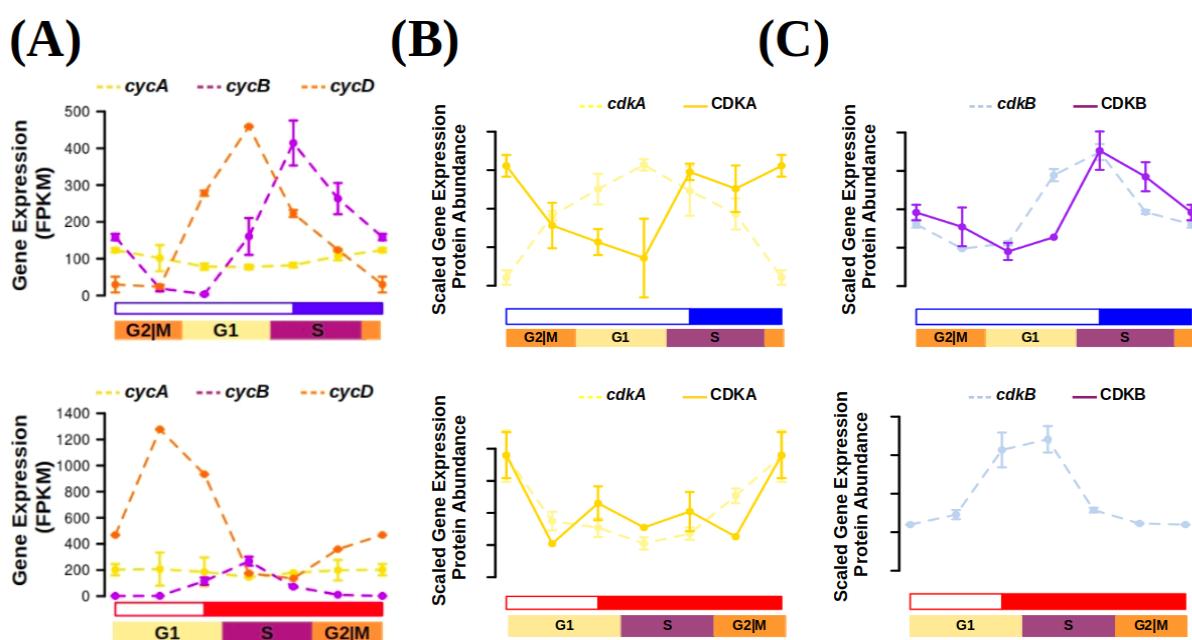
Gene expression and protein abundance profiles are integrated with the percentage of cells in each phase of the cell cycle in which the corresponding genes and proteins are present (Fig. 40). In addition to the globally observed transcript-protein offset, an overall offset is also identified between protein abundance and the execution of their physiological role. Notably, this offset appear to be more pronounced in certain phases such as G1 or G2|M (Fig. 40-A, C). However, during the S phase, once the proteins are available, the corresponding biological processes are promptly executed, resulting in a shorter offset between protein and physiological execution (Fig. 40-B).



**Figure 40: Integration of gene expression, protein abundance and cell population profiles for each phase of the cell cycle.** Violin plots represent the three molecular levels studied: “Genes” for transcriptomic data, “Proteins” for proteomic data and “Cells” for DNA content estimation by flow cytometry. (A) Involves G1 phase related data, (B) S phase and (C ) G2 phase.

The transcriptomic and proteomic data for cyclins and CDKs in this work (Fig. 41) align with previous transcriptomic data of *Ostreococcus* under neutral day conditions (Corellou et al., 2005; Monnier et al., 2010). In SD and LD entrained cultures, the transcription of the CYCD, that triggers the start of the S phase, occurs during the G1 phase and it is the first cyclin to be activated after sunrise. The expression of Cyclin D (during G1 phase) is followed by the expression of Cyclin B (during the S phase) (Fig 41-A), which aligns with the results obtained in neutral day condition. The accumulation and degradation of Cyclin B is critical to ensuring proper progression through mitosis, and its strong rhythmic expression profile has been also described in previous work in *Ostreococcus* (Corellou et al., 2005; Monnier et al.,

2010). Cyclin proteins were not detected in our proteomic analysis, but CDKA and CDKB protein abundance profiles are detected. The expression of CDKA encoding gene reaches its maximum level during G1, preceding an increase in CDKA protein abundance during the latter part of the G1 phase (Fig. 41-B). This increase, in conjunction with Cyclin D, facilitates the progression of the cell cycle into the S phase. During the S phase, the abundance of CDKB protein reaches its maximum level (Fig. 41-C), coinciding with Cyclin B transcript levels which, after being translated, will control the entrance and exit from the mitosis phase (Gutierrez, 2009). The transcript levels of Cyclin A are relatively low under both photoperiods (Fig. 41-A).



**Figure 41: Transcript and protein abundance profiles of the main cell cycle proteins under LD and SD in *Ostreococcus tauri*.** (A) Expression level of cyclins A (yellow), B (purple) and D (orange) genes are represented with discontinued lines under long (blue) and short (red) photoperiods; (B-C) Transcript (discontinued line) and protein (solid line) abundances of CDKA are represented in yellow (B) and CDKB in purple (C).

## Diel and seasonal rhythm of photosynthesis in *Ostreococcus tauri*

Photosynthesis constitutes a fundamental process for plants, wherein oxygen ( $O_2$ ), adenosine triphosphate (ATP), and reduced nicotinamide adenine dinucleotide phosphate (NADPH) are generated from water ( $H_2O$ ) and light. The generated oxygen is released into the atmosphere, while ATP serves as an essential energy source for cellular processes. Moreover, NADPH

plays a crucial role as a reducing agent, facilitating assimilatory processes such as the Calvin cycle, which mediates the fixation of atmospheric carbon dioxide ( $\text{CO}_2$ ) into organic carbon compounds. Within the genome of *Ostreococcus tauri*, all the indispensable proteins implicated in the electron transport chain of photosynthesis and carbon fixation are present. However, the number of copies in *Ostreococcus* is lower when compared with plants and other microalgae. Notably, the composition of light-harvesting complexes in *Ostreococcus* exhibits distinct characteristics. While light-harvesting complex proteins associated with photosystem I (LHCI) are present, LHCII are lacking. Instead, specific chlorophyll-binding proteins unique in prasinophytes are identified in *Ostreococcus* (Blanc-Mathieu et al., 2014; Derelle et al., 2006). This observation suggests the presence of LHCI within the green lineage from an evolutionary stage prior to *Ostreococcus* ancestors (Six et al., 2005).

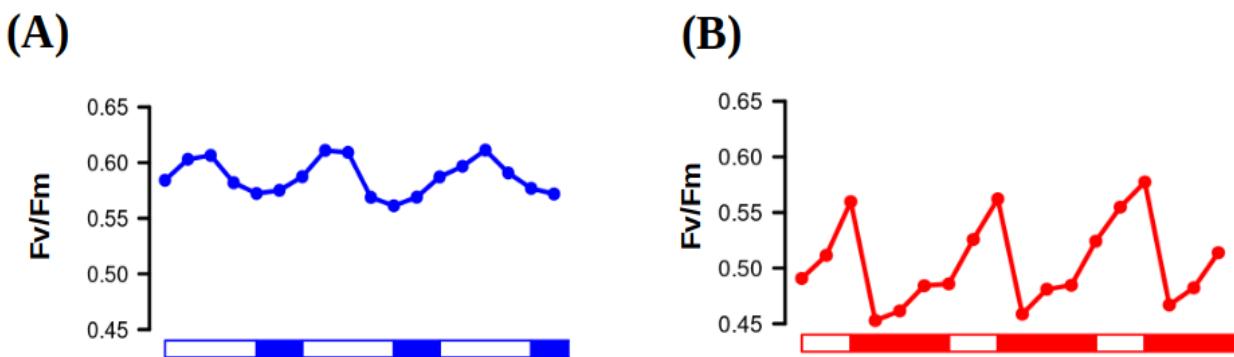
Understanding how photosynthesis adapts to diel and seasonal cycles can contribute to develop systems enhancing microalgae productivity, which is a relevant topic in microalgae biotechnology. The circadian regulation of photosynthesis was initially documented in marine algae, with the observation of circadian oscillations in oxygen production (Dodd et al., 2014; Sweeney & Haxo, 1961). After that discovery, circadian oscillations in various physiological phenomena associated with photosynthesis, including chloroplast ATP concentration, electron transport rate, starch content, and photosynthetic efficiency, have been observed in microalgae (Mackenzie & Morse, 2011; Ral et al., 2006; Sorek et al., 2013; Sweeney & Haxo, 1961). Furthermore, circadian oscillations in photosynthesis have also been documented in agriculturally relevant plant species (Feugier & Satake, 2013; Lonergan, 1981; Tucker et al., 2004). Nowadays, the application of omics techniques has enabled the elucidation of 24 hour rhythmic oscillations in genes involved in photosynthesis and carbon fixation in both microalgae and plants (Ferrari et al., 2019).

This work describes the maintenance of rhythmic expression profiles of genes involved in photosynthesis and carbon assimilation under different photoperiods and free-running conditions (constant light and constant darkness) in *Ostreococcus tauri*. The genes involved in photosynthesis have been described as *bona fide* circadian genes in Chapter 2 due to their sustained rhythmic expression profiles during both long and short photoperiods, as well as under free-running conditions (Fig. 25-C,B). Additionally, in Chapter 3, it has been observed that photosynthesis exhibited one of the shortest offsets between gene expression and translation in *Ostreococcus tauri* (Fig. 38-B).

These findings, derived from multi-omics analyses, are validated with photosynthetic efficiency measurements in LD and SD entrained cultures. This integration sheds light on the adaptation of photosynthesis, as well as other interconnected processes such as carbon fixation and starch metabolism in *Ostreococcus*, to accommodate varying seasonal conditions. Furthermore, it provides valuable insights about the conserved mechanisms governing circadian regulation and ultimately influencing photosynthetic activity within the green lineage.

### **Rhythmic oscillations of photosynthetic efficiency under summer and winter photoperiods**

Photosynthetic efficiency has been calculated using Pulse-Amplitude-Modulation (PAM) by estimating maximum quantum efficiency ( $F_v/F_m$ ), a widely accepted metric to assess the overall integrity and efficiency of Photosystem II (PSII). To determine rhythmicity in photosynthesis,  $F_v/F_m$  measurements are obtained from samples collected under 3 consecutive light-dark cycles (Appendix 6).



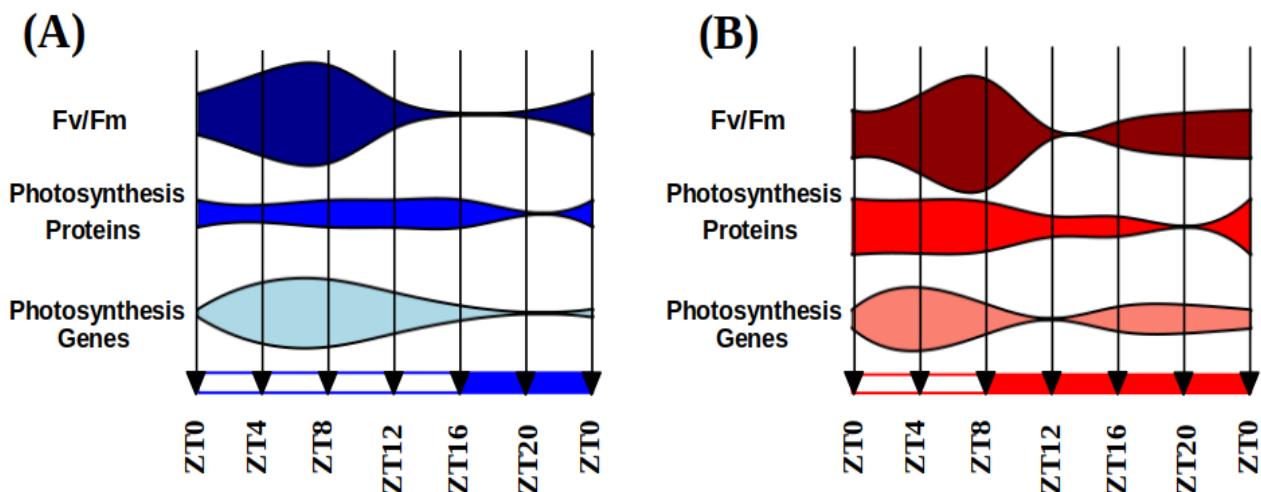
**Figure 42. Photosynthetic efficiency rhythmic oscillations.**  $F_v/F_m$  oscillating values used as an estimation of PSII performance and, thus, photosynthetic efficiency under three consecutive days of long (A) and short (B) photoperiods.

In LD entrained cultures, the  $F_v/F_m$  parameter exhibits a rhythmic profile with a 24 h period, displaying a significant p-value of  $7.07 \times 10^{-6}$  (Fig. 42-A). The maximum  $F_v/F_m$  value consistently occurs during the maximum irradiance hours, around ZT8. This observation indicates that the photosystems operate at their highest efficiency during that specific time of the day, consequently enhancing photosynthetic efficiency. In contrast, the  $F_v/F_m$  profile in SD entrained cultures displays a rhythmic pattern with two peaks every 24 h (Fig. 42-B). Both peaks demonstrate significant periodicity every 12 hours, supported by a p-value of  $1.5 \times 10^{-2}$  (Appendix 6). One of the peaks aligns with the  $F_v/F_m$  rhythmic profile observed in

LD, occurring at ZT8 as well. Whereas, the second peak represents a smaller increment in the  $F_v/F_M$  value, occurring more than 8 hours prior to sunrise. This finding suggests that the photosynthetic machinery in *Ostreococcus tauri* exhibits a circadian response, preparing itself in advance during the night to anticipate the limited daylight hours of the winter photoperiod. These outcomes underscore the ability of *Ostreococcus tauri* to anticipate cyclic variations in photoperiod, which is a widely accepted signal of robust circadian regulation (Roenneberg & Merrow, 2005).

### ***Integration of photosynthesis efficiency rhythmic profile with multi-omics data***

Photosynthesis entails an electron transport chain consisting of four major protein complexes: Photosystem II (PSII), cytochrome b6f, Photosystem I (PSI) and Ferredoxin NADP<sup>+</sup> oxidoreductase (FNR). The photosynthetic efficiency ( $F_v/F_M$ ) rhythmic profiles previously presented are now integrated with the proteomics and transcriptomics data associated with the PSII (Fig. 43).

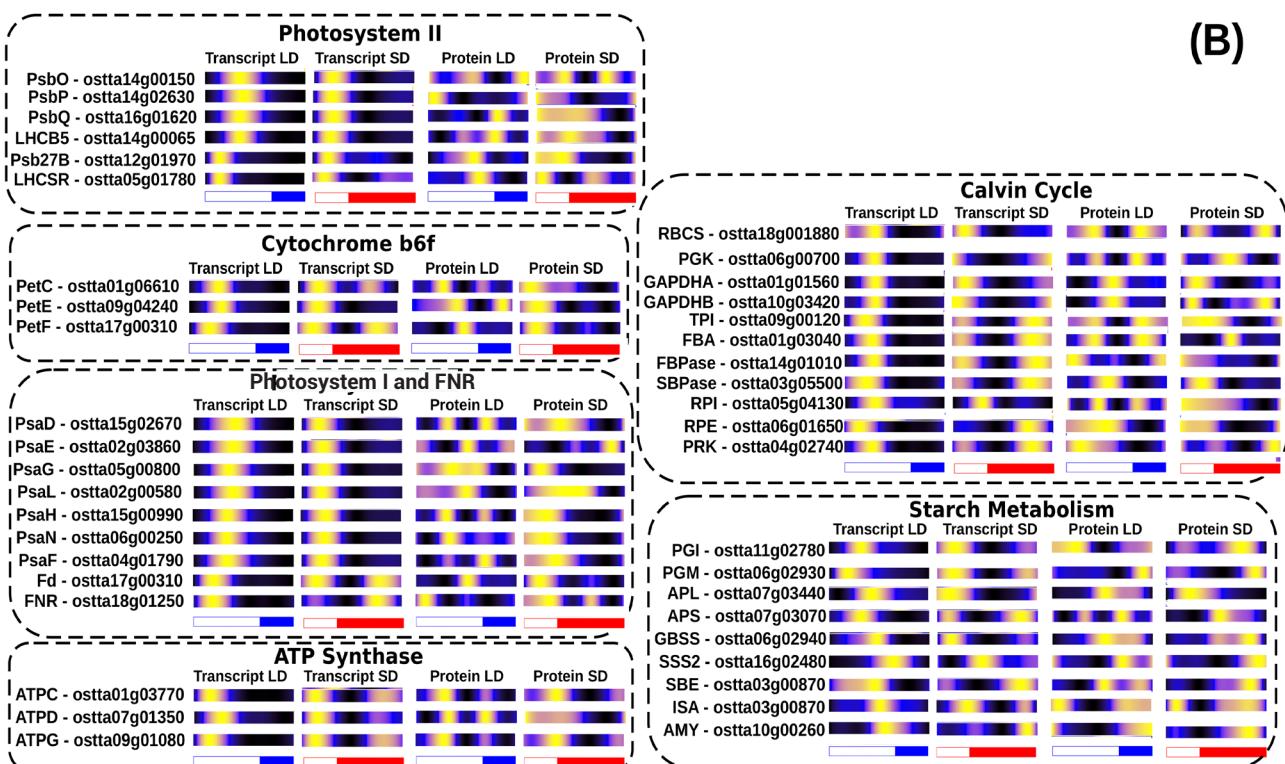
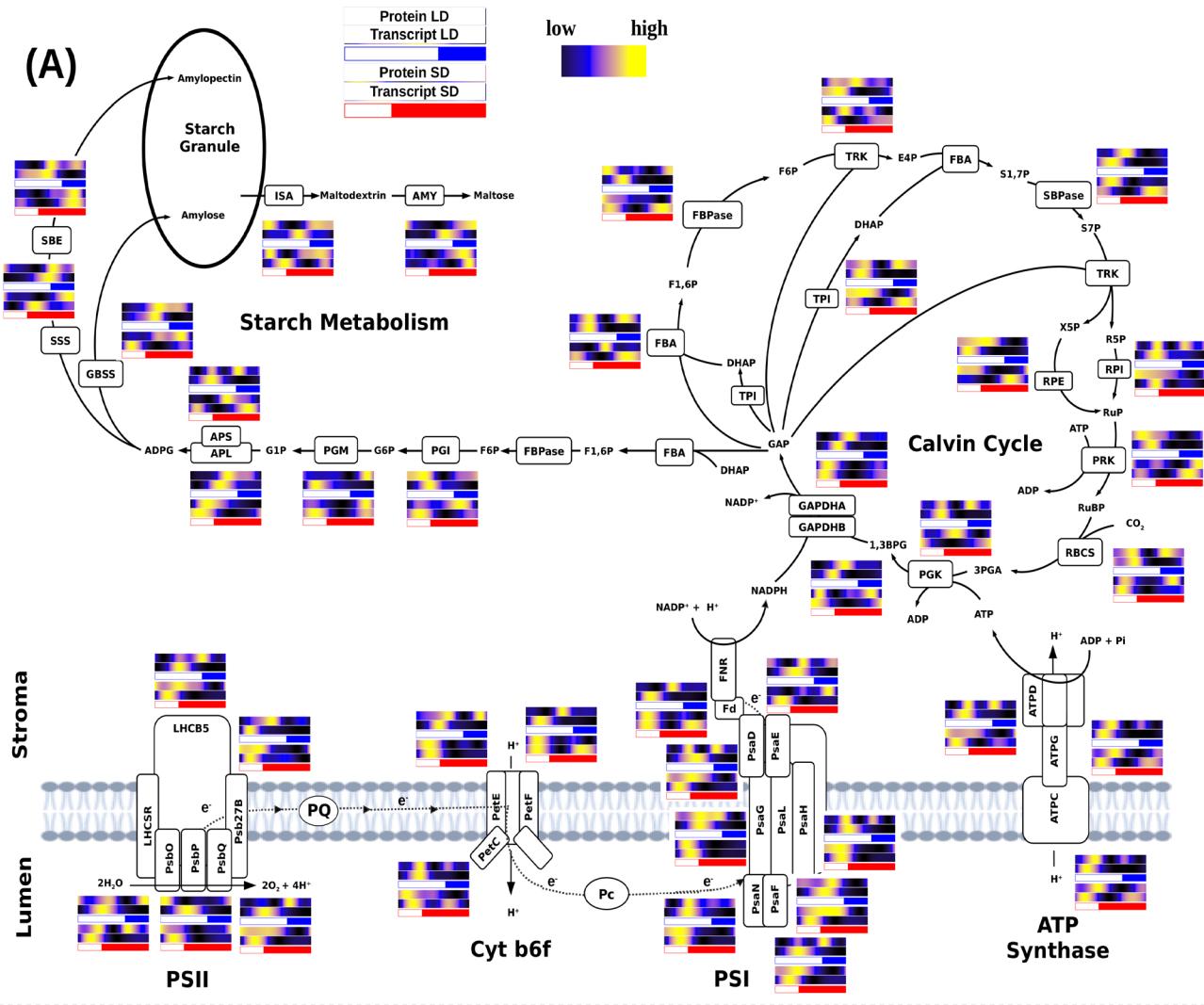


**Figure 43: Integration of  $F_v/F_m$  oscillations with multiomic data.**  $F_v/F_m$  measurements are integrated with multiomic data from proteins and genes related with photosynthetic efficiency under long (A) and short (B) photoperiod. Proteins and genes included in PSII are listed in Fig. 44-B.

