Doctoral Thesis

Multiomics Characterization of the Responses to Diurnal and Seasonal Cycles in the Marine Picoeukaryote *Ostreococcus tauri*

PhD Dissertation presented by Ana Belén Romero Losada to obtain the PhD Doctorate by the University of 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.

Dedicatoria . . .

word cloud

Contents

Index

Abstract

Abstract text

Abstract text

Introduction

Chronobiology.

Simply by looking around, it is easy to identify many cyclic processes. We are so much used to them that we usually don't perceive how they influence our lives and bodies.

There are four environmental cycles that we all know: tides (which repeat every 12 and half hours), lunar cycles (lasting 28.5 days), days (every 24 hours), lunar cycles (lasting 28.5 days) and years (every 365.25 days).

These cycles are also called circatidal, circalunar, circadian, or circannual (Numata et al., 2015) . These four rhythmic processes that arise from physical forces are quite important because they are extremely predictable. It is possible, for example, to know the exact time of a high tide years in advance. We constantly see how full moons are always followed by new moons, and everybody goes to bed sure that the sun will rise the next morning. The exact date for the next equinox is known and written in all calendars all over the world, so the change of season is never a surprise.

These four cycles affect earth with an overwhelming precision, it would be a foolish idea to think that earth living organisms don't react to such rhythmic changes. That's the main theme of Chronobiology: it is a young science that studies how these rhythmic environmental changes affect organisms(Edmunds, 1983; Kuhlman et al., 2018).

Living beings perceive environmental cyclic changes and they are able to react in advance generating endogenous biological rhythms thanks to an internal machinery that acts as a clock. Chronobiologists have found biological rhythms in a wide range of scales. It goes from a molecular level (transcription, translation, protein degradation, metabolites synthesis, etc), to a cellular and tissue level (cell division, synaptic connections, apoptosis, etc.) or even a complete organism or populations (Edmunds, 1983; Merrow et al., 2005; Sharma et al., 2022).

A surprising characteristic of theses rhythms is that they are self-sustained (Pittendrigh, 1960; Roenneberg & Merrow, 2005). For example, if we study a biological function that responds to light and light irradiance in our experiment is rhythmic, the measure of that biological function will also show a rhythmic profile. However, that rhythmic profile would not be maintained when the light input is turn off. But when we are studying a biological function that is regulated by that machinery called clock, it will maintain a rhythmic profile even under constant conditions and it will react in anticipation to the environmental change.

For that reason, every chronobiology experiment is designed as follows: several consecutive days where the organism is exposed to the rhythmic input (called zeitgeber, which is used as synchronizer) are followed by several consecutive days where the organism is exposed to constant conditions (called free-running conditions) (Kuhlman et al., 2018). In both scenarios, data is collected every few hours, minutes or seconds depending on the complexity of the data. For example, in the case of a circadian experiment, as the ones executed in this work, the zeitgeber would be a light-dark cycle and the free-running conditions would be constant light and constant darkness. Under that experimental design, circadian processes can be detected and discerned from light or dark responding processes.