In LD entrained cultures, no temporal offsets are observed between the time points of maximum transcript and protein abundances and the highest  $F_v/F_M$  values (Fig. 43-A). These observations suggest that, during summer photoperiod, transcripts are promptly translated as soon as genes are transcribed, leading to an increase in photosynthetic efficiency through

de-novo protein synthesis. Conversely, during winter photoperiod there is a longer transcript-protein temporal offset, as it has been discussed in Chapter 3 ([Fig. 37-D](#)), and the maximum photosynthetic efficiency occurs a few hours after the maximum gene expression level ([Fig. 43-B](#)). Rhythmic profiles with a 12-hour period (two peaks every 24 hours) are observed not only in  $F_v/F_M$  but also in the gene expression and protein abundance profiles of PSII proteins in SD entrained cultures ([Fig. 43-B](#)).

Besides PSII, the other three complexes are also crucial in the execution of photosynthesis. These complexes facilitate the transfer of electrons, generated through the cleavage of the water molecule to their final acceptor, NADP, generating the NADPH required for biosynthetic reactions. The energy derived from this transport enables the pumping of protons into the chloroplast lumen, ultimately leading to ATP synthesis as they return to the stroma via ATP synthase. NADPH and ATP are needed to fix  $\text{CO}_2$  and generate carbon compounds within the Calvin cycle, which can be stored as starch reserves ([Fig. 44-A](#)). In the data integrated in this work, 12 h period rhythmic profiles are found in transcripts and proteins abundance profiles from all these complexes. Specifically, profiles with two peaks are found in genes such as Protein Electron Transfer C (PetC, ostta01g06610), Ferredoxin (Fd, ostta17g00310) and ATPase delta subunit (ATPD, ostta07g01350) coding for key components of Cytochrome b6f, Photosystem I and ATP Synthase ([Fig. 44-B](#)). Also, numerous genes involved in Calvin cycle and starch metabolism exhibit these expression profiles as well ([Fig. 44-B](#)). In all these genes, one expression peak occurs during the early morning, enhancing photosynthetic efficiency during the hours of maximum irradiance. The second expression peak takes place during the night, inducing the anticipation of the photosynthetic machinery before sunrise. These findings suggest that the response of the photosynthetic machinery in anticipation to the photoperiod is transcriptionally regulated in *Ostreococcus tauri* and it affects all processes related with photosynthesis, including not only the electron transport chain but also the Calvin cycle. In fact, in plants, the circadian clock participates in coordinating different physiological processes like photophosphorilation, carbon fixation and starch biosynthesis (de los Reyes et al., 2017; Farré & Weise, 2012; Graf et al., 2010).



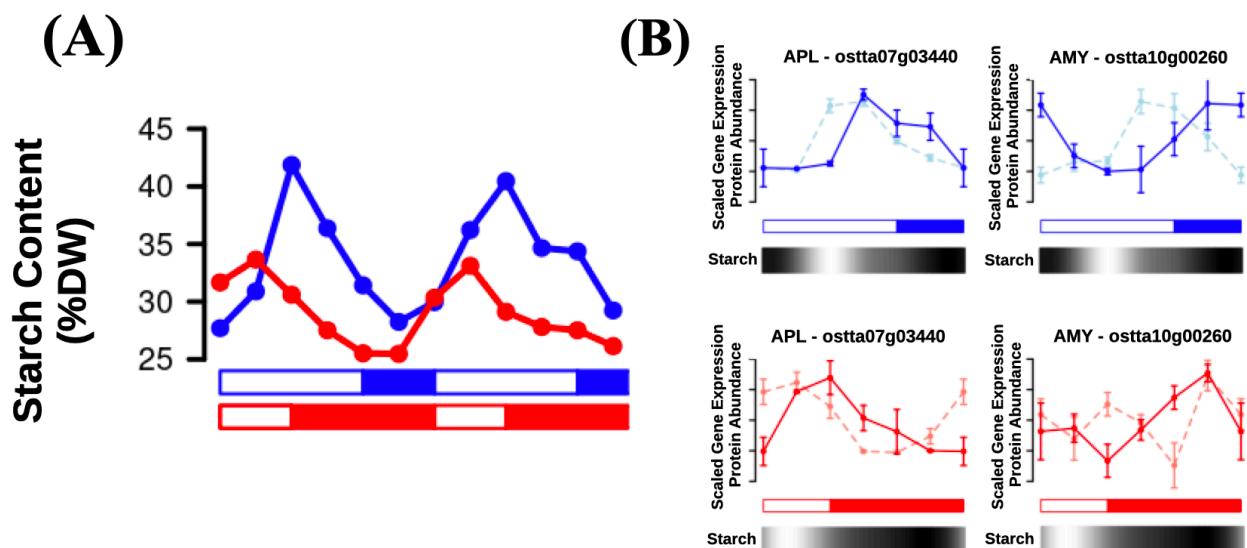
**Figure 43: Integration of multiomic data of enzymes and proteins involved in photosynthesis, Calvin cycle and starch biosynthesis.** (A) Schematic representation of the mentioned metabolic pathways. For each enzyme its corresponding protein and transcript abundances under LD, top in blue, and SD, bottom in red, are represented using heatmaps where black represents low, blue medium and yellow high abundances respectively. (B) Organized list of genes IDs associated with Photosystem II, cytochrome b6f, Photosystem I and FNR, ATP synthase, Calvin cycle and starch metabolism. Heatmaps represent, from left to right, transcript abundance under LD and SD, protein abundance under LD and SD.

However, most genes involved in photosynthetic electron transport chain and Calvin cycle are consistently expressed early in the morning, under both long and short photoperiods (Fig. 44-B). The expression of those genes seems to be unaffected by the photoperiod, since it has been also observed in *Ostreococcus* under a neutral photoperiod, consisting in 12 h of light and 12 h of dark cycles (Monnier et al., 2010). In fact, genes coding for key components of the Calvin Cycle such as Glyceraldehyde-3-phosphate dehydrogenase A (GAPDHA ostta01g01560) and Fructose-1,6-bisphosphate aldolase (FBA, ostta01g03040) are examples of *bona fide* circadian genes exhibiting rhythmicity under both long and short photoperiods, as well as free-running conditions of constant light and constant darkness.

In addition, protein abundance rhythmic profiles with 12 h period are observed in LD entrained cultures along numerous genes listed in Fig. 43-B. However, the secondary peaks observed in these proteins are not induced by an increment in the transcription level of their corresponding gene (Fig. 43-B), which is in agreement with previously published data in plants where these transcripts show a 24 h period (Pilgrim & McClung, 1993). Multi-omics integration of the results presented in this thesis shows that this phenomenon is present in all Calvin cycle enzymes, and some proteins from the photosynthetic electron transport chain (Fig. 43-B). These 12 h period profiles are not transcriptionally regulated as the ones described in short photoperiod, since its encoding transcripts describe only one peak of expression per day (Fig. 43-B). Instead, a post-transcription regulatory mechanism is probably causing this 12 h period profile in proteins involved in photosynthesis during long photoperiod in *Ostreococcus*.

## Integration of starch content diel oscillations with multiomic data

The accumulation and degradation of starch have been described to be circadian regulated, since periodic oscillations of its content are observed in photosynthetic organisms as *Chlamydomonas* and *Arabidopsis*, as well as a rhythmic gene expression profile of the enzymes involved in the process (Flis et al., 2019; Geigenberger, 2011; Kötting et al., 2010; Ral et al., 2006; Smith et al., 2004; Sulpice et al., 2014). *Chlamydomonas* reaches its maximum starch content few hours after sunset (Ral et al., 2006), while *Arabidopsis* plants reaches it exactly at sunset (Feugier & Satake, 2013; Kötting et al., 2010). In both organisms, starch accumulate during the light hours, until amylases (AMY) are activated and its degradation starts. Starch content must be considered as the result of a controlled balance between its degradation, by AMY, and its synthesis, by ADP-glucose pyrophosphorylase with its small and large subunit genes (APS and APL).



**Figure 45: Starch Content compared with transcript and protein abundance profiles of key enzymes.** (A) Starch content (%DW) of *Ostreococcus tauri* under LD (blue) and SD (red) conditions (B) Transcript (lighter color) and protein (darker color) abundance profiles for ADP-glucose pyrophosphorylase (APL, left) and amylase (AMY, right) under LD condition (top, blue) and SD condition (bottom, red). Heatmaps are incorporated below to represent the changes in starch content during the 24 h cycle. Black represents low content and white high content.

Starch content in *Ostreococcus tauri* under diel cycles is also rhythmic with a 24 h period (Fig. 45-A) with a p-value < 0.05 (Appendix 7), which align with the periodic oscillations described in *Arabidopsis* and *Chlamydomonas*. However, *Ostreococcus* do not reach its maximum starch content around sunset as they do. Instead, the maximum starch content aligns with the high

irradiance hours and it decreases gradually after that point, under both photoperiods (Fig. 45-A, Appendix 7). These results suggest that there is not a conserved starch content temporal programme throughout the green lineage, since its degradation starts at different times of the day depending on the organism. However, which seem to be conserved are the differences between photoperiods. In *Ostreococcus*, starch degradation during winter photoperiod seems to be executed more slowly than during summer photoperiod (Appendix 7), as it has been also described in plants (Feugier & Satake, 2013; Geigenberger, 2011; Kötting et al., 2010; Sulpice et al., 2014). The starch temporal programme observed in *Ostreococcus tauri* (Fig. 45-A) aligns with the transcriptomics and proteomics data generated and analyzed in this work. Under both long and short photoperiods, starch content reaches its maximum at midday, despite the abundance of enzymes involved in starch biosynthesis, like APL (ostta07g03440), peaking several hours later, toward end of the day (Fig. 45-B). The halt in starch increase and the subsequent decrease in its content could be attributed instead to the activation of the genes encoding enzymes involved in starch degradation, like AMY (ostta10g00260). The consequently increase in abundance of the corresponding proteins during the second half of the day under both photoperiod is, thus, coincident with the decrement in starch content (Fig. 45-B).

In Chapter 3, transcript-protein offsets have been hypothesized to be dependent on the photoperiod of entrainment and the biological process where those proteins are involved (Fig. 37-38). Our results show how temporal offsets of genes involved in the same biological process increase under short day condition, indicating a common regulation under seasonal cycles. In starch metabolism, offsets of APL are approximately 4 hours under both photoperiods, while those of AMY differ, increasing by 8 hours in LD and 12 hours in SD (Fig. 45-B). This support previous results in plants, that suggest that synthesis and degradation of starch are complex processes regulated by different mechanisms (Geigenberger, 2011; Hartman et al., 2023; Kötting et al., 2010). In addition, both APL and AMY protein abundance profiles are coincident with their gene expression profile except for the described offset (Fig. 45-B), showing that at least transcriptional regulation is one of the mechanisms that may contribute to starch synthesis-degradation balance in *Ostreococcus tauri* as it has been observed in *Arabidopsis* and other plants (Finegan et al., 2022; Geigenberger, 2011; Kötting et al., 2010; Sorokina et al., 2011).

## **Other metabolic pathways of *Ostreococcus tauri* showing periodic oscillations under diel and seasonal cycles**

### ***Carotenoids biosynthesis in *Ostreococcus tauri* under diel and seasonal cycles***

Carotenoids are a group of isoprenoid pigments that are widely distributed among various organisms, including microalgae and plants. Some of these pigments are associated with light-harvesting complexes and perform a crucial role in photosynthesis by efficiently absorbing light energy and transferring it to reaction centers. In addition, carotenoids exhibit antioxidant properties, safeguarding the organism against potential harm induced by excessive light exposure and environmental stress (García-Plazaola et al., 2017; Sun et al., 2022). The expression of genes associated with carotenoid biosynthesis is intricately regulated by the circadian clock in both plants and algae. This circadian regulation ensures that the production and accumulation of carotenoids aligns in time with the physiological demands and environmental conditions, maximizing their effectiveness in light absorption, energy transfer, and antioxidant protection. (Covington et al., 2008; García-Plazaola et al., 2017; Pan et al., 2009; Sun et al., 2010; Zhang et al., 2022). Besides their role in photosynthesis, carotenoids have considerable nutritional value for humans. Certain carotenoids, including β-carotene, possess the capacity to be converted into vitamin A, a vital nutrient essential for maintaining vision and bolstering the immune system. Furthermore, specific carotenoids, such as astaxanthin, have exhibited promising potential in promoting health, as they have been associated with reducing the risk of certain cancers and cardiovascular diseases (Eggersdorfer & Wyss, 2018).

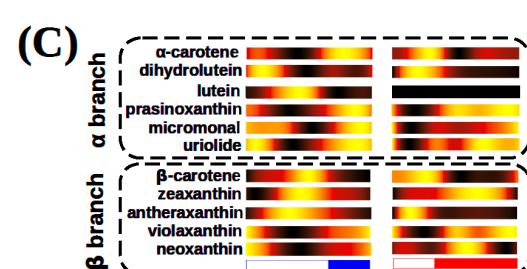
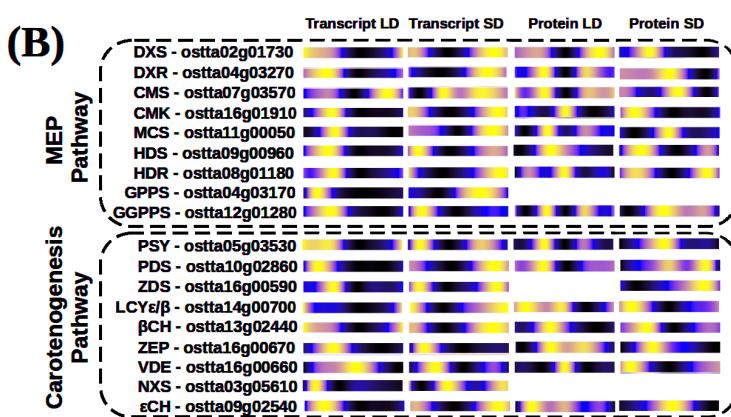
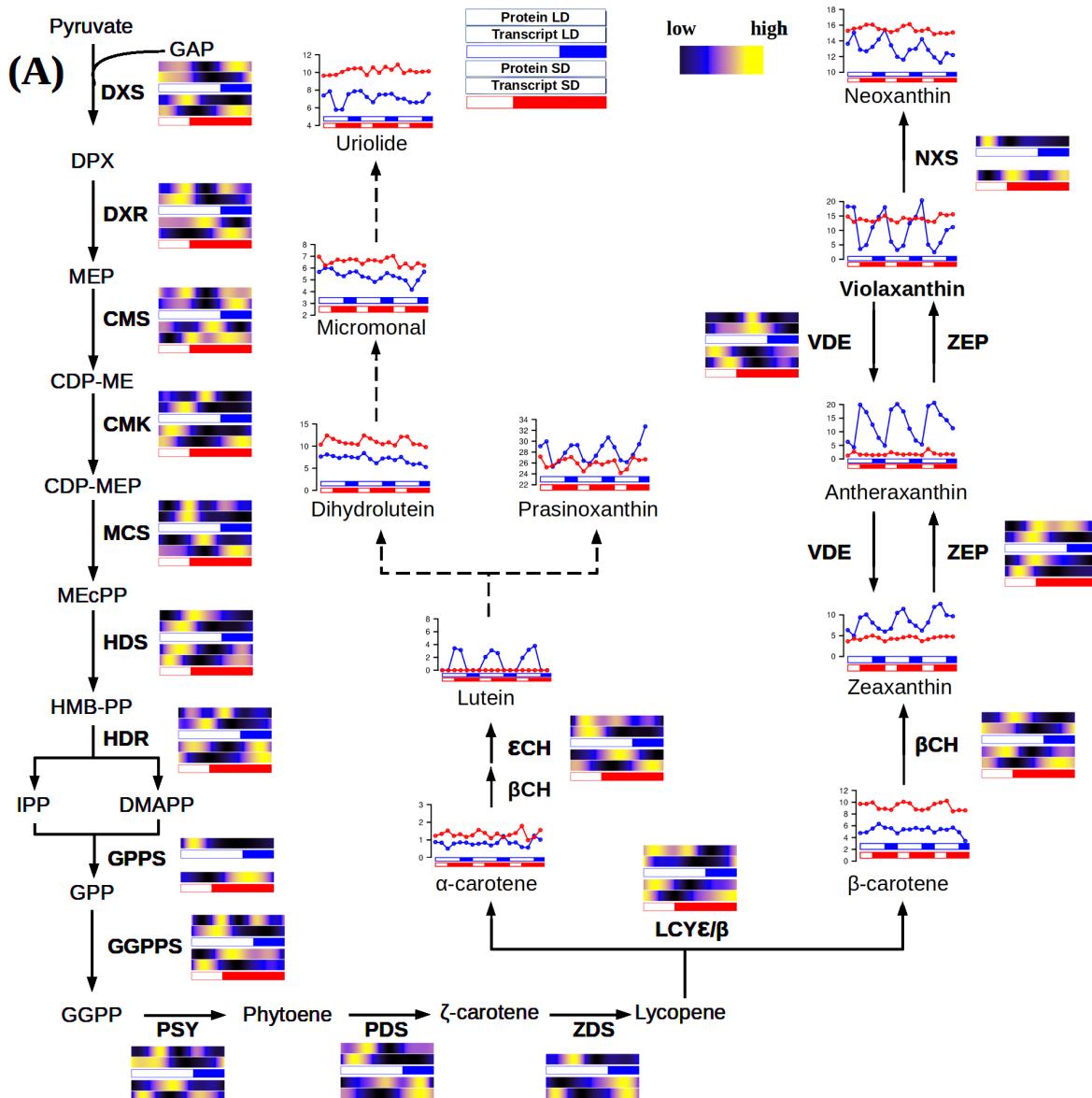
Microalgae have emerged as a promising source for large-scale production of carotenoids. However, the full potential of this technology has not been reached due to limited understanding of the molecular mechanisms underlying carotenoid biosynthesis. Over the past two decades, numerous research groups have been studying growth conditions, microalgae metabolism and optimizing photobioreactors design to maximize carotenoid production while minimizing associated costs (Del Campo et al., 2004; Hoys et al., 2021; Sierra et al., 2008). Industrial-scale cultivation of microalgae is predominantly achieved in outdoor settings. Therefore, comprehending the oscillation patterns of carotenoid biosynthesis under diel and seasonal cycles becomes crucial for ensuring the maximum carotenoid content at harvesting time, as well as identifying potential gene and protein targets for further optimization. To understand the

adaptive nature of carotenoid content to seasonal variations in diel cycles and its implications for optimizing light energy capture, photoprotection and, thus, it's possibilities of industrial optimization, the transcript and protein abundance profiles of carotenoid biosynthesis genes in *Ostreococcus* are examined. In addition, these profiles are integrated with carotenoid content profiles (Appendix 8-9) as biological validation of the results (Fig. 46).