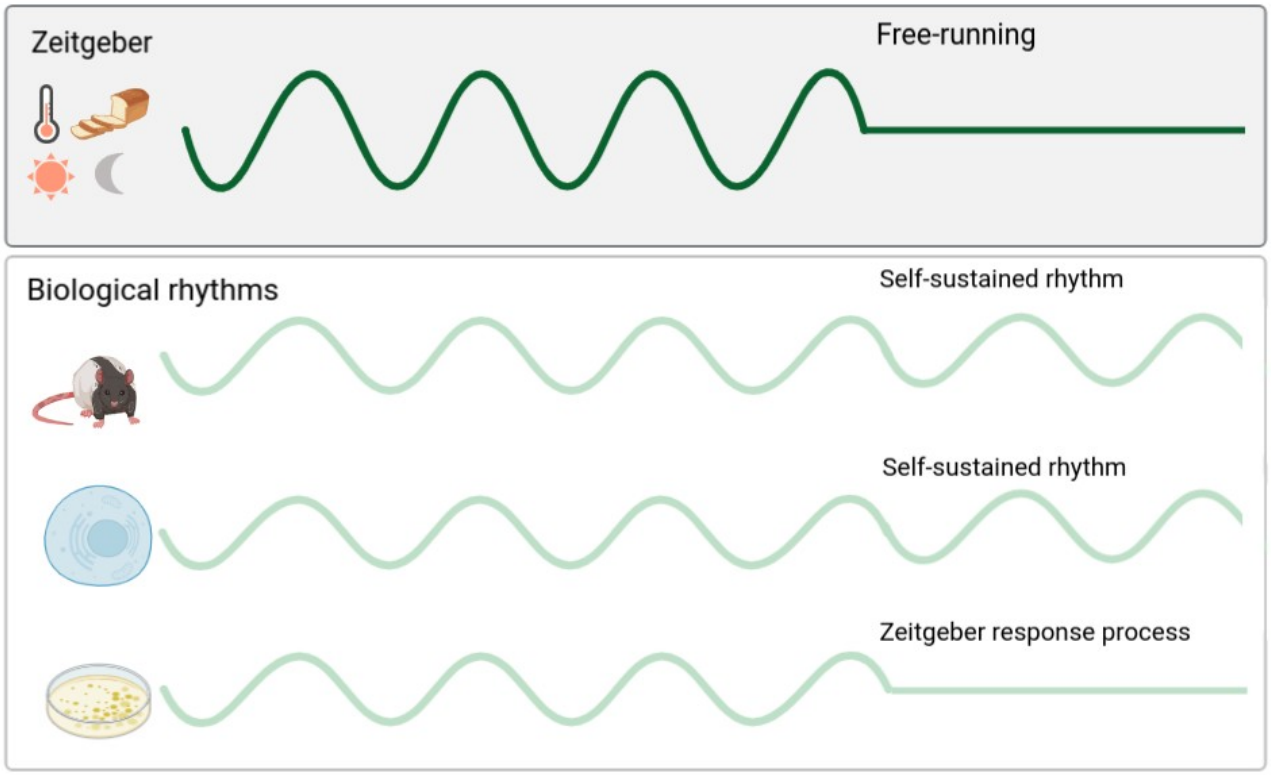

Figure 2: Under a zeitgeber (rhythmic environmental inputs as light/dark, temperature, food availability, etc) two kind of biological processes show rhythmic patterns: the ones that are self-sustained and thus regulated by an endogenous clock; the ones that are only responding to the zeitgeber. Self-sustained processes can be discerned from zeitgeber responding processes by changing the environmental cyclic condition to a constant one (free-running condition). Under free-running conditions, only the self-sustained processes will maintain their rhythmic profiles.

Following that experimental design, chronobiologists have described self-sustained biological rhythms reacting to the four different environmental rhythms mentioned above. For example, some marine organisms show self-sustained circatidal rhythms when they are kept in laboratory tanks without its zeitgeber, in this case, the tidal changes (Rock et al., 2022). As it was observed in the marine diatom *Hantzschia amphioxys* that descends to the sand at high tides and rises to the surface at low tides (Fauré-Fremiet, 1951). In contrast, the self-sustained circalunar processes are yet more unknown (Andreatta & Tessmar-Raible, 2020). One of the most famous ones is the larvae of the insect called Ant Lion, which build small holes in the sand as traps for insects. Scientist found that the size of the traps changes showing a circalunar profile that is maintained under constant conditions (Youthed & Moran, 1969).

However, the 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 are involved in activity-rest cycles (Roenneberg et al., 2022; Zee & Abbott, 2020), but hundreds of other parameters from behavior to gene expression also show circadian profiles. For example, the olfactory discrimination in mice is higher at night time, even under free-running conditions (Granados-Fuentes et al., 2006). Also fungus show circadian rhythmic phenomena, for example, *Neurospora crassa* generates asexual spores every 24 h even under constant darkness (Correa & Bell-Pedersen, 2002). Circadian rhythmic profiles in plants are found, for example, in leaf movement, growth rate, stomatal opening, as well as the expression of a wide range of genes (Merrow et al., 2005). Also all the organisms of the green lineage react in many different ways to circadian cycles (Noordally & Millar, 2015), one of the most famous ones is the 24 h-cyclic movement in the water column adjusted to their metabolism requirements (Lebert et al., 1999).

Day length or photoperiod, is a crucial signal for the circannual timing system. Surprisingly, seasonal changes of photoperiod have been strongly connected to reproduction. In fact, the entire animal reproductive systems are related to seasons, from gene expression profiles to anatomical structures. Hamsters kept in short days conditions have 10-fold smaller testes than the ones kept in long days conditions (Klante & Steinlechner, 1994; Nishiwaki-Ohkawa & Yoshimura, 2016). But also in plants, flowering and seed production is photoperiod-regulated (Brandoli et al., 2020; Serrano-Bueno et al., 2017).

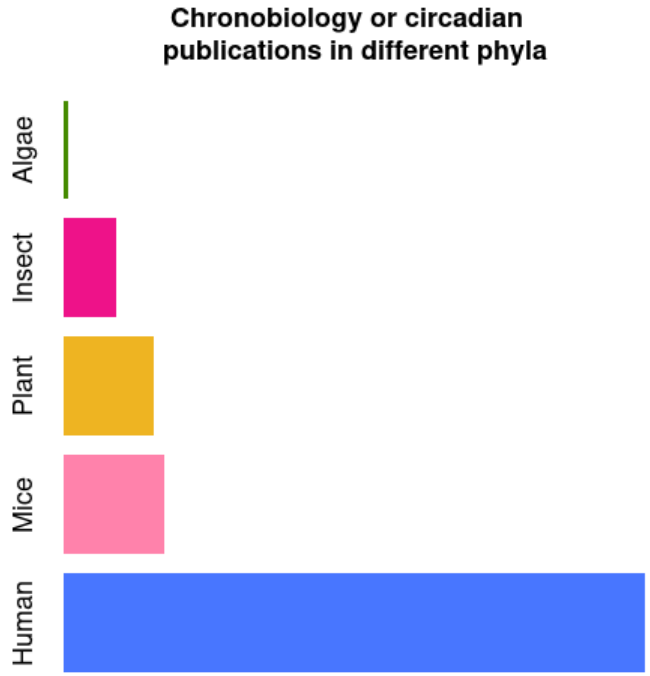

Figure 3: Number of publications found in PubMed using "chronobiology" or "circadian" keywords in their abstracts in July of 2022. For the algae group (using the generic term "algae" and "Ostreococcus tauri", "Chlamydomonas reinhardtii" and "polyedra" as main model organisms of this group), only 493 publications were found. For the insect group (including the generic term "insect" and "Drosophila melanogaster" as main model organism), 5717 publications were found. For plants (including the generic term "plant" and "Arabidopsis thaliana" as main model organism), 9734 publications were found. For mice and human, 10771 and 62305 publications were found, respectively.

Since photosynthetic organisms depend on light to ensure its success, they are highly synchronized with cyclic environmental changes involving light such as circadian rhythms and photoperiods. Nevertheless, the chronobiology of algae is yet barely studied (compared with other organisms) despite representing one of the largest polyphyletic groups in the eukaryotic domain. The already investigated genetic and molecular techniques used to identify clock components in other taxa have not been widely applied to algae yet (Noordally & Millar, 2015). This work aims to contribute to the chronobiology community by describing the circadian and photoperiodic changes in the microalgae *Ostreococcus tauri*.

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 all this phenomena 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 (as plants, fungus and algae ones) will also help in the comprehension of the circadian clock in humans and its variation among individuals.

The synchrony that exists between the sunrise/twilight and organisms have been so obvious for scientists that the underlying biological 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 light and darkness. Although De Mairan invited botanists to investigate his discovery to confirm that leaf movement was also maintained when temperature changes are avoid (which inside its closet was difficult to ensure). Nevertheless, it took 30 years to confirm what De Mairan observed, taking attention to temperature for the first time (Kuhlman et al., 2018; Roenneberg & Merrow, 2005).

The history of the circadian research have a lot of gaps between one discovery and another: the physiology of the endogenous nature of the clock in plants wasn't studied until 1832, despite De Mairan observations; similar observations in animals took another century, and 50 years more for humans (Kuhlman et al., 2018; McClung, 2006; Roenneberg & Merrow, 2005). Though circadian research and chronobiology was born in the 18th century, it took relevance and coherence in the 20th century. One of the breaking points in circadian research history was the international conference in Cold Spring Harbor, where 157 pioneers (Colin Pittendrigh, Patricia Decourse, Franz Halberg, etc) of this field met together for the fist time (Evans, 1961). From all the data shared, Pittendrigh summarized the qualities of circadian clocks in 16 generalizations that he predicted to be true in all organisms

(Pittendrigh, 1960). Nowadays, 62 years later, these generalizations are still being useful though circadian research has suffered a huge development. The new technology approaches have enabled a more controlled experimental design and analysis of the data. The circadian community has already find circadian rhythms in almost all kind of organisms, even non-photosynthetic prokaryotes (Eelderink-Chen et al., 2021). Also the physiology and genetics behind the circadian clock have been broadly studied in a wide range of phyla since the first gene of the clock was described in 1971 by Seymour Benzer and Ronald Konopka using mutant screening in *Drosophila melanogaster* (Konopka & Benzer, 1971; Takahashi, 2021).

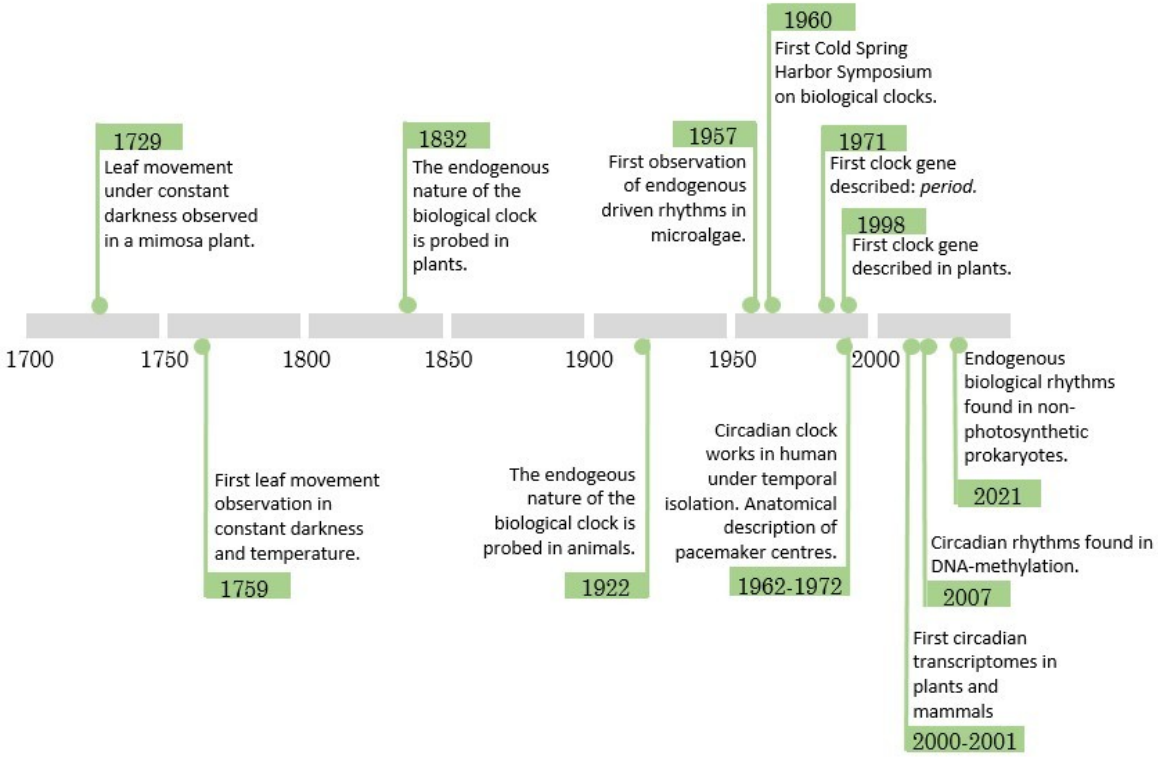

Figure 4: Timeline of circadian research. The main circadian discoveries have been listed in chronological order (De Mairan, 1729; Eelderink-Chen et al., 2021; Evans, 1961; Konopka & Benzer, 1971; Kuhlman et al., 2018; McClung, 2006; Ripperger & Merrow, 2011; Roenneberg & Merrow, 2005; Takahashi, 2021).