### ***Integration of multi-omics data with oscillations described by carotenoids content in *Ostreococcus tauri* under diel and seasonal cycles***

*Ostreococcus tauri* has some of the most widely distributed carotenoids among plants such as violaxanthin, antheraxanthin or zeaxanthin. Although specific carotenoids of *Mamiellophyceae* like micromonal, uriolide or prasinoxanthin are also found in this prasinophyte, being prasinoxanthin the most abundant one (Egelund et al., 1995; Guyon et al., 2018; Six et al., 2009). Carotenoids content during diel cycles under both LD and SD conditions have been estimated (Appendix 8-9). The rhythmicity analysis detect rhythmic abundance profiles with periods of 24 h in all carotenoids under both photoperiods, except for lutein and violaxanthin, that do not maintain their rhythmicity under SD condition (Appendix 10). In general, fluctuations on carotenoids content SD entrained cultures are less drastic, resulting in a decreased wave amplitude (Fig. 46-A).

*Ostreococcus* genome presents genes encoding for the Methylerythritol 4-phosphate (MEP) pathway (Derelle et al., 2006; L. Zhao et al., 2013), which derives pyruvate to the production of geranyl pyrophosphate (GPP), the main carotenoid precursor (Fig. 46-A). Most genes encoding enzymes involved in the MEP pathway peaked at sunrise under long photoperiod and during the last part of the night under short photoperiod, preceding the corresponding protein abundance peaks by 4 hours or less (Fig. 46-B). Similar patterns are observed in the first enzymes of the carotenogenesis pathway (Fig. 46-A). The progression of the MEP and carotenogenesis pathways allow the cell to produce the main precursors needed for carotenoids synthesis and seems to be transcriptionally regulated in a similar way. In both pathways, enzymes are transcribed during the night under SD ensuring the presence of its encoding proteins during the day (Fig. 46-B). From this point, the pathway diverges in two different branches, due to the Lycopene  $\epsilon/\beta$  cyclase (LCY $\epsilon/\beta$ , ostta14g00700) that seems to be regulated in specific ways depending of the biological function of the carotenoids produced:  $\beta$ -branch, including the xanthophylls cycle; and  $\alpha$ -branch, including the main antenna carotenoids in prasinophyte (Fig. 46-A) which biosynthesis pathways are still unknown (Guyon et al., 2018;



**Figure 46: Integration of multi-omics data from the complete carotenoids biosynthesis pathway and carotenoids content of *Ostreococcus tauri*. (A) Schematic MEP pathway and carotenogenesis pathway according to (Egeland et al., 1997).**

For each enzyme its corresponding protein and transcript abundances under LD, top in blue, and SD, red, are represented using heatmaps where black represents low, blue medium and yellow high abundances respectively; (B) Organized list of the gene IDs involved in these pathways including multiomic data. Heatmaps represent, from left to right, transcript abundance under LD and SD, protein abundance under LD and SD; (C) Content oscillations of each carotenoid from the  $\beta$  and  $\alpha$  branch represented using heatmaps where black represents low, red medium and yellow high abundances respectively.

The xanthophyll cycle, the interconversion between violaxanthin, antheraxanthin and zeaxanthin as a response to light intensity (Jahns et al., 2009), is especially active in LD, although its activity is also detected in SD. The changes in these xanthophylls coincide with the accumulation of transcripts and proteins encoded by genes associated with the xanthophyll cycle, with short temporal offsets (Fig. 46-A). In LD entrained cultures, the maximum protein abundance of violaxanthin de-epoxidase (VDE, ostta16g00660) (Fig. 46-B) match the increasing zeaxanthin content during the light hours and the increasing content of violaxanthin during the dark hours (Fig. 46-A). These enzymes are transcribed sequentially with a clear temporal regulation.  $\beta$ CH gene is expressed early in the morning, sequentially followed by ZEP and VDE in that specific order under both photoperiods (Fig. 46-B). However, there is not a significant amount of zeaxanthin being accumulated in SD entrained cultures. Instead, a high level of violaxanthin without drastic variations is maintained during diel cycles (Fig. 46-A). It suggests that xanthophylls cycle is not enhancing the production of zeaxanthin due to the limited daylight hours in winter photoperiod. In some studies regarding irradiance stress in *Ostreococcus* and other prasinophytes, variations in xanthophylls contents have been described, in a similar way to the ones observed in this thesis due to short light periods (Böhme et al., 2002; Guyon et al., 2018; Six et al., 2009). In general, the enzymes of the  $\beta$ -branch pathway seem to present a strong transcriptional regulation to sequentially achieve their roles at the right time. However, this transcriptional regulation does not reach the physiological level in SD conditions (Fig. 46-A), suggesting that, despite the strong temporal transcriptional regulation that is also reflected in proteins abundance profile (Fig. 46-C), there are other regulatory pathways that balance xanthophylls production in winter photoperiod.

Regarding pigment of the  $\alpha$ -branch biosynthesis pathway, a lack of lutein during short photoperiod is observed (Fig. 46-A). However, during long photoperiod, lutein is accumulated during the light hours, followed by the increment of prasinoxanthin content after sunset

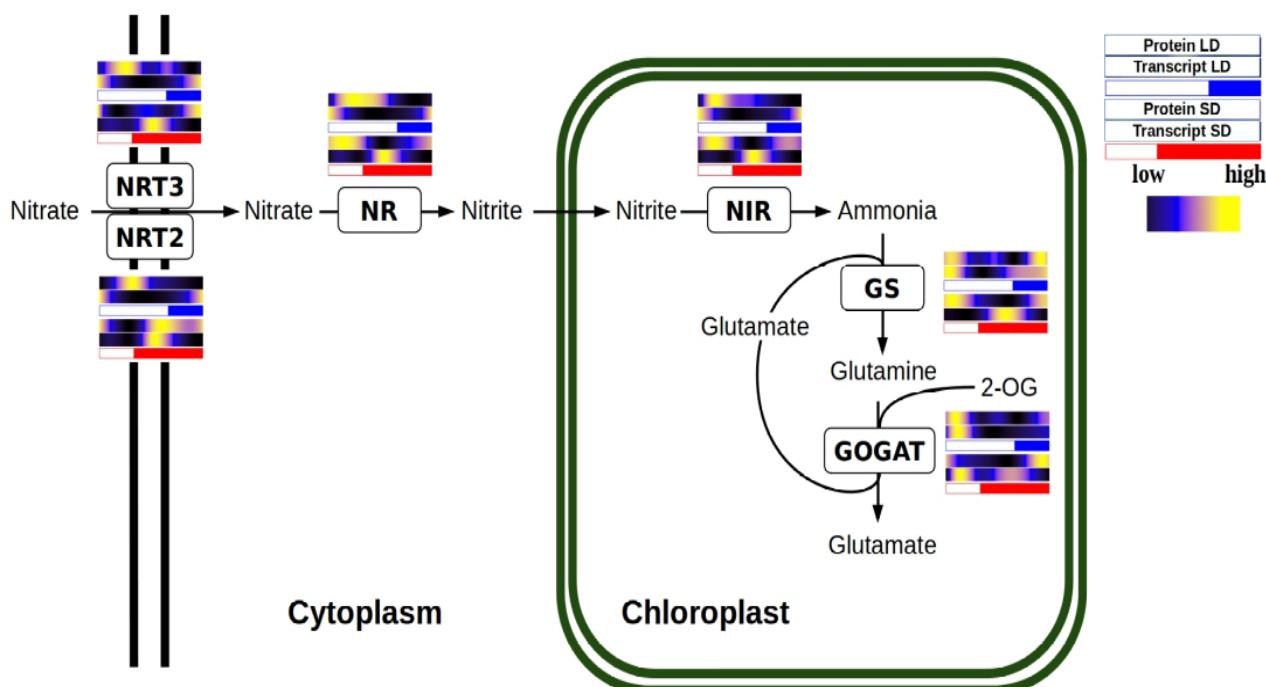
(Fig. 46-C). Lutein and prasinoxanthin contents behavior in LD entrained cultures are similar to the one described in irradiance stress experiments in other prasinophytes. In *Mantoniella squamata* the accumulation of lutein is linked to irradiance stress and its following conversion to prasinoxanthin when the stress condition is over (Böhme et al., 2002). Although the α-branch biosynthesis pathway is still unknown, there are some hypotheses about the inter-conversion of lutein-prasinoxanthin (Egelund et al., 1997). This could be supported by our results in *Ostreococcus* since the content of lutein and prasinoxanthin seem to be linked, not only under irradiance stress, as it has been described in previous published studies (Böhme et al., 2002), but under different photoperiods as well. However, the enzymes involved in the interconversion of these carotenoids remain to be identified, and a comparison of their gene expression and protein accumulation is not feasible in this study. In summary, *Ostreococcus tauri* carotenogenesis presents the common characteristics of a process regulated by the circadian clock, as being able to adapt to different photoperiods and presenting an anticipation to diel cycles at transcriptomic level.

### **Nitrate assimilation under diel and seasonal cycles in *Ostreococcus tauri***

Nitrogen is an essential component in biomolecules of all living beings. In the atmosphere, N<sub>2</sub> is the most abundant form of nitrogen. This gas is dissolved in water ecosystems but is inaccessible for microalgae. In fact, nitrogen is a major limiting nutrient of marine phytoplankton (Barros et al., 2005; Mittag, 2001; Sanz-Luque et al., 2015). *Ostreococcus tauri* have developed competitive mechanisms to ensure nitrogen assimilation in the marine ecosystem. It can grow on nitrate, ammonium and urea, and complete sets of genes allowing transport and assimilation of these substrates have been identified in its genome (Blanc-Mathieu et al., 2014; Derelle et al., 2006). In the previous chapters of this thesis, nitrate assimilation has been identified as one of the biological processes which genes and proteins present significant rhythmic profiles under diel cycles with large offset between gene expression and translation. In this section, enzymatic activities of two of the main enzymes involved in the nitrate assimilation pathway are integrated with the transcript and protein abundance profiles of the complete pathway. These results give new insights on the adaptive response of this assimilation process to seasonal variations in diel cycles and its implications for optimizing nutrient uptake and metabolism.

## **Integration of key enzyme activities from nitrate assimilation pathway with multi-omic data**

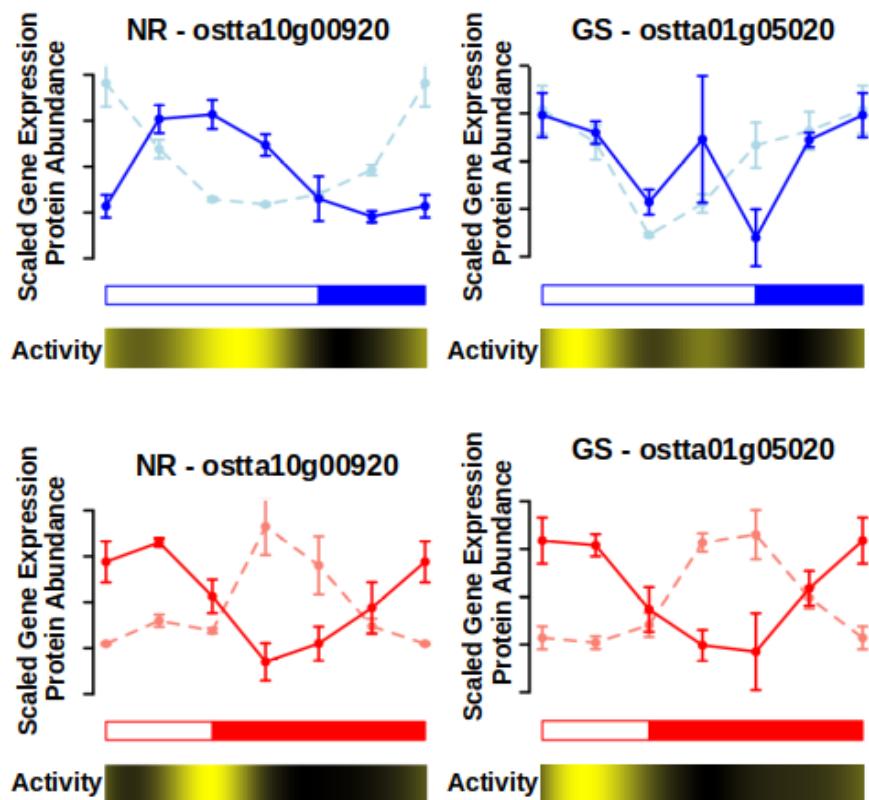
Nitrate is first transported into the cell by Nitrate Transporters 2 and 3 (NRT2, ostta10g00950 and NRT3, ostta10g00940), followed by its reduction to nitrite by Nitrate Reductase (NR, ostta10g00920). Nitrite is further reduced to ammonia by Nitrite Reductase (NIR, ostta10g00930). The central part of nitrate assimilation is played by the Glutamine Synthetase (GS, ostta01g05020) and Glutamine Oxoglutarate Aminotransferase (GOGAT, ostta14g01900) cycle, which converts inorganic nitrogen, ammonia, into glutamine and glutamate (Fig. 47), a central precursor for the biosynthesis of nitrogen-containing compounds such as amino acids and nucleotides (Sanz-Luque et al., 2015).



**Figure 47: Multi-omics integration of nitrate assimilation pathway.** Schematic nitrate assimilation pathway including protein abundance and gene expression profiles of each enzyme. For each enzyme its corresponding protein and transcript abundances under LD, top in blue, and SD, bottom in red, are represented using heatmaps where black represents low, blue medium and yellow high abundances respectively.

In LD entrained cultures, gene expression profiles of NRT2/3, NR and NIR reached their maximum at dawn (ZT0), while their protein abundances peaked 8 hours later at midday (ZT8), coinciding with the time point of maximum light irradiance. GS and GOGAT gene expression and protein abundance profiles are almost coincident, reaching their maximum at

the beginning of the day (ZT4) without noticeable temporal offsets between them. However, a slight increase in protein abundance was detected at the end of the day (ZT12) for both GS and GOGAT (Fig. 47). In contrast, in SD entrained cultures, all genes encoding the transporters and enzymes involved in nitrate assimilation show their maximum expression level during the first part of the night (ZT12-ZT16), preceding their protein abundance peaks at dawn (ZT0) or midday (ZT4) by 8 hours or more. Notably, GOGAT gene expression display a bimodal pattern under SD condition, maintaining the peak observed under LD at ZT4 besides the new peak at ZT16. Therefore, GOGAT gene expression constitutes an example of the emergence of complex expression patterns under winter photoperiod consisting of two gene expression peaks per day (Fig. 47).



**Figure 48: NR and GS rhythmic activity compared with its proteomic and transcriptomic data generated.** Transcript (lighter color) and protein (darker color) abundance profiles for Nitrate reductase (NR, left) and Glutamine Synthetase (GS, right) under LD condition (top, blue) and SD condition (bottom, red). Heatmaps are incorporated below to represent the changes in enzymatic activity of these enzymes. Black represents low activity and yellow high activity.

Circadian oscillations in expression and activity of the first enzyme of this pathway (NR) have been described in *Arabidopsis* and crop plants as maize or tomato (Lillo et al., 2001; Lillo & Ruoff, 1989; Tucker et al., 2004; Z. Yang & Midmore, 2005). In fact, light is apparently an important factor for NR to maintain its rhythmic behavior. Rhythms in NR activity or NR gene expression profiles are shown to persist only in continuous light in plants (Lillo et al., 2001; Lillo & Ruoff, 1989). These results are in agreement with the transcriptomic data obtained in this work, where NR gene expression rhythmic profiles, as well as other enzymes involved in this pathway, are maintained only under light-dark cycles and constant light. To validate these results, the enzymatic activity of NR and GS are measured throughout complete diel cycles under long and short photoperiods (Appendix 11). These measurements present a significant rhythmic profile with a p-value lower than 0.05 and an almost non-existent offset between their protein abundance profiles and their activity profiles is found (Fig. 48). The huge transcriptomic anticipation observed is adjusted by the clock taking in count the large offset between gene expression and translation described by the enzymes involved in this pathway. In fact, although genes are transcribed at different times under short and long photoperiod, proteins reach their maximum abundance level at a similar time of the day in both cases (Fig. 47, 48). This is an example of how *Ostreococcus* adjust its transcriptional programme in order to ensure the presence of proteins at the exact right time, in spite of their specific translation offset.



# Conclusions



- The web-based tools ALGAEFUN with MARACAS constitute a noteworthy contribution to the microalgae research community. The platform has garnered substantial traction, accumulating more than 3.40k visits, underscoring its utility and positive reception within the scientific community.
- Around 80 % of the *Ostreococcus* transcriptome present rhythmic expression patterns under diel cycles, but *bona fide* circadian genes represent only the 4.6 % of the transcriptome.
- Although *Ostreococcus* is a photosynthetic organism, the most of its transcripts present their maximum levels of abundance during dark periods and its transcriptome strongly depends on those dark periods to maintain rhythmic expression patterns. The effects of the free-running conditions observed over the transcriptome (greater desynchronization under LL, positive phase shifts under LL and negative phase shifts under DD) agree with the ones described in activity circadian records of nocturnal organisms.
- Seasonal cycles induce changes in gene expression profiles consisting of the emergence of 12 h period cycles, phase shifts and amplitude reductions.
- The 12 h period cycles (2 peaks every 24 h) emerged under short day are not a self-sustained rhythm, but a combination of two distinct rhythmic profiles: one depending on the photoperiod and another depending on the skotoperiod. These rhythmic profiles are coincident under long photoperiods, becoming out of phase as photoperiod gets shorter.
- Only 25 % of the *Ostreococcus* proteome present rhythmic protein abundances under diel cycles.
- Despite the nocturnal transcriptomic activity in *Ostreococcus tauri*, its proteome present a clear diurnal activity (the maximum protein abundance levels take place during the light hours).
- There is a temporal offset between gene expression and translation. The rhythmic wave described by a gene expression level and its encoding protein abundance are almost identical, except for their phase offsets.
- The offset between gene expression and translation increases during short day and its dependent on the physiological function in which proteins are involved. Therefore,

*Ostreococcus tauri* severely reorganizes its transcriptional temporal programme taking in count not only the current photoperiod but also the different transcript-proteins offsets that each biological function presents.

- There is a temporal offset between the proteome of *Ostreococcus tauri* and the execution of their physiological processes.
- Cell division cycle is strongly influenced by the circadian clock in *Ostreococcus*, it can anticipate cyclic changes like the short day in winter photoperiods. To anticipate it, the circadian clock guarantees the completion of the cellular division phase prior to sunrise, ensuring the optimization of light hours without wastage. The cell cycle temporal programme observed is in agreement with the transcriptomic and proteomic data generated.
- Photosynthetic efficiency rhythmic oscillations present a 12 h period cycles under short day, which is again in agreement with the transcriptomic and proteomic data generated. Photosynthesis machinery anticipation to photoperiod is transcriptionally regulated in *Ostreococcus tauri*.
- Starch content in *Ostreococcus tauri* presents a rhythmic 24 h cycle, reaching its maximum starch content at the high irradiance hours under both long and short day. Both APL and AMY protein abundance profiles are strongly coincident with their gene expression profile with a ~4 h offset, showing that starch circadian synthesis-degradation balance is possibly transcriptionally regulated.
- *Ostreococcus tauri* carotenogenesis and nitrate assimilation pathways present the common characteristics of processes regulated by the circadian clock (at a transcriptomic, proteomic and physiological level), as being able to adapt to different photoperiods and present an anticipation to diel cyclic changes.





# Bibliography



Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., Akiyama, J. A., Jammal, O. Al, Amrhein, H., Anderson, S. M., Andrews, G. R., Antoshechkin, I., Ardlie, K. G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A. A., Barnes, I. H. A., Barozzi, I. Myers, R. M. (2020). Perspectives on ENCODE. *Nature*, 583(7818). <https://doi.org/10.1038/s41586-020-2449-8>

Abinandan, S., Subashchandrabose, S. R., Venkateswarlu, K., & Megharaj, M. (2018). Nutrient removal and biomass production: advances in microalgal biotechnology for wastewater treatment. *Critical Reviews in Biotechnology*, 38(8), 1244–1260. <https://doi.org/10.1080/07388551.2018.1472066>

Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L. B., Coppersmith, J., Kwok, K., Peach, L., Orchard, E., Kalb, R., Xu, W., Carlson, T. J., Francis, K., Konigsfeld, K., Bartalis, J., Schultz, A., Lambert, W., Schwartz, A. S., Brown, R., & Moellering, E. R. (2017). Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nature Biotechnology*, 35, 647–652. <https://doi.org/10.1038/nbt.3865>

Akhtar, R. A., Reddy, A. B., Maywood, E. S., Clayton, J. D., King, V. M., Smith, A. G., Gant, T. W., Hastings, M. H., & Kyriacou, C. P. (2002). Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Current Biology*, 12(7). [https://doi.org/10.1016/S0960-9822\(02\)00759-5](https://doi.org/10.1016/S0960-9822(02)00759-5)

Andreatta, G., & Tessmar-Raible, K. (2020). The Still Dark Side of the Moon: Molecular Mechanisms of Lunar-Controlled Rhythms and Clocks. *Journal of Molecular Biology*, 432(12). <https://doi.org/10.1016/j.jmb.2020.03.009>

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25, 25–29. <https://doi.org/10.1038/75556>

Bachy, C., Wittmers, F., Muschiol, J., Hamilton, M., Henrissat, B., & Worden, A. Z. (2022). The Land-Sea Connection: Insights Into the Plant Lineage from a Green Algal Perspective. *Annual Review of Plant Biology*, 73. <https://doi.org/10.1146/annurev-arplant-071921-100530>

Ballance, H., & Zhu, B. (2021). Revealing the hidden reality of the mammalian 12-h ultradian rhythms. *Cellular and Molecular Life Sciences*, 78(7), 3127–3140. <https://doi.org/10.1007/s00018-020-03730-5>

Barros, M. P., Pinto, E., Sigaúd-Kutner, T. C. S., Cardozo, K. H. M., & Colepicolo, P. (2005). Rhythmicity and oxidative/nitrosative stress in algae. *Biological Rhythm Research*, 36(1–2). <https://doi.org/10.1080/09291010400028666>

Bartoszewicz, R., Chmielewska, D., Domoń, M., & Barbacka-Surowiak, G. (2010). Influence of short-term constant light on phase shift of mouse circadian locomotor activity rhythm induced by agonist and antagonist of serotonin. *Biological Rhythm Research*, 41(4). <https://doi.org/10.1080/09291010903018016>

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J. P., & Delignette-Muller, M. L. (2015). A toolbox for nonlinear regression in R: The package nlstools. *Journal of Statistical Software*, 66(5). <https://doi.org/10.18637/jss.v066.i05>

Becker, B., & Marin, B. (2009). Streptophyte algae and the origin of embryophytes. *Annals of Botany*, 103(7), 999–1004. <https://doi.org/10.1093/aob/mcp044>

Benites, L. F., Bucchini, F., Sanchez-Brosseau, S., Grimsley, N., Vandepoele, K., & Piganeau, G. (2021). Evolutionary Genomics of Sex-Related Chromosomes at the Base of the Green Lineage. *Genome Biology and Evolution*, 13(10). <https://doi.org/10.1093/gbe/evab216>

Berube, P. M., Biller, S. J., Hackl, T., Hogle, S. L., Satinsky, B. M., Becker, J. W., Braakman, R., Collins, S. B., Kelly, L., Berta-Thompson, J., Coe, A., Bergauer, K., Bouman, H. A., Browning, T. J., De Corte, D., Hassler, C., Hulata, Y., Jacquot, J. E., Maas, E. W., Chisholm, S. W. (2018). Data descriptor: Single cell genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine environments. *Scientific Data*, 5. <https://doi.org/10.1038/sdata.2018.154>

Biebach, H., Falk, H., & Krebs, J. R. (1991). The Effect of Constant Light and Phase Shifts on a Learned Time-Place Association in Garden Warblers (*Sylvia borin*): Hourglass or Circadian Clock? *Journal of Biological Rhythms*, 6(4). <https://doi.org/10.1177/074873049100600406>

Binkley, S., & Mosher, K. (1985). Direct and circadian control of sparrow behavior by light and dark. *Physiology and Behavior*, 35(5). [https://doi.org/10.1016/0031-9384\(85\)90413-5](https://doi.org/10.1016/0031-9384(85)90413-5)

Blaby, I. K., Blaby-Haas, C. E., Tourasse, N., Hom, E. F. Y., Lopez, D., Aksoy, M., Grossman, A., Umen, J., Dutcher, S., Porter, M., King, S., Witman, G. B., Stanke, M., Harris, E. H., Goodstein, D., Grimwood, J., Schmutz, J., Vallon, O., Merchant, S. S., & Prochnik, S. (2014). The *Chlamydomonas* genome project: A decade on. *Trends in Plant Science*, 19(10). <https://doi.org/10.1016/j.tplants.2014.05.008>

Blanc-Mathieu, R., Verhelst, B., Derelle, E., Rombauts, S., Bouget, F.-Y., Carré, I., Château, A., Eyre-Walker, A., Grimsley, N., Moreau, H., Piégu, B., Rivals, E., Schackwitz, W., Van de Peer, Y., & Piganeau, G. (2014). An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics*, 15(1), 1103. <https://doi.org/10.1186/1471-2164-15-1103>

Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D. D., Gurnon, J., Ladunga, I., Lindquist, E., Lucas, S., Pangilinan, J., Pröschold, T., Salamov, A., Schmutz, J., Weeks, D., Yamada, T., Lomsadze, A., Borodovsky, M., Claverie, J. M., van Etten, J. L. (2012). The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biology*, 13(5). <https://doi.org/10.1186/gb-2012-13-5-r39>

Bläsing, O. E., Gibon, Y., Günther, M., Höhne, M., Morcuende, R., Osuna, D., Thimm, O., Usadel, B., Scheible, W. R., & Stitt, M. (2005). Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell*, 17(12). <https://doi.org/10.1105/tpc.105.035261>

Böhme, K., Wilhelm, C., & Goss, R. (2002). Light regulation of carotenoid biosynthesis in the prasinophycean alga *Mantoniella squamata*. *Photochemical and Photobiological Sciences*, 1(8). <https://doi.org/10.1039/b204965c>

Bolaños, L. M., Karp-Boss, L., Choi, C. J., Worden, A. Z., Graff, J. R., Haëntjens, N., Chase, A. P., Della Penna, A., Gaube, P., Morison, F., Menden-Deuer, S., Westberry, T. K., O’Malley, R. T., Boss, E., Behrenfeld, M. J., & Giovannoni, S. J. (2020). Small phytoplankton dominate western North Atlantic biomass. *ISME Journal*, 14(7). <https://doi.org/10.1038/s41396-020-0636-0>

Borowitzka, M. A. (2013). High-value products from microalgae-their development and commercialisation. *Journal of Applied Phycology*, 25(3), 743–756. <https://doi.org/10.1007/s10811-013-9983-9>

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Grigoriev, I. V. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, 456(7219). <https://doi.org/10.1038/nature07410>

Brandoli, C., Petri, C., Egea-Cortines, M., & Weiss, J. (2020). Gigantea: Uncovering new functions in flower development. *Genes*, 11(10), 1–15. <https://doi.org/10.3390/genes11101142>

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5). <https://doi.org/10.1038/nbt.3519>

Bruce, V. G. (1970). The biological clock in *Chlamydomonas reinhardtii*. *J Protozool*, 17, 328–334. <https://doi.org/https://doi.org/10.1111/j.1550-7408.1970.tb02380.x>

Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N. L., Lewis, S. E., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., Westerfield, M. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1), D330–D338. <https://doi.org/10.1093/nar/gky1055>

Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., Lomax, J., Mungall, C., Hitz, B., Balakrishnan, R., Dolan, M., Wood, V., Hong, E., & Gaudet,

P. (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics*, 25(2), 288–289. <https://doi.org/10.1093/bioinformatics/btn615>

Carlson, M., & Pagès, H. (2019). AnnotationForge: Tools for building SQLite-based annotation data packages (1.26.0.).

Carneiro, A. K., Montessoro, P. da F., Fusaro, A. F., Araújo, B. G., & Hemerly, A. S. (2021). Plant cdks driving the cell cycle through climate change. *Plants*, 10(9). <https://doi.org/10.3390/plants10091804>

Chapman, R. L. (2013). Algae: The world's most important “plants”—an introduction. *Mitigation and Adaptation Strategies for Global Change*, 18(1). <https://doi.org/10.1007/s11027-010-9255-9>

Chen, H., Li, T., & Wang, Q. (2019). Ten years of algal biofuel and bioproducts: gains and pains. *Planta*, 249(1), 195–219. <https://doi.org/10.1007/s00425-018-3066-8>

Chen, M. X., Zhang, Y., Fernie, A. R., Liu, Y. G., & Zhu, F. Y. (2021). SWATH-MS-Based Proteomics: Strategies and Applications in Plants. *Trends in Biotechnology*, 39(5), 433–437. Elsevier Ltd. <https://doi.org/10.1016/j.tibtech.2020.09.002>

Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., Sun, W., Li, X., Xu, Y., Zhang, Y., Wittek, S., Reder, T., Günther, G., Gontcharov, A., Wang, S., Li, L., Liu, X., Wang, J., Yang, H., Melkonian, M. (2019). Genomes of Subaerial *Zygnematophyceae* Provide Insights into Land Plant Evolution. *Cell*, 179(5). <https://doi.org/10.1016/j.cell.2019.10.019>

Cock, J. M., & Coelho, S. M. (2011). Algal models in plant biology. *Journal of Experimental Botany*, 62(8). <https://doi.org/10.1093/jxb/err117>

Collado-Fabbri, S., Vaulot, D., & Ulloa, O. (2011). Structure and seasonal dynamics of the eukaryotic picophytoplankton community in a wind-driven coastal upwelling ecosystem. *Limnology and Oceanography*, 56(6). <https://doi.org/10.4319/lo.2011.56.6.2334>

Corellou, F., Camasses, A., Ligat, L., Peaucellier, G., & Bouget, F. Y. (2005). Atypical regulation of a green lineage-specific B-type cyclin-dependent kinase. *Plant Physiology*, 138(3). <https://doi.org/10.1104/pp.105.059626>

Correa, A., & Bell-Pedersen, D. (2002). Distinct signaling pathways from the circadian clock participate in regulation of rhythmic conidiospore development in *Neurospora crassa*. *Eukaryotic Cell*, 1(2). <https://doi.org/10.1128/EC.1.2.273-280.2002>

Corteggiani Carpinelli, E., Telatin, A., Vitulo, N., Forcato, C., D'Angelo, M., Schiavon, R., Vezzi, A., Giacometti, G. M., Morosinotto, T., & Valle, G. (2014). Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Molecular Plant*, 7(2). <https://doi.org/10.1093/mp/sst120>

Covington, M. F., Maloof, J. N., Straume, M., Kay, S. A., & Harmer, S. L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biology*, 9(8). <https://doi.org/10.1186/gb-2008-9-8-r130>

Craig, R. J., Hasan, A. R., Ness, R. W., & Keightley, P. D. (2021). Comparative genomics of *Chlamydomonas*. *Plant Cell*, 33(4). <https://doi.org/10.1093/plcell/koab026>

Critchley, C., & Russell, A. W. (1994). Photoinhibition of photosynthesis in vivo: The role of protein turnover in photosystem II. *Physiologia Plantarum*, 92(1). <https://doi.org/10.1111/j.1399-3054.1994.tb06670.x>

Cui, Y., Thomas-Hall, S. R., & Schenk, P. M. (2019). *Phaeodactylum tricornutum* microalgae as a rich source of omega-3 oil: Progress in lipid induction techniques towards industry adoption. *Food Chemistry*, 297. <https://doi.org/10.1016/j.foodchem.2019.06.004>

Cumming, B. G., & Wagner, E. (1968). Rhythmic processes in plants. *Annual Review of Plant Physiology*, 19(1). <https://doi.org/10.1146/annurev.pp.19.060168.002121>

De Candolle, M. (1832). Physiologie végétale. *Bechet Jeune*.

De Keersmaecker, S. C. J., Thijs, I. M. V., Vanderleyden, J., & Marchal, K. (2006). Integration of omics data: How well does it work for bacteria?. *Molecular Microbiology*, 62(5). <https://doi.org/10.1111/j.1365-2958.2006.05453.x>

de los Reyes, P., Romero-Campero, F. J., Ruiz, M. T., Romero, J. M., Valverde, F., Teresa Ruiz, M., Romero, J. M., & Valverde, F. (2017). Evolution of daily gene co-expression patterns from algae to plants. *Frontiers in Plant Science*, 8(1217), 1–22. <https://doi.org/10.3389/fpls.2017.01217>

De Mairan, J. J. . (1729). Observation Botanique. *Histoire de l'Academie Royale Des Sciences*.

Del Campo, J. A., Rodríguez, H., Moreno, J., Vargas, M. Á., Rivas, J., & Guerrero, M. G. (2004). Accumulation of astaxanthin and lutein in *Chlorella zofingiensis* (*Chlorophyta*). *Applied Microbiology and Biotechnology*, 64(6). <https://doi.org/10.1007/s00253-003-1510-5>

Demir-Hilton, E., Sudek, S., Cuvelier, M. L., Gentemann, C. L., Zehr, J. P., & Worden, A. Z. (2011). Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. *ISME Journal*, 5(7). <https://doi.org/10.1038/ismej.2010.209>

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynie, S., Cooke, R., Saeys, Y., Wuyls, J., Jabbari, K., Bowler, C., Panaud, O., Piegu, B., Ball, S. G., Ral, J.-P., Bouget, F.-Y., Moreau, H. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences*, 103(31), 11647–11652. <https://doi.org/10.1073/pnas.0604795103>

Ditz, B., Boekhoudt, J. G., Aliee, H., Theis, F. J., Nawijn, M., Brandsma, C.-A., Hiemstra, P. S., Timens, W., Tew, G. W., Grimaldeston, M. A., Neighbors, M., Guryev, V., van den Berge, M., & Faiz, A. (2021). Comparison of genome-wide gene expression profiling by RNA Sequencing versus microarray in bronchial biopsies of COPD patients before and after inhaled corticosteroid treatment: does it provide new insights?. *ERJ*

Dodd, A. N., Kusakina, J., Hall, A., Gould, P. D., & Hanaoka, M. (2014). The circadian regulation of photosynthesis. *Photosynthesis Research*, 119(1–2). <https://doi.org/10.1007/s11120-013-9811-8>

Donnan, L., & John, P. C. L. (1983). Cell cycle control by timer and sizer in *Chlamydomonas*. *Nature*, 304(5927). <https://doi.org/10.1038/304630a0>

Duhamel du Monceau, H. L. (1759). La physique des arbres. *H. L. Guerin & L. F. Delatour*.