As it can be observed in the timeline, circadian rhythms discoveries are very diverse. Since biological rhythms can be found in a wide range of scales, as it was mentioned in the chronobiology introduction, chronobiologists studying them usually also have strong roots in other fields such as anatomy, physiology, molecular biology, genetics, ecology or

even mathematics. During all these years of research, the knowledge obtained from each field have been shared in order to obtain a better picture of circadian rhythms. In the circadian community, anatomy studies have strong bases in physiology, ecology studies have been explained with genetics (Klante & Steinlechner, 1994; Merrow et al., 2005; Nishiwaki-Ohkawa & Yoshimura, 2016), and so on. Mathematics, however, have influenced the complete research field since the study of the biological rhythms as waves have been crucial. A wave's shape repeat itself over and over, maintaining several characteristics that define the wave as it is. Those characteristics are called wave parameters and are used to quantitative compare different waves. There are several waves parameters that are used in mathematics: period (the time between two peaks of a wave) or wavelength (distance between two peaks of a wave), frequency (number of wave repetitions per seconds), amplitude (how much high the wave reach), phase (time point where the wave reaches its maximum high), mesor (mean level from which the wave fluctuates around), etc.

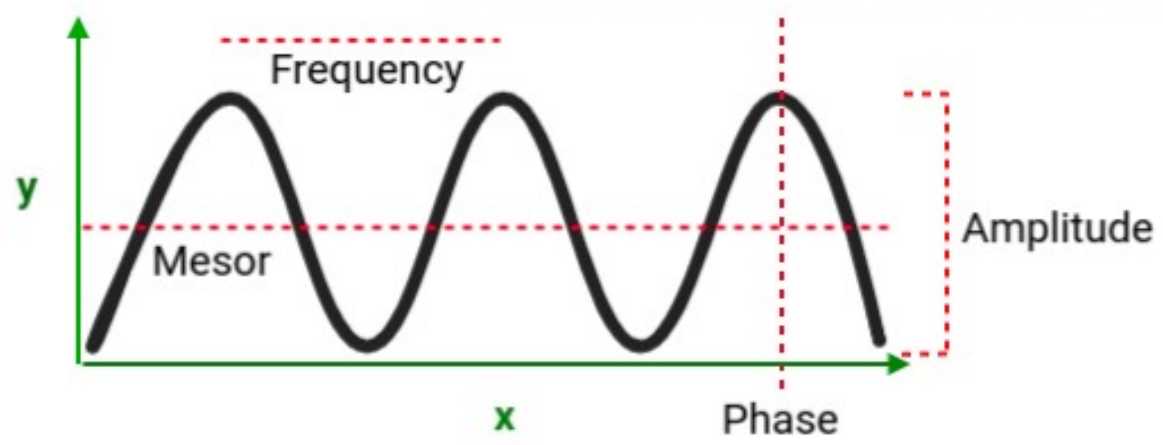

Figure 5: Graphical representation of classic waves parameters: period (the time between two peaks of a wave) or wavelength (distance between two peaks of a wave), frequency (number of wave repetitions per seconds), amplitude (how much high the wave reach), phase (time point where the wave reaches its maximum high) and mesor (mean level from which the wave fluctuates around).

The most used ones for chronobiologists are amplitude, phase, period and mesor. The parameterization of waves using these four parameters enables to mathematically compare different groups of rhythmic data (McClung, 2006; Parsons et al., 2020). A deep study of circadian waves in *Ostreococcus tauri* was achieved during this work, using theses parameters to find and statistically validates differences between rhythms found in three biological levels (mRNAs, proteins and physiology).

Ostreococcus tauri.

The green linage (Viridiplantae), comprehends two of the most important groups of oxygen photosynthetic eukaryotes: green algae and their descendants terrestrial plants. Nowadays, the development of high-throughput sequencing has allowed researchers to clarify the evolution history of these lineage (Bachy et al., 2022; Becker & Marin, 2009; Benites et al., 2021; Leliaert et al., 2012; Merchant et al., 2007). An early divergence of two discrete clades from an ancestral green microalgae is hypothesized. These two main clades are: the Streptophyta, including land plants and charophyte (green algae that are their closest ancestors/older sisters); and the Chlorophyta, comprising the core chlorophytes and their closest ancestors/older sisters the prasinophytes (Bachy et al., 2022; Leliaert et al., 2012; Tragin & Vaulot, 2019).