Edmunds, L. N. (1983). Chronobiology at the cellular and molecular levels: Models and mechanisms for circadian timekeeping. *American Journal of Anatomy*, 168(4). <https://doi.org/10.1002/aja.1001680404>

Edmunds, L. N., & Laval-Martin, D. L. (2019). Cell Division Cycles And Circadian Oscillators In *Euglena*. Intracellular space as oligogenetic ecosystem. *Proceedings*, 1(1), 1-9 <https://doi.org/10.1515/9783110841237-033>

Edwards, K. D., Akman, O. E., Knox, K., Lumsden, P. J., Thomson, A. W., Brown, P. E., Pokhilko, A., Kozma-Bognar, L., Nagy, F., Rand, D. A., & Millar, A. J. (2010). Quantitative analysis of regulatory flexibility under changing environmental conditions. *Molecular Systems Biology*, 6. <https://doi.org/10.1038/msb.2010.81>

Eelderink-Chen, Z., Bosman, J., Sartor, F., Dodd, A. N., Kovács, Á. T., & Merrow, M. (2021). A circadian clock in a nonphotosynthetic prokaryote. *Science Advances*, 7(2). <https://doi.org/10.1126/sciadv.abe2086>

Egeland, E. S., Eikrem, W., Throndsen, J., Wilhelm, C., Zapata, M., & Liaaen-Jensen, S. (1995). Carotenoids from further prasinophytes. *Biochemical Systematics and Ecology*, 23(7–8). [https://doi.org/10.1016/0305-1978\(95\)00075-5](https://doi.org/10.1016/0305-1978(95)00075-5)

Egeland, E. S., Guillard, R. R. L., & Liaaen-Jensen, S. (1997). Additional carotenoid prototype representatives and a general chemosystematic evaluation of carotenoids in *Prasinophyceae* (*Chlorophyta*). *Phytochemistry*, 44(6) [https://doi.org/10.1016/S0031-9422\(96\)00650-4](https://doi.org/10.1016/S0031-9422(96)00650-4)

Eggersdorfer, M., & Wyss, A. (2018). Carotenoids in human nutrition and health. *Archives of Biochemistry and Biophysics*, 652. <https://doi.org/10.1016/j.abb.2018.06.001>

Eissenberg, J. C., & Elgin, S. C. R. (2006). Marking time. *Nature Genetics*, 38(3). <https://doi.org/10.1038/ng0306-276>

Engel, S. R., Dietrich, F. S., Fisk, D. G., Binkley, G., Balakrishnan, R., Costanzo, M. C., Dwight, S. S., Hitz, B. C., Karra, K., Nash, R. S., Weng, S., Wong, E. D., Lloyd, P., Skrzypek, M. S., Miyasato, S. R., Simison, M., & Cherry, J. M. (2014). The Reference Genome Sequence of *Saccharomyces cerevisiae*: Then and Now. *G3: Genes, Genomes, Genetics*, 4(3). <https://doi.org/10.1534/g3.113.008995>

Evans, D. R. (1961). Biological Clocks. Volume XXV. Cold Spring Harbor Symposia on Quantitative Biology. *The Quarterly Review of Biology*, 36(3). <https://doi.org/10.1086/403447>

Fauré-Fremiet, E. (1951). The tidal rhythm of the diatom *Hantzschia amphioxys*. *Biological Bulletin*, 100(3), 173–177.

Ferrari, C., Proost, S., Janowski, M., Becker, J., Nikoloski, Z., Bhattacharya, D., Price, D., Tohge, T., Bar-Even, A., Fernie, A., Stitt, M., & Mutwil, M. (2019). Kingdom-wide comparison reveals the evolution of diurnal gene expression in *Archaeplastida*. *Nature Communications*, 10(737). <https://doi.org/10.1038/s41467-019-08703-2>

Feugier, F. G., & Satake, A. (2013). Dynamical feedback between circadian clock and sucrose availability explains adaptive response of starch metabolism to various photoperiods. *Frontiers in Plant Science*, 3. <https://doi.org/10.3389/fpls.2012.00305>

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, 281(5374). <https://doi.org/10.1126/science.281.5374.237>

Finegan, C., Boehlein, S. K., Leach, K. A., Madrid, G., Hannah, L. C., Koch, K. E., Tracy, W. F., & Resende, M. F. R. (2022). Genetic perturbation of the starch

biosynthesis in maize endosperm reveals sugar-responsive gene networks. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.800326>

Flis, A., Mengin, V., Ivakov, A. A., Mugford, S. T., Hubberten, H. M., Encke, B., Krohn, N., Höhne, M., Feil, R., Hoefgen, R., Lunn, J. E., Millar, A. J., Smith, A. M., Sulpice, R., & Stitt, M. (2019). Multiple circadian clock outputs regulate diel turnover of carbon and nitrogen reserves. *Plant Cell and Environment*, 42(2). <https://doi.org/10.1111/pce.13440>

Flis, A., Sulpice, R., Seaton, D. D., Ivakov, A. A., Liput, M., Abel, C., Millar, A. J., & Stitt, M. (2016). Photoperiod-dependent changes in the phase of core clock transcripts and global transcriptional outputs at dawn and dusk in *Arabidopsis*. *Plant Cell and Environment*, 39(9). <https://doi.org/10.1111/pce.12754>

Foà, A., & Bertolucci, C. (2001). Temperature cycles induce a bimodal activity pattern in ruin lizards: Masking or clock-controlled event? A seasonal problem. *Journal of Biological Rhythms*, 16(6). <https://doi.org/10.1177/074873001129002268>

Frazee, A. C., Pertea, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., & Leek, J. T. (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature Biotechnology*, 33, 243. <https://doi.org/10.1038/nbt.3172>

Fu, L., Patel, M. S., Bradley, A., Wagner, E. F., & Karsenty, G. (2005). The molecular clock mediates leptin-regulated bone formation. *Cell*, 122(5). <https://doi.org/10.1016/j.cell.2005.06.028>

Fung-Uceda, J., Lee, K., Seo, P. J., Polyn, S., De Veylder, L., & Mas, P. (2018). The Circadian Clock Sets the Time of DNA Replication Licensing to Regulate Growth in *Arabidopsis*. *Developmental Cell*, 45(1). <https://doi.org/10.1016/j.devcel.2018.02.022>

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Alvarez, R. V., Landsman, D., & Koonin, E. V. (2021). COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, 49(D1). <https://doi.org/10.1093/nar/gkaa1018>

García-Cubero, R., Moreno-Fernández, J., & García-González, M. (2018). Potential of *Chlorella vulgaris* to abate flue gas. *Waste and Biomass Valorization*, 9(11), 2015–2019. <https://doi.org/10.1007/s12649-017-9987-9>

García-Domínguez, M., & Florencio, F. J. (1997). Nitrogen availability and electron transport control the expression of glnB gene (encoding PII protein) in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Molecular Biology*, 35(6). <https://doi.org/10.1023/A:1005846626187>

García-Plazaola, J. I., Fernández-Marín, B., Ferrio, J. P., Alday, J. G., Hoch, G., Landais, D., Milcu, A., Tissue, D. T., Voltas, J., Gessler, A., Roy, J., & Resco de Dios, V. (2017). Endogenous circadian rhythms in pigment composition induce changes in photochemical efficiency in plant canopies. *Plant Cell and Environment*, 40(7). <https://doi.org/10.1111/pce.12909>

Gaspar, J. M. (2018). Improved peak-calling with MACS2. *bioRxiv*, 496521. <https://doi.org/10.1101/496521>

Geigenberger, P. (2011). Regulation of starch biosynthesis in response to a fluctuating environment. *Plant Physiology*, 155(4). <https://doi.org/10.1104/pp.110.170399>

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1), D1178–D1186. <https://doi.org/10.1093/nar/gkr944>

Goto, K., & Johnson, C. H. (1995). Is the cell division cycle gated by a circadian clock? The case of *Chlamydomonas reinhardtii*. *Journal of Cell Biology*, 129(4). <https://doi.org/10.1083/jcb.129.4.1061>

Granados-Fuentes, D., Tseng, A., & Herzog, E. D. (2006). A circadian clock in the olfactory bulb controls olfactory responsivity. *Journal of Neuroscience*, 26(47). <https://doi.org/10.1523/JNEUROSCI.3445-06.2006>

Grigoriev, I. V., Hayes, R. D., Calhoun, S., Kamel, B., Wang, A., Ahrendt, S., Dusheyko, S., Nikitin, R., Mondo, S. J., Salamov, A., Shabalov, I., & Kuo, A. (2021). PhycoCosm, a comparative algal genomics resource. *Nucleic Acids Research*, 49(D1). <https://doi.org/10.1093/nar/gkaa898>

Gutierrez, C. (2009). The *Arabidopsis* Cell Division Cycle. *The Arabidopsis Book*, 7. <https://doi.org/10.1199/tab.0120>

Guyon, J.-B., Vergé, V., Schatt, P., Lozano, J.-C., Liennard, M., & Bouget, F.-Y. (2018). Comparative analysis of culture conditions for the optimization of carotenoid production in several strains of the picoeukaryote *Ostreococcus*. *Marine Drugs*, 16(3). <https://doi.org/10.3390/md16030076>

Hagiwara, S., Bolige, A., Zhang, Y., Takahashi, M., Yamagishi, A., & Goto, K. (2002). Circadian gating of photoinduction of commitment to cell-cycle transitions in relation to photoperiodic control of cell reproduction in *Euglena*. *Photochemistry and Photobiology*, 76(1). [https://doi.org/10.1562/0031-8655\(2002\)076<0105:cgopoc>2.0.co;2](https://doi.org/10.1562/0031-8655(2002)076<0105:cgopoc>2.0.co;2)

Hartman, M. D., Rojas, B. E., Iglesias, A. A., & Figueroa, C. M. (2023). The involvement of allosteric effectors and post-translational modifications in the control of plant central carbon metabolism. *Plant Journal*, 114(5). <https://doi.org/10.1111/tpj.16215>

Hastings, J. W., & Sweeney, B. M. (1957). The luminescent reaction in extracts of the marine dinoflagellate, *Gonyaulax polyedra*. *Journal of Cellular Physiology*, 49(2). <https://doi.org/10.1002/jcp.1030490205>

Hayes, K. R., Beatty, M., Meng, X., Simmons, C. R., Habben, J. E., & Danilevskaya, O. N. (2010). Maize global transcriptomics reveals pervasive leaf diurnal rhythms but rhythms in developing ears are largely limited to the core oscillator. *PLoS ONE*, 5(9). <https://doi.org/10.1371/journal.pone.0012887>

Henderson, G. P., Gan, L., & Jensen, G. J. (2012). 3-D Ultrastructure of *O. tauri*: electron cryotomography of an entire eukaryotic cell. *PLoS ONE*, 2(1). <https://doi.org/10.1371/journal.pone.0000749>

Herrero A, Flores E, Guerrero MG. (1981) Regulation of nitrate reductase levels in the cyanobacteria *Anacystis nidulans*, *Anabaena* sp. strain 7119, and *Nostoc* sp. strain 6719. *J Bacteriol*, 145(1), 175-80. <https://doi.org/10.1128/jb.145.1.175-180.1981>.

Homma, K., & Hastings, J. W. (1989). The S phase is discrete and is controlled by the circadian clock in the marine dinoflagellate *Gonyaulax polyedra*. *Experimental Cell Research*, 182(2). [https://doi.org/10.1016/0014-4827\(89\)90265-6](https://doi.org/10.1016/0014-4827(89)90265-6)

Hoopes, G. M., Zarka, D., Feke, A., Acheson, K., Hamilton, J. P., Douches, D., Buell, C. R., & Farré, E. M. (2022). Keeping time in the dark: Potato diel and circadian rhythmic gene expression reveals tissue-specific circadian clocks. *Plant Direct*, 6(7). <https://doi.org/10.1002/pld3.425>

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., Moriyama, T., Ikeuchi, M., Watanabe, M., Wada, H., Kobayashi, K., Saito, M., Masuda, T., Sasaki-Sekimoto, Y., Mashiguchi, K., Ohta, H. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms4978>

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Ridwan Amode, M., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*, 49(D1). <https://doi.org/10.1093/nar/gkaa942>

Hoys, C., Romero-Losada, A. B., del Río, E., Guerrero, M. G., Romero-Campero, F. J., & García-González, M. (2021). Unveiling the underlying molecular basis of astaxanthin accumulation in *Haematococcus* through integrative metabolomic-transcriptomic analysis. *Bioresource Technology*, 332. <https://doi.org/10.1016/j.biortech.2021.125150>

Hundahl, C. A., Fahrenkrug, J., Hay-Schmidt, A., Georg, B., Faltoft, B., & Hannibal, J. (2012). Circadian behaviour in neuroglobin deficient mice. *PLoS ONE*, 7(4). <https://doi.org/10.1371/journal.pone.0034462>

Ideker, T., Galitski, T., & Hood, L. (2001). A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics*, 2. <https://doi.org/10.1146/annurev.genom.2.1.343>

Imai, R., Makino, H., Katoh, T., Kimura, T., Kurita, T., Hokamura, K., Umemura, K., & Nakajima, Y. (2020). Desflurane anesthesia shifts the circadian rhythm phase depending on the time of day of anesthesia. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-75434-6>

Jahns, P., Latowski, D., & Strzalka, K. (2009). Mechanism and regulation of the violaxanthin cycle: The role of antenna proteins and membrane lipids. *Biochimica et Biophysica Acta - Bioenergetics*, 1787(1). <https://doi.org/10.1016/j.bbabiobio.2008.09.013>

Jamers, A., Blust, R., & De Coen, W. (2009). Omics in algae: Paving the way for a systems biological understanding of algal stress phenomena?. *Aquatic Toxicology*, 92(3). <https://doi.org/10.1016/j.aquatox.2009.02.012>

Joyce, A. R., & Palsson, B. (2006). The model organism as a system: Integrating “omics” data sets. *Nature Reviews Molecular Cell Biology*, 7(3). <https://doi.org/10.1038/nrm1857>

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>

Karahalil, B. (2016). Overview of systems biology and omics technologies. *Current Medicinal Chemistry*, 23(37), 4221–4230. <https://doi.org/10.2174/0929867323666160926150617>

Kato, S., & Nam, H. G. (2021). The cell division cycle of *Euglena gracilis* indicates that the level of circadian plasticity to the external light regime changes in prolonged-stationary cultures. *Plants*, 10(7). <https://doi.org/10.3390/plants10071475>

Kay, H., Grünwald, E., Feord, H. K., Gil, S., Peak-Chew, S. Y., Stangherlin, A., O'Neill, J. S., & van Ooijen, G. (2021). Deep-coverage spatiotemporal proteome of the picoeukaryote *Ostreococcus tauri* reveals differential effects of environmental

and endogenous 24-hour rhythms. *Communications Biology*, 4(1), 1–11. <https://doi.org/10.1038/s42003-021-02680-3>

Kester, D. R., Duedall, I. W., Connors, D. N., & Pytkowicz, R. M. (1967). Preparation of artificial seawater. *Limnology and Oceanography*, 12(1), 176–179. <https://doi.org/10.4319/LO.1967.12.1.0176>

Khadaroo, B., Robbens, S., Ferraz, C., Derelle, E., Eychenié, S., Cooke, R., Peaucellier, G., Delseny, M., Demaille, J., Van De Peer, Y., Picard, A., & Moreau, H. (2004). The first green lineage cdc25 dual-specificity phosphatase. *Cell Cycle*, 3(4). <https://doi.org/10.4161/cc.3.4.815>

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12, 357–360. <https://doi.org/10.1038/nmeth.3317>

Klante, G., & Steinlechner, S. (1994). Light irradiance and wavelength as seasonal cues for djungarian hamsters. *Biological Rhythm Research*, 25(4). <https://doi.org/10.1080/09291019409360310>

Konopka, R. J., & Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 68(9). <https://doi.org/10.1073/pnas.68.9.2112>

Kötting, O., Kossmann, J., Zeeman, S. C., & Lloyd, J. R. (2010). Regulation of starch metabolism: The age of enlightenment?. *Current Opinion in Plant Biology*, 13(3). <https://doi.org/10.1016/j.pbi.2010.01.003>

Krumholz, E. W., Yang, H., Weisenhorn, P., Henry, C. S., & Libourel, I. G. L. (2012). Genome-wide metabolic network reconstruction of the picoalga *Ostreococcus*. *Journal of Experimental Botany*, 63(6), 2353–2362. <https://doi.org/10.1093/jxb/err407>

Kuhlman, S. J., Craig, L. M., & Duffy, J. F. (2018). Introduction to chronobiology. *Cold Spring Harbor Perspectives in Biology*, 10(9). <https://doi.org/10.1101/cshperspect.a033613>

Kyorku, C., & Brady, J. (1994). A free-running bimodal circadian rhythm in the tsetse fly *Glossina longipennis*. *Journal of Insect Physiology*, 40(1). [https://doi.org/10.1016/0022-1910\(94\)90112-0](https://doi.org/10.1016/0022-1910(94)90112-0)

Laloum, D., & Robinson-Rechavi, M. (2020). Methods detecting rhythmic gene expression are biologically relevant only for strong signal. *PLoS Computational Biology*, 16(3). <https://doi.org/10.1371/journal.pcbi.1007666>

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., & Huala, E. (2012). The *Arabidopsis* Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, 40(D1). <https://doi.org/10.1093/nar/gkr1090>

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4). <https://doi.org/10.1038/nmeth.1923>

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, 9(8). <https://doi.org/10.1371/journal.pcbi.1003118>

Le Bihan, T., Martin, S. F., Chirnside, E. S., van Ooijen, G., Barrios-Llerena, M. E., O'Neill, J. S., Shliaha, P. V., Kerr, L. E., & Millar, A. J. (2011). Shotgun proteomic analysis of the unicellular alga *Ostreococcus tauri*. *Journal of Proteomics*, 74(10). <https://doi.org/10.1016/j.jprot.2011.05.028>

Lebert, M., Porst, M., & Häder, D. P. (1999). Circadian rhythm of gravitaxis in *Euglena gracilis*. *Journal of Plant Physiology*, 155(3). [https://doi.org/10.1016/S0176-1617\(99\)80115-1](https://doi.org/10.1016/S0176-1617(99)80115-1)

Leconte, J., Benites, L. F., Vannier, T., Wincker, P., Piganeau, G., & Jaillon, O. (2020). Genome resolved biogeography of mamiellales. *Genes*, 11(1). <https://doi.org/10.3390/genes11010066>

Lelandais, G., Scheiber, I., Paz-Yepes, J., Lozano, J.-C., Botebol, H., Pilátová, J., Žárský, V., Léger, T., Blaiseau, P.-L., Bowler, C., Bouget, F.-Y., Camadro, J.-M., Sutak, R., & Lesuisse, E. (2016). *Ostreococcus tauri* is a new model green alga for studying iron metabolism in eukaryotic phytoplankton. *BMC Genomics*, 17(1), 319. <https://doi.org/10.1186/s12864-016-2666-6>

Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., & De Clerck, O. (2012). Phylogeny and molecular evolution of the green algae. *Critical Reviews in Plant Sciences*, 31(1). <https://doi.org/10.1080/07352689.2011.615705>

Leliaert, F., Verbruggen, H., & Zechman, F. W. (2011). Into the deep: New discoveries at the base of the green plant phylogeny. *BioEssays*, 33(9). <https://doi.org/10.1002/bies.201100035>

Lillo, C., Meyer, C., & Ruoff, P. (2001). The nitrate reductase circadian system. The central clock dogma contra multiple oscillatory feedback loops. *Plant Physiology*, 125(4). <https://doi.org/10.1104/pp.125.4.1554>

Lillo, C., & Ruoff, P. (1989). An unusually rapid light-induced nitrate reductase mRNA pulse and circadian oscillations. *Naturwissenschaften*, 76(11). <https://doi.org/10.1007/BF00374129>

Lonergan, T. A. (1981). A circadian rhythm in the rate of light-induced electron flow in three leguminous species. *Plant Physiology*, 68(5). <https://doi.org/10.1104/pp.68.5.1041>

López, D., Casero, D., Cookus, S. J., Merchant, S. S., & Pellegrini, M. (2011). Algal Functional Annotation Tool: A web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics*, 12(282). <https://doi.org/10.1186/1471-2105-12-282>

Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B. C., & Aebersold, R. (2018). Data-independent acquisition-based SWATH - MS for quantitative proteomics: a tutorial. *Molecular Systems Biology*, 14(8). <https://doi.org/10.15252/msb.20178126>

Ma, D., Przybylski, D., Abruzzi, K. C., Schlichting, M., Li, Q., Long, X., & Rosbash, M. (2021). A transcriptomic taxonomy of *Drosophila* circadian neurons around the clock. *ELife*, 10. <https://doi.org/10.7554/eLife.63056>

Mack, G. A., & Wolfe, D. A. (1981). K-sample rank tests for umbrella alternatives. *Journal of the American Statistical Association*, 76(373). <https://doi.org/10.1080/01621459.1981.10477625>

Mackenzie, T. D. B., & Morse, D. (2011). Circadian photosynthetic reductant flow in the dinoflagellate *Lingulodinium* is limited by carbon availability. *Plant, Cell and Environment*, 34(4). <https://doi.org/10.1111/j.1365-3040.2010.02271.x>

Masseroli, M., Martucci, D., & Pincioli, F. (2004). GFINDER: Genome Function INtegrated Discoverer through dynamic annotation, statistical analysis, and mining. *Nucleic Acids Research*, 32. <https://doi.org/10.1093/nar/gkh432>

Matsuo, T., Yamaguchi, S., Mitsui, S., Emi, A., Shimoda, F., & Okamura, H. (2003). Control mechanism of the circadian clock for timing of cell division in vivo. *Science*, 302(5643). <https://doi.org/10.1126/science.1086271>

Mazzocchi, F. (2012). Complexity and the reductionism-holism debate in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 4(5), 413–427. <https://doi.org/10.1002/wsbm.1181>

Mazzoccoli, G., Vinciguerra, M., Carbone, A., & Relógio, A. (2020). The circadian clock, the immune system, and viral infections: The intricate relationship between biological time and host-virus interaction. *Pathogens*, 9(2). <https://doi.org/10.3390/pathogens9020083>

McClung, C. R. (2006). Plant circadian rhythms. *Plant Cell*, 18(4). <https://doi.org/10.1105/tpc.106.040980>

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., Marshall, W. F., Qu, L. H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., Zhou, K. (2007). The *Chlamydomonas*

genome reveals the evolution of key animal and plant functions. *Science*, 318(5848). <https://doi.org/10.1126/science.1143609>

Merchante, C., Stepanova, A. N., & Alonso, J. M. (2017). Translation regulation in plants: an interesting past, an exciting present and a promising future. *Plant Journal*, 90(4). <https://doi.org/10.1111/tpj.13520>

Mermet, J., Yeung, J., & Naef, F. (2017). Systems chronobiology: Global analysis of gene regulation in a 24-hour periodic world. *Cold Spring Harbor Perspectives in Biology*, 9(3). <https://doi.org/10.1101/cshperspect.a028720>

Merrow, M., Spoelstra, K., & Roenneberg, T. (2005). The circadian cycle: Daily rhythms from behaviour to genes. *EMBO Reports*, 6(10). <https://doi.org/10.1038/sj.embor.7400541>

Messager, S., Hazlerigg, D. G., Mercer, J. G., & Morgan, P. J. (2000). Photoperiod differentially regulates the expression of Per1 and ICER in the pars tuberalis and the suprachiasmatic nucleus of the Siberian hamster. *European Journal of Neuroscience*, 12(8). <https://doi.org/10.1046/j.1460-9568.2000.00174.x>

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L. P., Mushayamaha, T., & Thomas, P. D. (2021). PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research*, 49(D1). <https://doi.org/10.1093/nar/gkaa1106>

Miller, B. H., McDearmon, E. L., Panda, S., Hayes, K. R., Zhang, J., Andrews, J. L., Antoch, M. P., Walker, J. R., Esser, K. A., Hogenesch, J. B., & Takahashi, J. S. (2007). Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(9). <https://doi.org/10.1073/pnas.0611724104>

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1). <https://doi.org/10.1093/nar/gkaa913>

Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S., & Arai, T. (1986). Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature*, 323(6090). <https://doi.org/10.1038/323720a0>

Mittag, M. (2001). Circadian rhythms in microalgae. *International Review of Cytology*, 206. [https://doi.org/10.1016/S0074-7696\(01\)06023-5](https://doi.org/10.1016/S0074-7696(01)06023-5)

Mohabir, G., & Edmunds, L. N. (1999). Circadian clock regulation of the bimodal rhythm of cyclic AMP in wild-type *Euglena*. *Cellular Signalling*, 11(2). [https://doi.org/10.1016/S0898-6568\(98\)00046-1](https://doi.org/10.1016/S0898-6568(98)00046-1)

Monnier, A., Liverani, S., Bouvet, R., Jesson, B., Smith, J. Q., Mosser, J., Corellou, F., & Bouget, F. Y. (2010). Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. *BMC Genomics*, 11(1). <https://doi.org/10.1186/1471-2164-11-192>

Moreau, H., Grimsley, N., Derelle, E., Ferraz, C., Escande, M.L., Eychenié, S., Cooke, R., Piganeau, G., Desdevises, Y., Bellec, L. (1995). A new marine picoeucaryote: *Ostreococcus tauri* gen. et sp. nov. (*Chlorophyta, Prasinophyceae*). 34(4), 285–292. <https://doi.org/10.2216/i0031-8884-34-4-285.1>

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M. F., Piganeau, G., Rouzé, P., Da Silva, C., Wincker, P., Van de Peer, Y., & Vandepoele, K. (2012). Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biology*, 13(8). <https://doi.org/10.1186/gb-2012-13-8-r74>

Morimoto, D., Yoshida, T., & Sawayama, S. (2020). Draft genome sequence of the astaxanthin-producing microalga *Haematococcus lacustris* Strain NIES-144. *Microbiology Resource Announcements*, 9(23). <https://doi.org/10.1128/mra.00128-20>

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35(2), 182–185. <https://doi.org/10.1093/nar/gkm321>

Morris, A. R., Stanton, D. L., Roman, D., & Liu, A. C. (2020). Systems level understanding of circadian integration with cell physiology. *Journal of Molecular Biology*, 432(12). <https://doi.org/10.1016/j.jmb.2020.02.002>

Moulager, M., Corellou, F., Vergé, V., Escande, M. L., & Bouget, F. Y. (2010). Integration of light signals by the retinoblastoma pathway in the control of S phase entry in the picophytoplanktonic cell *Ostreococcus*. *PLoS Genetics*, 6(5), 11. <https://doi.org/10.1371/journal.pgen.1000957>

Moulager, M., Monnier, A., Jesson, B., Bouvet, R., Mosser, J., Schwartz, C., Garnier, L., Corellou, F., & Bouget, F. Y. (2007). Light-dependent regulation of cell division in *Ostreococcus*: Evidence for a major transcriptional input. *Plant Physiology*, 144(3), 1360–1369. <https://doi.org/10.1104/pp.107.096149>

Nelson, C. J., Alexova, R., Jacoby, R. P., & Harvey Millar, A. (2014). Proteins with high turnover rate in barley leaves estimated by proteome analysis combined with in planta isotope labeling. *Plant Physiology*, 166(1). <https://doi.org/10.1104/pp.114.243014>

Nelson, C. J., & Millar, A. H. (2015). Protein turnover in plant biology. *Nature Plants*, 1. <https://doi.org/10.1038/nplants.2015.17>

Ngan, C. Y., Wong, C. H., Choi, C., Yoshinaga, Y., Louie, K., Jia, J., Chen, C., Bowen, B., Cheng, H., Leonelli, L., Kuo, R., Baran, R., Garcíá-Cerdán, J. G., Pratap, A., Wang, M., Lim, J., Tice, H., Daum, C., Xu, J., Wei, C. L. (2015). Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae. *Nature Plants*, 1(15107). <https://doi.org/10.1038/nplants.2015.107>

Nishiwaki-Ohkawa, T., & Yoshimura, T. (2016). Molecular basis for regulating seasonal reproduction in vertebrates. *Journal of Endocrinology*, 229(3). <https://doi.org/10.1530/JOE-16-0066>

Noordally, Z. B., & Millar, A. J. (2015). Clocks in algae. *Biochemistry*, 54(2), 171–183. <https://doi.org/10.1021/bi501089x>

Numata, H., Miyazaki, Y., & Ikeno, T. (2015). Common features in diverse insect clocks. *Zoological Letters*, 1(1). <https://doi.org/10.1186/s40851-014-0003-y>

O'Kelly, C. J., Sieracki, M. E., Thier, E. C., & Hobson, I. C. (2003). A transient bloom of *Ostreococcus* (*Chlorophyta, Prasinophyceae*) in West Neck Bay, Long Island, New York. *Journal of Phycology*, 39(5). <https://doi.org/10.1046/j.1529-8817.2003.02201.x>

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/27.1.29>

Ohara, T., Fukuda, H., & Tokuda, I. T. (2015). An extended mathematical model for reproducing the phase response of *Arabidopsis thaliana* under various light conditions. *Journal of Theoretical Biology*, 382. <https://doi.org/10.1016/j.jtbi.2015.07.016>

Ottesen, E. A., Young, C. R., Eppley, J. M., Ryan, J. P., Chavez, F. P., Scholin, C. A., & DeLong, E. F. (2013). Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proceedings of the National Academy of Sciences of the United States of America*, 110(6). <https://doi.org/10.1073/pnas.1222099110>

Paajanen, P., Lane de Barros Dantas, L., & Dodd, A. N. (2021). Layers of crosstalk between circadian regulation and environmental signalling in plants. *Current Biology*, 31(8). <https://doi.org/10.1016/j.cub.2021.03.046>

Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E. E., McCarren, J., Paulsen, I., Dufresne, A., Partensky, F., Webb, E. A., & Waterbury, J. (2003). The genome of a motile marine *Synechococcus*. *Nature*, 424(6952). <https://doi.org/10.1038/nature01943>

Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K., Otellar, R., Merchant, S. S., Podell, S., Gaasterland, T., Napoli, C., Gandler, K., Manuell, A., Tai, V., Grigoriev, I. V. (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18). <https://doi.org/10.1073/pnas.0611046104>

Pan, Y., Ballance, H., Meng, H., Gonzalez, N., Kim, S. M., Abdurehman, L., York, B., Chen, X., Schnytzer, Y., Levy, O., Dacso, C. C., McClung, C. A., O’Malley, B. W., Liu, S., & Zhu, B. (2020). 12-h clock regulation of genetic information flow by XBP1s. *PLoS Biology*, 18(1). <https://doi.org/10.1371/journal.pbio.3000580>

Pan, Y., Michael, T. P., Hudson, M. E., Kay, S. A., Chory, J., & Schuler, M. A. (2009). Cytochrome P450 monooxygenases as reporters for circadian-regulated pathways. *Plant Physiology*, 150(2). <https://doi.org/10.1104/pp.108.130757>

Panter, P. E., Muranaka, T., Cuitun-Coronado, D., Graham, C. A., Yochikawa, A., Kudoh, H., & Dodd, A. N. (2019). Circadian regulation of the plant transcriptome under natural conditions. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.01239>

Parsons, R., Parsons, R., Garner, N., Oster, H., & Rawashdeh, O. (2020). CircaCompare: A method to estimate and statistically support differences in mesor, amplitude and phase, between circadian rhythms. *Bioinformatics*, 36(4). <https://doi.org/10.1093/bioinformatics/btz730>

Pereira, H., Sá, M., Maia, I., Rodrigues, A., Teles, I., Wijffels, R. H., Navalho, J., & Barbosa, M. (2021). Fucoxanthin production from *Tisochrysis lutea* and *Phaeodactylum tricornutum* at industrial scale. *Algal Research*, 56. <https://doi.org/10.1016/j.algal.2021.102322>

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., & Vizcaíno, J. A. (2022). The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research*, 50(D1). <https://doi.org/10.1093/nar/gkab1038>

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, 11(9). <https://doi.org/10.1038/nprot.2016.095>

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of

a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33, 290–295. <https://doi.org/10.1038/nbt.3122>

Pfeuty, B., Thommen, Q., Corellou, F., Djouani-Tahri, E. B., Bouget, F. Y., & Lefranc, M. (2012). Circadian clocks in changing weather and seasons: Lessons from the picoalga *Ostreococcus tauri*. *BioEssays*, 34(9), 781–790. <https://doi.org/10.1002/bies.201200012>

Pilgrim, M. L., & McClung, C. R. (1993). Differential involvement of the circadian clock in the expression of genes required for ribulose-1,5-bisphosphate carboxylase/oxygenase synthesis, assembly, and activation in *Arabidopsis thaliana*. *Plant Physiology*, 103(2), 553–564. <https://doi.org/10.1104/pp.103.2.553>

Pittendrigh, C. S. (1960). Circadian rhythms and the circadian organization of living systems. *Cold Spring Harbor Symposia on Quantitative Biology*, 25. <https://doi.org/10.1101/SQB.1960.025.01.015>

Polle, J. E. W., Barry, K., Cushman, J., Schmutz, J., Tran, D., Hathwaik, L. T., Yim, W. C., Jenkins, J., McKie-Krisberg, Z., Prochnik, S., Lindquist, E., Dockter, R. B., Adam, C., Molina, H., Bunkenborg, J., Jin, E. S., Buchheim, M., & Magnuson, J. (2017). Draft nuclear genome sequence of the halophilic and beta-carotene-accumulating green alga *Dunaliella salina* strain CCAP19/18. *Genome Announcements*, 5(43). <https://doi.org/10.1128/genomeA.01105-17>

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, 46(1). <https://doi.org/10.1093/nar/gky448>

Prabhakaran, P. M., & Sheeba, V. (2012). Sympatric drosophilid species melanogaster and ananassae differ in temporal patterns of activity. *Journal of Biological Rhythms*, 27(5). <https://doi.org/10.1177/0748730412458661>

Prendergast, S. C., Strobl, A. C., Cross, W., Pillay, N., Strauss, S. J., Ye, H., Lindsay, D., Tirabosco, R., Chalker, J., Mahamdallie, S. S., Sosinsky, A., Flanagan, A. M., & Amary, F. (2020). Sarcoma and the 100,000 Genomes Project: our experience and changes to practice. *Journal of Pathology: Clinical Research*, 6(4). <https://doi.org/10.1002/cjp2.174>

Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L. K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S. A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A., Rokhsar, D. S. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*, 329(5988). <https://doi.org/10.1126/science.1188800>

Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., Settlage, R. E., Boore, J. L., & Posewitz, M. C. (2012). Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nature Communications*, 3. <https://doi.org/10.1038/ncomms1688>

Ral, J. P., Colleoni, C., Wattebled, F., Dauvillée, D., Nempong, C., Deschamps, P., Li, Z., Morell, M. K., Chibbar, R., Purton, S., D'Hulst, C., & Ball, S. G. (2006). Circadian clock regulation of starch metabolism establishes GBSSI as a major contributor to amylopectin synthesis in *Chlamydomonas reinhardtii*. *Plant Physiology*, 142(1). <https://doi.org/10.1104/pp.106.081885>

Rayko, E., Maumus, F., Maheswari, U., Jabbari, K., & Bowler, C. (2010). Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytologist*, 188(1). <https://doi.org/10.1111/j.1469-8137.2010.03371.x>

Richter, C. P. (1922). A behavioristic study of the activity of the rat. *Comparative Psychology Monographs*, 1, 2-56. <https://doi.org/10.5962/bhl.title.151527>

Ripperger, J. A., & Merrow, M. (2011). Perfect timing: Epigenetic regulation of the circadian clock. *FEBS Letters*, 585(10). <https://doi.org/10.1016/j.febslet.2011.04.047>

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7). <https://doi.org/10.1093/nar/gkv007>

Robbens, S., Khadaroo, B., Camasses, A., Derelle, E., Ferraz, C., Inzé, D., Van De Peer, Y., & Moreau, H. (2005). Genome-wide analysis of core cell cycle genes in the unicellular green alga *Ostreococcus tauri*. *Molecular Biology and Evolution*, 22(3). <https://doi.org/10.1093/molbev/msi044>

Robles, M. S., Cox, J., & Mann, M. (2014). In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genetics*, 10(1). <https://doi.org/10.1371/journal.pgen.1004047>

Rock, A., Wilcockson, D., & Last, K. S. (2022). Towards an understanding of circatidal clocks. *Frontiers in Physiology*, 13. <https://doi.org/10.3389/fphys.2022.830107>

Roenneberg, T., Foster, R. G., & Klerman, E. B. (2022). The circadian system, sleep, and the health/disease balance: a conceptual review. *Journal of Sleep Research*, 31(4). <https://doi.org/10.1111/jsr.13621>

Roenneberg, T., & Merrow, M. (2005). Circadian clocks - The fall and rise of physiology. *Nature Reviews Molecular Cell Biology*, 6(12). <https://doi.org/10.1038/nrm1766>

Roenneberg, T., & Merrow, M. (2016). The circadian clock and human health. *Current Biology*, 26(10). <https://doi.org/10.1016/j.cub.2016.04.011>

Roenneberg, T., Pilz, L. K., Zerbini, G., & Winnebeck, E. C. (2019). Chronotype and social jetlag: A self-critical review. *Biology*, 8(3). <https://doi.org/10.3390/biology8030054>

Romero-Campero, F. J., Perez-Hurtado, I., Lucas-Reina, E., Romero, J. M., & Valverde, F. (2016). ChlamyNET: A *Chlamydomonas* gene co-expression network reveals global properties of the transcriptome and the early setup of key co-expression patterns in the green lineage. *BMC Genomics*, 17(227). <https://doi.org/10.1186/s12864-016-2564-y>