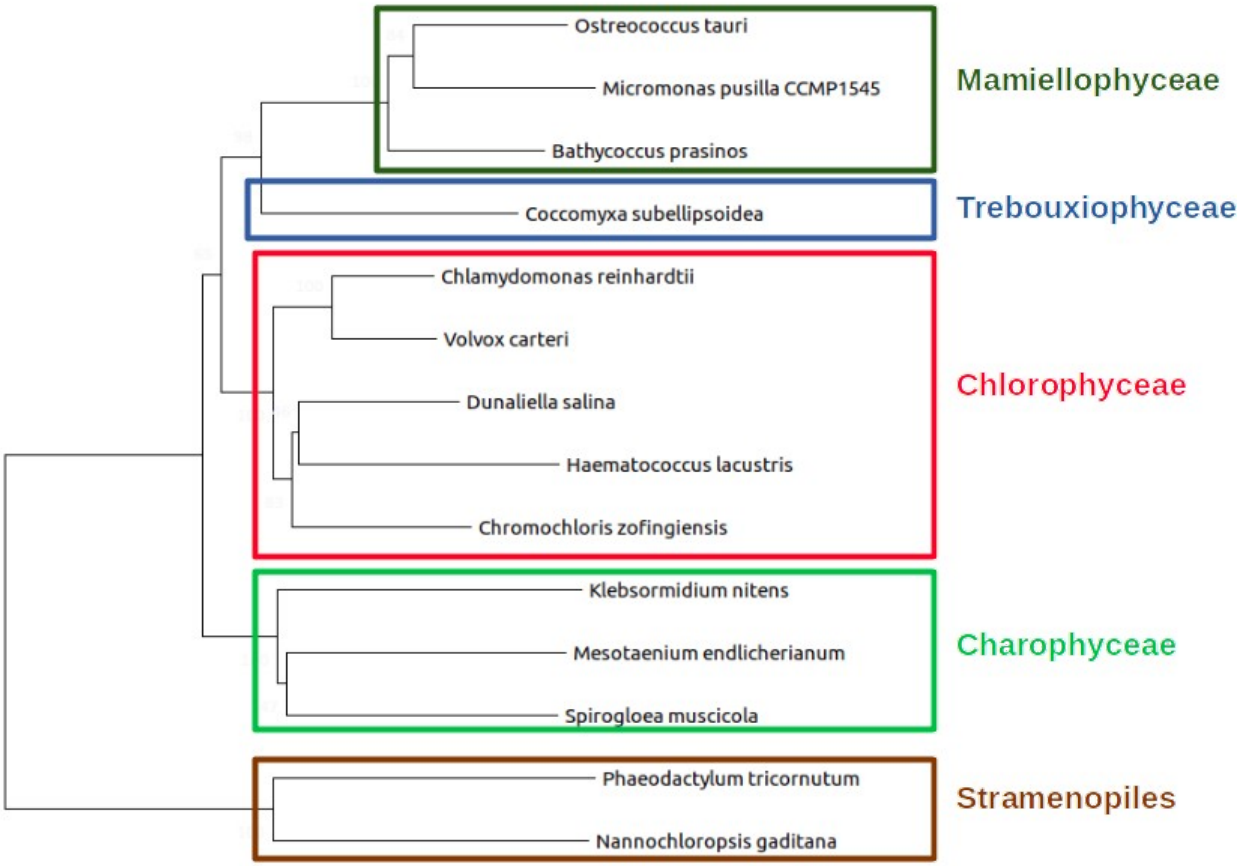

Figure 6: (mejorar arbol y escribir pie de foto)

Previous hypotheses posited the domination of the marine waters by the so-called red lineage conformed by diatoms and dinoflagellates. While the green lineage was thought to

have less importance in marine water than in terrestrial environments (Worden et al., 2004). These two lineages established their differences a long time ago when the endosymbiotic events took place, giving rise to the green lineage (one single endosymbiotic event) and the red lineage (two or even more endosymbiotic events) (Leliaert et al., 2012).

However, in the last decade, metabarcoding studies have completed the previous hypotheses based in microscopic and traditional molecular techniques. Theses studies have confirmed the importance of the green lineage in marine waters. Also, the cosmopolitan distribution of prasinophytes has been described, especially the older sisters from the order Mamiellales (Collado-Fabbri et al., 2011; Demir-Hilton et al., 2011; Leconte et al., 2020; Tragin & Vaulot, 2019; Worden et al., 2004).

In that taxonomic context is placed Ostreococcus tauri, which is considered a green mamiellale algae. Its contribution to the marine phytoplankton is crucial in a wide range of oceans and seas all over the world (Benites et al., 2021; Collado-Fabbri et al., 2011; Demir-Hilton et al., 2011). As an example, different Ostreococcus strains have caused blooms in the Atlantic and Pacific Oceans (O'Kelly et al., 2003; Worden et al., 2004). Furthermore, a certain Ostreococcus strain was found to be the most prevalent Mamiellales in the Mediterranean Sea (Tragin & Vaulot, 2019). In spite of its ecologically important role (Chapman, 2013; Worden et al., 2004) and presence in natural environments, Ostreococcus tauri is 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 described in a bloom that took place on the french Thau Lagoon. By that time, field researchers described it as "undetectable" so the cells were discovered by flow cytometry. As they correctly described, it contains a nucleus, only one chloroplast and one mitochondrion, a starch granule and a very reduced or almost non-existent cytoplasmic compartment (Moreau, H, Grimsley, N. Derelle, E, Ferraz, C, Escande, ML, Eychenié, S , Cooke, R, Piganeau, G, Desdevises, Y, Bellec, 1995).

Meanwhile, genome sequencing approaches to understand marine phytoplankton were focus on the prokaryote component of it (Berube et al., 2018; Palenik et al., 2003). It wasn't until 2006 when the genome of Ostreococcus tauri was sequenced by the first time (Derelle et al., 2006). From that point, an increasing number of picoeukaryotes genomes were sequenced, including a second version of the genome of Ostreococcus tauri (Blanc-Mathieu et al., 2014). These contributions allowed the identification of unique aspects of its genome, showing that its simplicity goes among its reduced cell structure.