Romero-Losada, A. B., Arvanitidou, C., de los Reyes, P., García-González, M., & Romero-Campero, F. J. (2022). ALGAEFUN with MARACAS, microALGAE FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq AnalysiS. *BMC Bioinformatics*, 23(1). <https://doi.org/10.1186/s12859-022-04639-5>

Roth, M. S., Cokus, S. J., Gallaher, S. D., Walter, A., Lopez, D., Erickson, E., Endelman, B., Westcott, D., Larabell, C. A., Merchant, S. S., Pellegrini, M., & Niyogi, K. K. (2017). Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. *Proceedings of the National Academy of Sciences of the United States of America*, 114(21). <https://doi.org/10.1073/pnas.1619928114>

Rufty, T. W., & Huber, S. C. (1983). Changes in Starch Formation and Activities of Sucrose Phosphate Synthase and Cytoplasmic Fructose-1,6-bisphosphatase in Response to Source-Sink Alterations. *Plant Physiology*, 72(2). <https://doi.org/10.1104/pp.72.2.474>

Sanz-Luque, E., Chamizo-Ampudia, A., Llamas, A., Galvan, A., & Fernandez, E. (2015). Understanding nitrate assimilation and its regulation in microalgae. *Frontiers in Plant Science*, 6. <https://doi.org/10.3389/fpls.2015.00899>

Sekimoto, H. (2017). Sexual reproduction and sex determination in green algae. *Journal of Plant Research*, 130(3). <https://doi.org/10.1007/s10265-017-0908-6>

Serrano-Bueno, G., Romero-Campero, F. J., Lucas-Reina, E., Romero, J. M., & Valverde, F. (2017). Evolution of photoperiod sensing in plants and algae. *Current Opinion in Plant Biology*, 37, 10–17. <https://doi.org/10.1016/j.pbi.2017.03.007>

Serrano-Pérez, E., Romero-Losada, A. B., Morales-Pineda, M., García-Gómez, M. E., Couso, I., García-González, M., & Romero-Campero, F. J. (2022). Transcriptomic and metabolomic response to high light in the charophyte alga *Klebsormidium nitens*. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.855243>

Serrano, G., Herrera-Palau, R., Romero, J. M., Serrano, A., Coupland, G., & Valverde, F. (2009). *Chlamydomonas CONSTANS* and the evolution of plant photoperiodic signaling. *Current Biology*, 19(5). <https://doi.org/10.1016/j.cub.2009.01.044>

Sharma, A., Tripathi, V., & Kumar, V. (2022). Control and adaptability of seasonal changes in behavior and physiology of latitudinal avian migrants: Insights from laboratory studies in Palearctic-Indian migratory buntings. *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, 337(9-10), 902-918. <https://doi.org/10.1002/jez.2631>

Shen, M., Chang, Y. T., Wu, C. T., Parker, S. J., Saylor, G., Wang, Y., Yu, G., Van Eyk, J. E., Clarke, R., Herrington, D. M., & Wang, Y. (2022). Comparative assessment and novel strategy on methods for imputing proteomics data. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-04938-0>

Sierra, E., Acién, F. G., Fernández, J. M., García, J. L., González, C., & Molina, E. (2008). Characterization of a flat plate photobioreactor for the production of microalgae. *Chemical Engineering Journal*, 138(1–3). <https://doi.org/10.1016/j.cej.2007.06.004>

Six, C., Sherrard, R., Lionard, M., Roy, S., & Campbell, D. A. (2009). Photosystem II and Pigment Dynamics among ecotypes of the green alga *Ostreococcus*. *Plant Physiology*, 151(1). <https://doi.org/10.1104/pp.109.140566>

Six, C., Worden, A. Z., Rodríguez, F., Moreau, H., & Partensky, F. (2005). New insights into the nature and phylogeny of prasinophyte antenna proteins: *Ostreococcus tauri*, a case study. *Molecular Biology and Evolution*, 22(11), 2217–2230. <https://doi.org/10.1093/molbev/msi220>

Smith, S. M., Fulton, D. C., Chia, T., Thorneycroft, D., Chapple, A., Dunstan, H., Hylton, C., Zeeman, S. C., & Smith, A. M. (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 Physiology*, 136(1). <https://doi.org/10.1104/pp.104.044347>

Somers, D. E., Webb, A. A. R., Pearson, M., & Kay, S. A. (1998). The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in *Arabidopsis thaliana*. *Development*, 125(3). <https://doi.org/10.1242/dev.125.3.485>

Sorek, M., Yacobi, Y. Z., Roopin, M., Berman-Frank, I., & Levy, O. (2013). Photosynthetic circadian rhythmicity patterns of *Symbiodium*, the coral endosymbiotic algae. *Proceedings of the Royal Society B: Biological Sciences*, 280(1759). <https://doi.org/10.1098/rspb.2012.2942>

Sorokina, O., Corellou, F., Dauvillée, D., Sorokin, A., Goryanin, I., Ball, S., Bouget, F. Y., & Millar, A. J. (2011). Microarray data can predict diurnal changes of starch content in the picoalga *Ostreococcus*. *BMC Systems Biology*, 5(1), 36. <https://doi.org/10.1186/1752-0509-5-36>

Spudich, J. L., & Sager, R. (1980). Regulation of the chlamydomonas cell cycle by light and dark. *Journal of Cell Biology*, 85(1). <https://doi.org/10.1083/jcb.85.1.136>

Sulpice, R., Flis, A., Ivakov, A. A., Apelt, F., Krohn, N., Encke, B., Abel, C., Feil, R., Lunn, J. E., & Stitt, M. (2014). *Arabidopsis* coordinates the diurnal regulation of carbon allocation and growth across a wide range of photoperiods. *Molecular Plant*, 7(1). <https://doi.org/10.1093/mp/sst127>

Sumová, A., Jáč, M., Sládek, M., Šauman, I., & Illnerová, H. (2003). Clock gene daily profiles and their phase relationship in the rat suprachiasmatic nucleus are affected by photoperiod. *Journal of Biological Rhythms*, 18(2). <https://doi.org/10.1177/0748730403251801>

Sun, T. H., Liu, C. Q., Hui, Y. Y., Wu, W. K., Zhou, Z. G., & Lu, S. (2010). Coordinated Regulation of Gene Expression for Carotenoid Metabolism in *Chlamydomonas reinhardtii*. *Journal of Integrative Plant Biology*, 52(10). <https://doi.org/10.1111/j.1744-7909.2010.00993.x>

Sun, T., Rao, S., Zhou, X., & Li, L. (2022). Plant carotenoids: recent advances and future perspectives. *Molecular Horticulture*, 2(1). <https://doi.org/10.1186/s43897-022-00023-2>

Swanson, W. J., Aagaard, J. E., Vacquier, V. D., Monné, M., Sadat Al Hosseini, H., & Jovine, L. (2011). The molecular basis of sex: Linking yeast to human. *Molecular Biology and Evolution*, 28(7). <https://doi.org/10.1093/molbev/msr026>

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., & Huala, E. (2008). The *Arabidopsis* Information Resource (TAIR): Gene structure and function annotation. *Nucleic Acids Research*, 36(1). <https://doi.org/10.1093/nar/gkm965>

Sweeney, B. M., & Haxo, F. T. (1961). Persistence of a photosynthetic rhythm in enucleated *Acetabularia*. *Science*, 134(3487). <https://doi.org/10.1126/science.134.3487.1361>

Takahashi, J. S. (2021). The 50th anniversary of the konopka and benzer 1971 paper in PNAS: “Clock mutants of *Drosophila melanogaster*”. *Proceedings of the National Academy of Sciences of the United States of America*, 118(39). <https://doi.org/10.1073/pnas.2110171118>

Thaben, P. F., & Westermark, P. O. (2014). Detecting rhythms in time series with rain. *Journal of Biological Rhythms*, 29(6). <https://doi.org/10.1177/0748730414553029>

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., & Su, Z. (2017). AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, 45(W1), W122–W129. <https://doi.org/10.1093/nar/gkx382>

Tragin, M., & Vaulot, D. (2019). Novel diversity within marine *Mamiellophyceae* (*Chlorophyta*) unveiled by metabarcoding. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-41680-6>

Tucker, D. E., Allen, D. J., & Ort, D. R. (2004). Control of nitrate reductase by circadian and diurnal rhythms in tomato. *Planta*, 219(2). <https://doi.org/10.1007/s00425-004-1213-x>

van Dongen, H. P. A., Kerkhof, G. A., & Klöppel, H. B. (1997). Seasonal covariation of the circadian phases of rectal temperature and slow wave sleep onset. *Journal of Sleep Research*, 6(1). <https://doi.org/10.1046/j.1365-2869.1997.00021.x>

van Gelderen, K. (2020). The rhythm of the light: How light and the clock drive cycling of transcript levels in barley. *Plant Physiology*, 183, (2). <https://doi.org/10.1104/pp.20.00360>

Vatakis, A., Balci, F., Di Luca, M., & Correa, Á. (2018). Circadian timing: from genetics to behavior. *Timing and Time Perception: Procedures, Measures, & Applications*, 1-31. [https://doi.org/10.1163/9789004280205\\_002](https://doi.org/10.1163/9789004280205_002)

Veenstra, T. D. (2021). Omics in systems biology: current progress and future outlook. *Proteomics*, 21(3–4). <https://doi.org/10.1002/pmic.202000235>

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1). <https://doi.org/10.1038/nrg2484>

Watanabe, T., Naito, E., Nakao, N., Tei, H., Yoshimura, T., & Ebihara, S. (2007). Bimodal clock gene expression in mouse suprachiasmatic nucleus and peripheral tissues under a 7-hour light and 5-hour dark schedule. *Journal of Biological Rhythms*, 22(1). <https://doi.org/10.1177/0748730406295435>

Watson, J. V., Chambers, S. H., & Smith, P. J. (1987). A pragmatic approach to the analysis of DNA histograms with a definable G1 peak. *Cytometry*, 8(1), 1–8. <https://doi.org/10.1002/CYTO.990080101>

Weckwerth, W. (2011). Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *Journal of Proteomics*, 75(1). <https://doi.org/10.1016/j.jprot.2011.07.010>

Wenden, B., Toner, D. L. K., Hodge, S. K., Grima, R., & Millar, A. J. (2012). Spontaneous spatiotemporal waves of gene expression from biological clocks in the leaf. *Proceedings of the National Academy of Sciences of the United States of America*, 109(17). <https://doi.org/10.1073/pnas.1118814109>

Willforss, J., Chawade, A., & Levander, F. (2019). NormalizerDE: online tool for improved normalization of omics expression data and high-sensitivity differential expression analysis. *Journal of Proteome Research*, 18(2). <https://doi.org/10.1021/acs.jproteome.8b00523>

Worden, A. Z., Lee, J. H., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L., Allen, A. E., Cuvelier, M. L., Derelle, E., Everett, M. V., Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S. M., Parker, M. S., Rombauts, S., Salamov, A., Grigoriev, I. V. (2009). Green evolution and dynamic adaptations

revealed by genomes of the marine picoeukaryotes micromonas. *Science*, 324(5924). <https://doi.org/10.1126/science.1167222>

Worden, A. Z., Nolan, J. K., & Palenik, B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. *Limnology and Oceanography*, 49(1). <https://doi.org/10.4319/lo.2004.49.1.0168>

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3). <https://doi.org/10.1016/j.xinn.2021.100141>

Wucher, V., Sodaei, R., Amador, R., Irimia, M., & Guigó, R. (2022). Day-night and seasonal variation of human gene expression across tissues. *BioRxiv*. <https://doi.org/10.1101/2021.02.28.433266>

Yakir, E., Hassidim, M., Melamed-Book, N., Hilman, D., Kron, I., & Green, R. M. (2011). Cell autonomous and cell-type specific circadian rhythms in *Arabidopsis*. *Plant Journal*, 68(3). <https://doi.org/10.1111/j.1365-313X.2011.04707.x>

Yang, M., Lin, X., Liu, X., Zhang, J., & Ge, F. (2018). Genome annotation of a model diatom *Phaeodactylum tricornutum* using an integrated proteogenomic pipeline. *Molecular Plant*, 11(10). <https://doi.org/10.1016/j.molp.2018.08.005>

Yang, Z., & Midmore, D. J. (2005). A model for the circadian oscillations in expression and activity of nitrate reductase in higher plants. *Annals of Botany*, 96(6). <https://doi.org/10.1093/aob/mci254>

Youthed, G. J., & Moran, V. C. (1969). The lunar-day activity rhythm of *Myrmeleontid* larvae. *Journal of Insect Physiology*, 15(7). [https://doi.org/10.1016/0022-1910\(69\)90235-2](https://doi.org/10.1016/0022-1910(69)90235-2)

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : A Journal of Integrative Biology*, 16(5). <https://doi.org/10.1089/omi.2011.0118>

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382–2383. <https://doi.org/10.1093/bioinformatics/btv145>

Zee, P. C., & Abbott, S. M. (2020). Circadian rhythm sleep-wake disorders. *Psychiatric Clinics of North America*, 38(4), 805-823. <https://doi.org/10.1016/j.psc.2015.07.012>

Zhang, Z., Han, T., Sui, J., & Wang, H. (2022). Cryptochrome-mediated blue-light signal contributes to carotenoids biosynthesis in microalgae. *Frontiers in Microbiology*, 13(1083387). <https://doi.org/10.3389/fmicb.2022.1083387>

Zhao, L., Chang, W., Xiao, Y., Liu, H., & Liu, P. (2013). Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annual Review of Biochemistry*, 82, 497–530. <https://doi.org/10.1146/annurev-biochem-052010-100934>

Zhao, X., Rastogi, A., Deton Cabanillas, A. F., Ait Mohamed, O., Cantrel, C., Lombard, B., Murik, O., Genovesio, A., Bowler, C., Bouyer, D., Loew, D., Lin, X., Veluchamy, A., Vieira, F. R. J., & Tirichine, L. (2021). Genome wide natural variation of H3K27me3 selectively marks genes predicted to be important for cell differentiation in *Phaeodactylum tricornutum*. *New Phytologist*, 229(6). <https://doi.org/10.1111/nph.17129>

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., Banf, M., Dai, X., Martin, G. B., Giovannoni, J. J., Zhao, P. X., Rhee, S. Y., & Fei, Z. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular Plant*, 9(12). <https://doi.org/10.1016/j.molp.2016.09.014>

Zhu, B., Zhang, Q., Pan, Y., Mace, E. M., York, B., Antoulas, A. C., Dacso, C. C., & O’Malley, B. W. (2017). A cell-autonomous mammalian 12 hr clock coordinates metabolic and stress rhythms. *Cell Metabolism*, 25(6). <https://doi.org/10.1016/j.cmet.2017.05.004>

Zhu, L. J. (2013). Integrative analysis of ChIP-chip and ChIP-seq dataset. *Methods in Molecular Biology*, 1067. [https://doi.org/10.1007/978-1-62703-607-8\\_8](https://doi.org/10.1007/978-1-62703-607-8_8)

Zones, J. M., Blaby, I. K., Merchant, S. S., & Umen, J. G. (2015). High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. *Plant Cell*, 27(10). <https://doi.org/10.1105/tpc.15.00498>

Zurbriggen, M. D., Moor, A., & Weber, W. (2012). Plant and bacterial systems biology as platform for plant synthetic biotechnology. *Journal of Biotechnology*, 160(1–2). <https://doi.org/10.1016/j.jbiotec.2012.01.014>





# **Publications**



Publications directly related to this thesis as first author or co-author:

† These authors have contributed equally to this work and share first authorship.

Hoys, C. †, Romero-Losada, A.B. †, del Río, E., Guerrero, M. G., Romero-Campero, F. J., & García-González, M. (2021). Unveiling the underlying molecular basis of astaxanthin accumulation in *Haematococcus* through integrative metabolomic-transcriptomic analysis. *Bioresource Technology*, 332. <https://doi.org/10.1016/j.biortech.2021.125150>

Romero-Losada, A.B., Arvanitidou, C., de los Reyes, P., García-González, M., & Romero-Campero, F. J. (2022). ALGAEFUN with MARACAS, microALGAE FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq AnalysiS. *BMC Bioinformatics*, 23(1). <https://doi.org/10.1186/s12859-022-04639-5>

Serrano-Pérez E.† , Romero-Losada A.B.† , Morales-Pineda M., García-Gómez M.E., Couso I., García-González M. and Romero-Campero F.J. (2022) Transcriptomic and metabolomic response to high light in the charophyte alga *Klebsormidium nitens*. *Frontiers in Plant Science*, 13(855243). <https://doi.org/10.3389/fpls.2022.855243>

Romero-Losada A.B., Arvanitidou C., García-Gómez E., Morales-Pineda M., Castro-Pérez M., García-González M., Romero-Campero F.J. (2023) Multiomics responses to seasonal variations in diel cycles in the marine phytoplanktonic picoeukaryote *Ostreococcus tauri*. Under review in *Plant cell*.

Other related publications:

Santamaría-Gómez J., Rubio M.Á., López-Igual R., **Romero-Losada A.B.**, Delgado-Chaves F.M., Bru-Martínez R., Romero-Campero F.J., Herrero A., Ibba M., Ochoa de Alda J.A.G., Luque I. (2021) Role of a cryptic tRNA gene operon in survival under translational stress. *Nucleic Acids Research*, 49(15), 8757-8776. <https://doi.org/10.1093/nar/gkab661>

Liu C., Mentzelopoulou A., Hatzianestis I.H., Tzagkarakis E., Skaltsogiannis V., Ma X., Michalopoulou V.A., Romero-Campero F.J., **Romero-Losada A.B.**, Sarris P.F., Marhavy P., Bölter B., Kanterakis A., Gutierrez-Beltran E., Moschou P.N. (2023)

A proxitome-RNA-capture approach reveals that processing bodies repress co-regulated hub genes. *Plant Cell*, 16(288). <https://doi.org/10.1093/plcell/koad288>

Arvanitidou C. †, Ramos-González M. †, **Romero-Losada A.B.** †, García-Gómez M. E., García-González M., Romero-Campero F.J. (2023) Transcriptomic characterization of the response to a microalgae extract in *Arabidopsis thaliana* and *Solanum lycopersicum*. Under review in *Journal of the Science of Food and Agriculture*.