A genome size of 12.56 Mb distributed in 20 chromosomes. Nothing remarkable at first sight since Saccharomyces has similar numbers. The fascinating fact comes when the number of genes of theses two organisms are compared. 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). This makes Ostreococcus tauri the most gene dense free-living eukaryote known to date.

In addition to its short intergenic regions, Ostreococcus tauri has reduced the size of gene families keeping only one copy of each gene or even merged different genes in 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. But there is more perplexing data from the chromosome 19 and also from chromosome 2. They contain 77% of the transposable elements of the genome, they have lower G+C content and even a different codon usage in low G+C content loci of chromosome 2 (Blanc-Mathieu et al., 2014; Derelle et al., 2006). The first hypotheses about these chromosomes were that they had a different origin that the rest of the genome. Currently, only the chromosome 19 is considered an alien chromosome since the most of its protein coding genes aren't related to the green lineage. However, chromosome 2 protein coding genes are essential housekeeping genes not duplicated and related to the green lineage, so from that point it was considered 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 lacking in most unicellular eukaryotic lineages. These type of genomic regions so-called mating-type has been characterized in other Chlorophyta (Sekimoto, 2017) and, recently, in Ostreococcus tauri which appears to encode two highly divergent haplotypes. These Mamiellales mating-types regions candidates are likely to be the oldest mating-type loci described to date (Benites et al., 2021; Leconte et al., 2020).

All in all, 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 lot of diversity and actually it is difficult to find a model organism that can represent the hole lineage (Cock & Coelho, 2011). However, its taxonomy classification makes Ostreococcus tauri a potential green lineage ancestor (Derelle et al., 2006; Leliaert et al., 2012) and the knowledge gained using it can be easily extrapolated to a wide range of photosynthetic organisms.

Also, studies in Systems Molecular Biology often deal with the problem that complex organisms maximize the issue to study since massive data is generated in order to study complete biological systems (De Keersmaecker et al., 2006; Jamers et al., 2009; Joyce & Palsson, 2006; Weckwerth, 2011).

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; 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.
Arabidopsis thaliana	135	27474	1779
Klebsormidium nitens	0	17055	286
Chalmydononas reinhardtii	110	14000	279
Phaeodactylum tricornutum	27.4	10567	212
Ostreococcus tauri	12.56	6275	102

Nevertheless, when Ostreococcus tauri is compared with another microalgae that are already sequenced and had been used as model organisms, it turns clear the simplification that it brings to Systems Molecular Biology studies (de los Reyes et al., 2017; Derelle et al., 2006; Krumholz et al., 2012; Le Bihan et al., 2011; Lelandais et al., 2016). *Ostreococcus tauri* allowed us to generate, analyze and interpret massive data from the complete system with less computational and experimental costs, so an holistic understanding of the chronobiology in this mammiellale was finally achieved.

Systems Biology.

For a scientist like me, it has been a passionate quarter of a century to live. First-generation DNA sequencing methods were developed only about 45 years ago. In addition, one of the most biggest international projects, the Human Genome Project, started about 20 years ago. (Ideker et al., 2001; Veenstra, 2021) In this project an extremely ambitious idea was pursued: to sequence the hole human genome. However, its impact in science transcends beyond that goal. Different groups cooperated all over the world, not only to sequence the complete human genome, but to develop new sequencing methods in order to make the process easier and cheaper (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 methods, around 100000 human genomes have been sequenced in the last 6 years. The Illumina HiSeq System generates around 500 gigabase sequences per sun, dropping the cost of sequencing a complete human genome to $1500 in only 10 years of difference (Prendergast et al., 2020; Veenstra, 2021). Ultimately, it changed how science was approached.

All this advances led to a mass development of methods to sequence and identify the complete transcriptome (total mRNAs transcripts). Sequencing of complete transcriptomes offered a massive amount of information never seen before. However, genes and their products (proteins) are highly related since they interact and regulate each other forming positive and negative feedback loops, so still a lot of information was missing. Researchers began to understand organisms and process as systems with important modules (mRNA, proteins, metabolites, etc) that interact with each other forming parts of large networks.(Joyce & Palsson, 2006; Veenstra, 2021; Weckwerth, 2011).

While this holistic view of biological systems were gaining strength, traditional reductionist methods (focusing in only one gene, protein or metabolite) were the most popular and asequible ones. Scientific research was limited by the time and effort needed to complete integrate all functions that occur simultaneously within a biological system from a set of individual results, which has been almost impossible to achieve (Karahalil, 2016; Mazzocchi, 2012; Veenstra, 2021). There is a ancient Indian fable that illustrate how an 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;
"Tis 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 them concluded that the elephant was like a snake (he was touching only the elephant’s trunk), another one concluded that it was

like a fan (since he was only touching the animal's ear), and so on. Each of them were sure about their findings, but reaching an agreement was impossible since they didn't treated their collected data as parts of a complex system, instead of independent truths.

During a long time in scientific research, systems biology studies that would bring that holistic view of living systems have been impossible to achieve due to a lack of technologies available. The development of systems biology is related to technology advances such as computational science, artificial intelligence and, the already mentioned, next-generation sequencing methods (Ideker et al., 2001; Karahalil, 2016; Veenstra, 2021; Weckwerth, 2011).

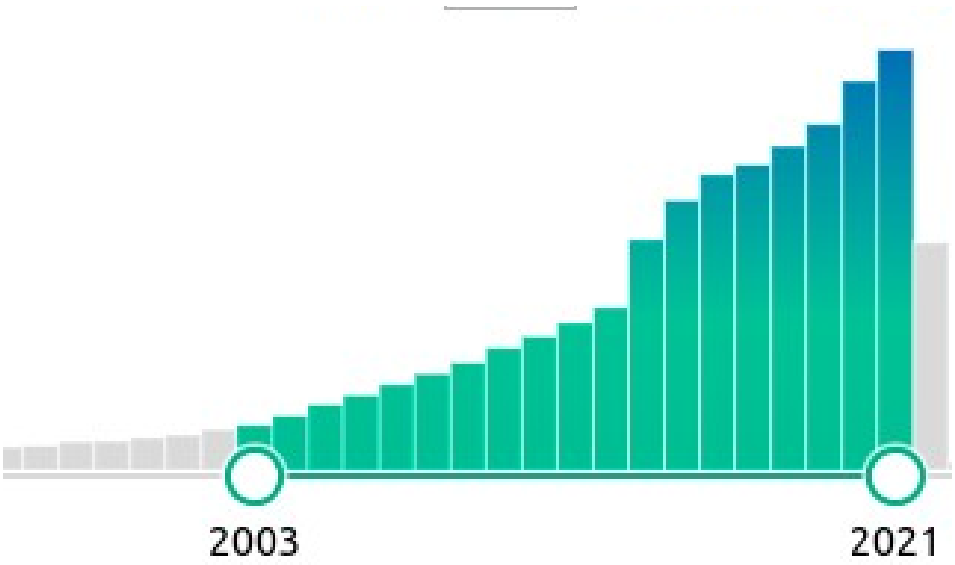

Figure 7: (es provisional, haría una mas bonita y mejor) Exponential increase in the number of publication using the term "systems biology" in PubMed since the year when the Genome Human Project was completed.

Nowadays, the so-called omics methods allow to mass measure all those important modules or biological levels that form part of the studied systems: transcriptomics, proteomics, metabolomics, etc. Systems biology aims to improve our understanding of living systems through the integration of our knowledge on how different biological components work simultaneously. The typical methodology of a systems biology study starts by obtaining omic data and its subsequent integration. Then, the results are computationally and statistically analyzed and the phenomenon observed is experimentally probed. Consequently, predictive models can be developed in order to simulate how a biological system would behave

when perturbed (Jamers et al., 2009; Veenstra, 2021; Weckwerth, 2011; Zurbriggen et al., 2012).

During the progression of this work, the generation and analysis of massive data from two different omics techniques (transcriptomics and proteomics) is achieved. The main purpose is studying *Ostreococcus tauri* as a biological system by integrating the omics results with traditional physiological measurements as a biological validation of the effects observed in the computational analysis.

Specifically, the transcriptomic method used was RNA-seq, which is the main contemporary method used for this omic. The development of next-generation sequencing methods has contributed to RNA-Seq analyses enabling working with a wide variety of different classes of RNAs, not requiring transcript-specific probes (unlike microarrays, the almost obsolete previous method used in transcriptomic analyses), and not only to identify but also to quantify abundance of transcripts (Ditz et al., 2021; Veenstra, 2021; Wang et al., 2009).

Meanwhile, the field of proteomics have been directly connected with the development of mass spectrometry (MS) technology. During the first part of the 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, technology has enable 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). In this study, a large scale proteomic analysis is achieved by SWATH using liquid chromatography MS / MS. SWATH proteomics enables protein identification and characterization, as well as label free relative quantification (M. X. Chen et al., 2021; Ludwig et al., 2018).

Currently, the era of systems biology is increasing the amount of data generated per study and consequently, the computational and mathematical knowledge required to analyze and integrate these results increases too. Unfortunately, most laboratories are not historically designed to incorporate this requirements yet due to a lack of qualified researchers gathering solid knowledge from the different disciplines needed: computational sciences, mathematics/statistics and molecular biology. The lasts new generations of young researchers are working to develop software applications, efficient data analysis algorithms and user-friendly app-tools to enable the progress of systems biology studies making it

more accessible for the hole scientific community (Coleto-Alcudia & Vega-Rodríguez, 2020; Romero-Campero et al., 2016; Romero-Losada et al., 2022).

However, systems biology studies in microalgae were recently started and there is a lack of tools for microalgae to analyze and interpret omics data. Consequently, during the progression of my doctoral thesis I aim to contribute to the progression of systems biology studies in the microalgae research community developing the web-app ALGAEFUN with MARACAS. In that way, any researcher can analyze and functional annotate RNA-seq and ChIP-seq data without previous knowledge in computational or mathematical analysis (Romero-Losada et al., 2022).

In summary, my doctoral thesis aims to contribute to the microalgae research community with two major items: the development of free and open source tools that facilitates systems biology studies in microalgae; and the understanding of the diurnal and seasonal rhythmic changes in *Ostreococcus tauri* as a complete system knowing all the genes expressed, all the proteins present and how the altered functions are being executed.