# **Appendix**



**Appendix 1. Number of reads and percentage of mapped reads over the nuclear genome of each sample.**

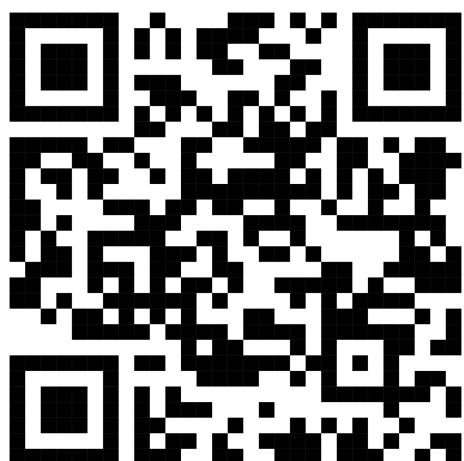
Sample Name	Number of reads	Percentage of mapped reads to the nuclear genome
<i>LD ZT0 1</i>	9,446,680	94.09%
<i>LD ZT4 1</i>	9,123,089	84.33%
<i>LD ZT8 1</i>	9,382,983	88.28%
<i>LD ZT12 1</i>	11,090,892	93.81%
<i>LD ZT16 1</i>	10,365,158	91.13%
<i>LD ZT20 1</i>	9,611,641	94.90%
<i>LD ZT0 2</i>	9,911,801	95.07%
<i>LD ZT4 2</i>	10,373,968	92.62%
<i>LD ZT8 2</i>	9,361,305	94.61%
<i>LD ZT12 2</i>	11,531,047	96.77%
<i>LD ZT16 2</i>	10,096,408	91.95%
<i>LD ZT20 2</i>	9,840,739	95.96%
<i>LD ZT0 3</i>	9,271,993	74.71%
<i>LD ZT4 3</i>	12,219,220	73.30%
<i>LD ZT8 3</i>	10,295,962	57.05%
<i>LD ZT12 3</i>	10,776,249	73.88%
<i>LD ZT16 3</i>	9,694,347	71.37%
<i>LD ZT20 3</i>	9,150,451	86.44%
<i>LD ZT0 4</i>	10,040,606	80.48%
<i>LD ZT4 4</i>	13,076,912	98.51%
<i>LD ZT8 4</i>	14,287,019	98.70%
<i>LD ZT12 4</i>	14,578,626	98.48%
<i>LD ZT16 4</i>	12,754,119	98.09%
<i>LD ZT20 4</i>	11,247,222	92.83%
<i>LD ZT0 5</i>	10,652,932	97.36%
<i>LD ZT4 5</i>	10,829,705	97.46%
<i>LD ZT8 5</i>	7,714,090	98.19%
<i>LD ZT12 5</i>	12,194,496	98.34%
<i>LD ZT16 5</i>	10,793,769	98.17%
<i>LD ZT20 5</i>	12,165,205	97.94%
<i>SD ZT0 1</i>	12,493,583	98.05%
<i>SD ZT4 1</i>	13,622,449	98.27%

<i>SD ZT8 1</i>	11,074,612	98.07%
<i>SD ZT12 1</i>	11,907,980	96.64%
<i>SD ZT16 1</i>	10,879,032	97.80%
<i>SD ZT20 1</i>	11,149,175	97.88%
<i>SD ZT0 2</i>	14,740,881	95.05%
<i>SD ZT4 2</i>	12,296,826	97.94%
<i>SD ZT8 2</i>	12,311,981	97.84%
<i>SD ZT12 2</i>	11,499,122	96.45%
<i>SD ZT16 2</i>	10,968,505	97.49%
<i>SD ZT20 2</i>	12,299,619	97.76%
<i>SD ZT0 3</i>	7,072,975	95.81%
<i>SD ZT4 3</i>	11,914,969	84.17%
<i>SD ZT8 3</i>	8,560,034	95.06%
<i>SD ZT12 3</i>	8,898,970	97.89%
<i>SD ZT16 3</i>	9,814,400	97.30%
<i>SD ZT20 3</i>	9,370,821	95.88%
<i>SD ZT0 4</i>	8,980,904	96.01%
<i>SD ZT4 4</i>	9,370,821	95.88%
<i>SD ZT8 4</i>	14,529,619	96.44%
<i>SD ZT12 4</i>	11,373,712	98.13%
<i>SD ZT16 4</i>	13,933,214	92.04%
<i>SD ZT20 4</i>	9,836,745	97.33%
<i>SD ZT0 5</i>	10,287,353	94.11%
<i>SD ZT4 5</i>	9,758,563	97.61%
<i>SD ZT8 5</i>	9,652,648	88.35%
<i>SD ZT12 5</i>	10,409,149	87.66%
<i>SD ZT16 5</i>	9,221,180	97.26%
<i>SD ZT20 5</i>	8,460,013	93.39%

**Appendix 2. Changes in the time point of maximum gene expression level (phase shift) and variation in gene maximum expression level (amplitude) as a consequence of seasonal photoperiod or day lengthening/shortening (skotoperiod or night shortening/lengthening).** Animation illustrating how seasonal day or photoperiod lengthening (night or skotoperiod shortening) may result in a gradual forward phase shift or an advancement in gene maximum expression level time point coupled to a gradual increase in amplitude or maximum gene expression level. Accordingly, seasonal day or photoperiod shortening (night or skotoperiod lengthening) may result in a gradual backward phase shift or an advancement in gene maximum expression level time point coupled to a gradual decrease in amplitude or maximum gene expression level. White rectangles represent photoperiods (light periods or days) and color filled rectangles correspond to skotoperiods (dark periods or nights). Transitions from red to blue colors and viceversa represent seasonal changes in photoperiods and skotoperiods.



**Appendix 3. Emergence of two peaks gene expression patterns as a consequence of seasonal photoperiod or day shortening (skotoperiod or night lengthening). Animation illustrating how seasonal day or photoperiod shortening (night or skotoperiod lengthening) may result in the emergence of two peaks gene expression patterns as a consequence of the two constituent gene expression profiles becoming out of phase.** The observed gene expression pattern represented by a continuous thick line can result from the combination of two distinct expression profiles represented by thin dotted and dashed lines. One of these expression profiles (dotted line) may not respond to seasonal changes in photoperiod/skotoperiod whereas the other one (dashed line) may experience phase shifts and amplitude changes. Under long day conditions (LD, alternating 16h light / 8h dark) the phases or maximum expression level time point of both expression profiles may coincide resulting in a single peak expression pattern. Whereas under short day conditions (SD, alternating 8h light / 16h dark) the phases may be reached at different time points producing a two peaks expression pattern. Transitions from red to blue colors and viceversa represent seasonal changes in photoperiods and skotoperiods resulting in gradual transitions from two peaks to a single peak expression pattern.



**Appendix 4. Percentage of cells in G1, S and G2|M phases in each time point under LD conditions.** The p-values as a result of the significant rhythmicity analysis are included in the lowest row.

Time	% of cells			
	G1	S	G2 M	
1 <sup>st</sup> day	ZT00	69.60	7.03	18.60
	ZT04	73.40	5.62	16.40
	ZT08	82.00	4.10	11.20
	ZT12	78.40	10.70	10.60
	ZT16	58.50	18.20	21.70
	ZT20	56.40	17.40	22.00
2 <sup>nd</sup> day	ZT00	68.80	6.65	19.30
	ZT04	76.20	7.19	13.60
	ZT08	79.50	5.85	13.70
	ZT12	72.60	9.70	13.80
	ZT16	70.00	9.98	17.60
	ZT20	65.20	13.90	18.50
3 <sup>rd</sup> day	ZT00	70.50	5.67	21.50
	ZT04	73.10	5.75	18.80
	ZT08	76.10	3.75	19.00
	ZT12	73.60	7.01	16.00
	ZT16	64.80	10.40	19.10
	ZT20	67.20	12.60	17.80
Rhythmicity p-value		9.14x10 <sup>-6</sup>	4.45x10 <sup>-6</sup>	
			1.53x10 <sup>-3</sup>	

**Appendix 5. Percentage of cells in G1, S and G2|M phases in each time point under SD conditions.** The p-values as a result of the significant rhythmicity analysis are included in the lowest row.

Time	% of cells		
	G1	S	G2 M
1 <sup>st</sup> day	ZT00	75.50	4.34
	ZT04	79.60	2.52
	ZT08	74.00	4.24
	ZT12	68.40	10.30
	ZT16	71.40	8.41
	ZT20	69.80	5.87
2 <sup>nd</sup> day	ZT00	71.90	5.31
	ZT04	75.10	6.54
	ZT08	76.30	5.05
	ZT12	68.50	10.70
	ZT16	67.20	8.72
	ZT20	71.10	9.85
3 <sup>rd</sup> day	ZT00	71.3	9.40
	ZT04	76.50	6.19
	ZT08	72.60	6.59
	ZT12	67.10	13.10
	ZT16	65.20	13.40
	ZT20	68.10	8.29
Rhythmicity p-value		9.07x10 <sup>-6</sup>	8.69x10 <sup>-4</sup>
			6.42x10 <sup>-2</sup>

**Appendix 6.  $F_v/F_m$  obtained at each time point under LD and SD conditions.** The p-values as a result of the significant rhythmicity analysis are included in the lowest row, for rhythms with 24 h and 12 h periods.

Time		$F_v/F_m$	
		LD	SD
1 <sup>st</sup> day	<b>ZT0</b>	0.584	0.490
	<b>ZT4</b>	0.603	0.525
	<b>ZT8</b>	0.607	0.556
	<b>ZT12</b>	0.582	0.455
	<b>ZT16</b>	0.572	0.423
	<b>ZT20</b>	0.575	0.484
2 <sup>nd</sup> day	<b>ZT0</b>	0.587	0.491
	<b>ZT4</b>	0.611	0.511
	<b>ZT8</b>	0.609	0.556
	<b>ZT12</b>	0.569	0.453
	<b>ZT16</b>	0.561	0.462
	<b>ZT20</b>	0.569	0.484
3 <sup>rd</sup> day	<b>ZT0</b>	0.587	0.486
	<b>ZT4</b>	0.597	0.526
	<b>ZT8</b>	0.611	0.561
	<b>ZT12</b>	0.591	0.459
	<b>ZT16</b>	0.577	0.481
	<b>ZT20</b>	0.572	0.485
Rhythmicity p-value	24 h period	7.07x10 <sup>-6</sup>	2.70x10 <sup>-11</sup>
	12 h period	1.00	1.50x10 <sup>-2</sup>

**Appendix 7. Starch content as percentage of Dry Weight under Long Day and Short Day conditions.** The p-values as a result of the significant rhythmicity analysis are included in the lowest row.

		% DW	
Time		LD	SD
1 <sup>st</sup> day	ZT00	27.70	31.67
	ZT04	30.89	33.65
	ZT08	41.87	30.6
	ZT12	36.37	27.52
	ZT16	31.41	25.52
	ZT20	28.24	25.47
2 <sup>nd</sup> day	ZT00	29.96	30.34
	ZT04	36.22	33.11
	ZT08	40.46	29.12
	ZT12	34.66	27.8
	ZT16	34.37	27.52
	ZT20	29.24	26.14
Rhythmicity p-value		1.2x10 <sup>-4</sup>	1.38x10 <sup>-4</sup>

**Appendix 8. Carotenoid content (% of total carotenoids) in Long Day entrained cultures.** The abbreviations used for each carotenoid are: AC for  $\alpha$ -carotene; L for lutein, P for prasinoxanthin; DL for dihydrolutein; M for micromonal; U for Uriolide; BC for  $\beta$ -carotene; Z for zeaxanthin; A for antheraxanthin; V for violaxanthin and N for neoxanthin.

		% of total carotenoids										
Time		AC	L	P	DL	M	U	BC	Z	A	V	N
1 <sup>st</sup> day	ZT00	0.88	0.00	29.08	7.63	5.67	7.39	4.76	6.34	6.30	18.30	13.63
	ZT04	0.84	0.00	29.97	8.10	6.00	7.86	4.89	4.96	4.24	18.09	15.05
	ZT08	0.51	3.42	25.31	7.74	5.97	5.77	5.53	9.39	19.94	3.55	12.87
	ZT12	0.79	3.15	26.23	7.31	5.48	5.81	6.32	10.09	17.20	4.93	12.68
	ZT16	0.86	0.00	27.88	7.74	5.31	7.54	5.67	8.11	12.59	11.06	13.24
	ZT20	0.85	0.00	29.26	7.50	5.65	7.87	5.59	6.69	7.74	14.69	14.18
2 <sup>nd</sup> day	ZT00	0.74	0.00	29.29	7.33	5.71	7.92	4.71	5.99	4.91	18.02	15.38
	ZT04	0.78	2.07	26.41	8.42	5.28	7.22	5.40	6.70	18.17	6.10	13.45
	ZT08	0.84	3.10	25.95	7.06	5.18	6.63	5.38	10.47	20.18	3.23	11.98
	ZT12	0.69	2.66	27.34	6.13	4.82	7.50	5.64	11.44	17.45	4.72	11.60
	ZT16	0.82	0.00	29.18	7.19	5.14	7.50	5.34	8.42	11.11	12.41	12.87
	ZT20	1.22	0.00	30.69	7.39	5.58	7.59	5.70	7.39	6.77	14.67	12.99
3 <sup>rd</sup> day	ZT00	0.81	0.00	28.91	6.84	5.33	7.03	4.86	6.23	5.32	20.44	14.23
	ZT04	0.86	1.92	26.51	7.54	5.16	7.03	5.45	8.16	19.44	5.08	12.86
	ZT08	0.59	3.18	26.15	6.27	4.95	6.63	5.32	11.91	20.64	2.45	11.91
	ZT12	0.57	3.79	27.5	5.86	4.16	6.60	5.69	12.65	16.24	5.72	11.23
	ZT16	1.24	0.00	29.46	6.04	4.96	6.67	4.91	9.91	14.24	10.13	12.44
	ZT20	1.02	0.00	32.72	5.27	5.69	7.59	3.43	9.68	11.27	11.15	12.19

**Appendix 9. Carotenoid content (% of total carotenoids) in Short Day entrained cultures.** The abbreviations used for each carotenoid are: AC for  $\alpha$ -carotene; L or lutein, P for prasinoxanthin; DL for dihydrolutein; M for micromonal; U for Uriolide; BC for  $\beta$ -carotene; Z for zeaxanthin; A for antheraxanthin; V for violaxanthin and N for neoxanthin.

		% of total carotenoids										
Time		AC	L	P	D	M	U	BC	Z	A	V	N
1 <sup>st</sup> day	ZT00	1.23	0.00	27.17	10.33	6.97	9.64	9.72	3.62	1.23	14.80	15.29
	ZT04	1.34	0.00	25.20	12.39	6.22	9.68	9.71	4.28	2.68	12.93	15.56
	ZT08	1.52	0.00	25.50	11.62	6.43	9.73	9.96	4.01	1.56	13.99	15.67
	ZT12	1.23	0.00	26.43	10.98	6.72	10.06	8.86	4.66	1.57	13.44	16.06
	ZT16	1.32	0.00	26.72	10.62	6.59	10.36	8.89	5.00	1.39	13.03	16.07
	ZT20	1.17	0.00	27.09	10.58	6.76	10.45	8.70	4.50	1.48	13.76	15.50
2 <sup>nd</sup> day	ZT00	1.26	0.00	25.93	10.33	6.71	10.47	9.68	3.64	1.43	15.14	15.42
	ZT04	1.57	0.00	24.46	12.41	6.34	9.71	10.07	4.26	2.47	13.59	15.12
	ZT08	1.40	0.00	25.66	11.72	6.68	10.57	9.83	4.20	1.85	12.72	15.35
	ZT12	1.10	0.00	26.15	10.93	6.66	9.97	8.76	4.58	1.56	14.37	15.92
	ZT16	1.36	0.00	25.74	10.45	6.55	10.63	8.67	4.87	1.75	13.86	16.12
	ZT20	1.15	0.00	26.19	10.86	6.89	10.31	8.89	4.66	1.65	14.18	15.23
3 <sup>rd</sup> day	ZT00	1.28	0.00	26.48	10.16	7.03	10.89	9.70	3.68	1.38	14.11	15.29
	ZT04	1.39	0.00	24.16	12.11	6.05	9.93	9.95	4.14	3.63	13.12	15.52
	ZT08	1.79	0.00	24.83	12.15	6.38	10.22	10.24	4.56	2.01	12.97	14.85
	ZT12	0.99	0.00	27.03	10.45	5.98	10.04	8.45	4.77	1.55	15.74	15.01
	ZT16	1.15	0.00	26.52	10.37	6.41	10.08	8.66	4.82	1.82	15.24	14.93
	ZT20	1.56	0.00	26.67	9.77	6.22	10.12	8.61	4.77	1.61	15.59	15.07

**Appendix 10. Results of the significant rhythmicity analysis for the carotenoids profiles in LD and SD conditions.**

Carotenoid	P-value	
	LD	SD
<b>α-carotene</b>	4.98x10 <sup>-2</sup>	3.00x10 <sup>-3</sup>
<b>Lutein</b>	1.54x10 <sup>-3</sup>	-
<b>Prasinoxanthin</b>	1.48x10 <sup>-6</sup>	7.37x10 <sup>-4</sup>
<b>Dihydrolutein</b>	4.28x10 <sup>-2</sup>	6.93x10 <sup>-6</sup>
<b>Micromonal</b>	2.04x10 <sup>-2</sup>	2.77x10 <sup>-2</sup>
<b>Uriolide</b>	3.25x10 <sup>-4</sup>	7.63x10 <sup>-2</sup>
<b>β-carotene</b>	1.10x10 <sup>-3</sup>	6.24x10 <sup>-5</sup>
<b>Zeaxanthin</b>	8.75x10 <sup>-6</sup>	2.28x10 <sup>-7</sup>
<b>Antheraxanthin</b>	7.70x10 <sup>-8</sup>	1.48x10 <sup>-3</sup>
<b>Violaxanthin</b>	5.44x10 <sup>-9</sup>	8.28x10 <sup>-2</sup>
<b>Neoxanthin</b>	1.65x10 <sup>-4</sup>	8.80x10 <sup>-1</sup>

**Appendix 11. Nitrate Reductase (NR) and Glutamine Synthetase (GS) enzymatic activities (U) for each time point in long day and short day entrained cultures. The p-values as a result of the significant rhythmicity analysis are included in the lowest row.**

		LD		SD	
Time		NR (U)	GS (U)	NR (U)	GS (U)
1 <sup>st</sup> day	ZT00	1.13	0.06	2.76	0.16
	ZT04	1.14	0.07	2.96	0.22
	ZT08	1.26	0.04	3.28	0.16
	ZT12	1.07	0.07	2.38	0.17
	ZT16	0.93	0.04	2.04	0.17
	ZT20	0.85	0.03	2.77	0.16
2 <sup>nd</sup> day	ZT00	1.25	0.04	2.46	0.19
	ZT04	1.10	0.08	2.32	0.19
	ZT08	1.50	0.05	3.10	0.18
	ZT12	1.27	0.04	2.52	0.12
	ZT16	1.02	0.04	2.48	0.14
	ZT20	1.15	0.04	1.87	0.15
Rhythmicity p-value		2.35x10 <sup>-2</sup>	4.53x10 <sup>-2</sup>	3.24x10 <sup>-2</sup>	2.04x10 <sup>-2</sup>





**Long live the walls we crashed through**