Materials and Methods

Algal Material, Growth Conditions, and Sample Collection

The sequenced strain of *Ostreococcus tauri,* RCC4221, was used in this study. Autoclaved artificial see water (ASW) (Kester et al., 1967) supplemented filter sterilized (0.2µm) nitrogen source, potassium source and vitamins were used as growing medium. A complete list of media components and concentrations needed to grow *Ostreococcus tauri* can be found in table 2. Cells were grown in photobiorreactors, water jacketed bubble columns with 2 L capacity (7 cm diameter, 50 cm height) containing 1.8 L of cell culture, under continuous regime and neumatic agitation (bubbled with air supplemented with 1% (v/v) CO2 as carbon source). Temperature was constantly maintained at 20ºC by the flow of water from the jacket to an external cooling system. Other culture growth conditions were constantly measured and computationally controlled by a LabJack: pH, fresh flow rate (dilution rate) and light regime. A pH probe connected to the LabJack keep pH at 8 by controlling a CO_2 electrovalve on demand. The dilution rate were maintained constant at 0.3 d^{-1} (overflow is descarded in orded to avoid a volume increase) to keep the cells in the culture in exponential phase. In adittion, instead of a sudden transition from dark to light and from light to dark, our LabJack controlled system gradually increased light until an irradiance of 2500 µE m^{-2} s^{-1} during the hours of daylight that corresponds to each simulated photoperiod. Then, light irradiance is decreased gradually until total darkness is reached. Each photochemostat is illuminated using six Phillips PL-32 W/840/4p white-light fluorescent lamps. Furthermore, they are surrounded by a wooden box and a completely opaque fabric to avoid any external light input. Long day light regimen consists of 16h of light and 8h of darkness, while short day light regimen consists of 8h of light and 16h of darkness.

Cultures at exponential phase with 45 µg/mL chlorophyll content were used in our experiments. Cells were harvested at specific times in the daily cycle, expressed in zeitgeber time (ZT), where ZT0 corresponds to dawn, ZT4 to 4h after dawn, ZT8 to 8h after down, and so on. In this study, samples were taken every 4h (from ZT0 to ZT20), so there is a total of 6 samples for each day of sampling. Cells were collected using different methods depending on the experiment.

Table 2: (falta añadir descripción)

	Concentration in solution	Concentration in medium
Solución I	400 g/L de $NaNO_3$	222.22 mg/L de $NaNO_3$
Solución II	2,8 g/L de Na_2HPO_4	1.56 mg/L de Na_2HPO_4
	10 g/L of K_2HPO_4	5.56 mg/L of K_2HPO_4
Solución III	5.36 g/L of NH_4Cl	2.98 mg/L of NH_4Cl
	10.4 g/L of Fe-EDTA	5.78 mg/L of Fe-EDTA
	74.4 g/L of Na_2-EDTA	41.33 mg/L of Na_2-EDTA
	$4.6 \cdot 10^{-2}$ g/L of $ZnSO_4$	$2.56 \cdot 10^{-2}$ mg/L of $ZnSO_4$
	$2.8 \cdot 10^{-2}$ g/L of $CoSO_4$	$1.56 \cdot 10^{-2}$ mg/L of $CoSO_4$
	$1.6 \cdot 10^{-2}$ g/L of $Na_2MoO_4 \cdot 2H_2O$	$8.89 \cdot 10^{-2}$ mg/L of $Na_2MoO_4 \cdot 2H_2O$
	$5.0 \cdot 10^{-3}$ g/L of $CuSO_4$	$2.78 \cdot 10^{-3}$ mg/L of $CuSO_4$
	$3.4 \cdot 10^{-2}$ g/L of H_2SeO_3	$1.89 \cdot 10^{-2}$ mg/L of H_2SeO_3
	$3.6 \cdot 10^{-2}$ g/L of $MnCl_2 \cdot 4H_2O$	$2.0 \cdot 10^{-2}$ mg/L of $MnCl_2 \cdot 4H_2O$
Solución IV	0,2 g/L of Thiamin-HCl	0.22 mg/L of Thiamin-HCl
	$1.50 \cdot 10^{-3}$ g/L of Biotin	$1.67 \cdot 10^{-3}$ mg/L of Biotin
	$1.50 \cdot 10^{-3}$ g/L of Vitamin B12	$1.67 \cdot 10^{-3}$ mg/L of Vitamin B12

Klebsormidium nitens and *Haematococcus lacustris* cultures were maintained as described in (Hoys et al., 2021; Serrano-Pérez et al., 2022).

RNA-Seq Data Generation and Processing

A volume of 50mL of cell suspension were pelleted (4 min centrifugation at 5000 x g and 4ºC) for each time point. After a washing step with Phosphate-buffered saline solution (PBS), pellets were flash frozen with liquid Nitrogen and stored at −80 ºC.

For cell disruption prior to RNA extraction, frozen pellets were resuspended in 400 µL of STET buffer (García-Domínguez & Florencio, 1997) and directly added to a 1,5 mL Eppendorf tube (RNAse free and phenol-proof) that contains 400 µL of phenol:chloroform 1:1 and 100 µL of acid washed glass beads (0.25–0.3 mm diameter; Braun, Melsungen, Germany). Mechanical disruption was performed by 30 min of repeated cycles of 60 s of vortexing and 60 s of incubating on ice.

For RNA extraction, in order to separate the different phases, the tubes were centrifuged (4 ºC) for 15 min at 13000 x g. Phenolization process was repeated three times more with centrifugations of 5 min, followed by a step using only chloroform instead of phenol:chloroform 1:1 to avoid phenol contamination of the samples. Finally the supernatan is incubated overnight at -20 ºC in a solution of 80 µL ClLi (10 M) and 550 µL of 100% EtOH for RNA precipitation.

After overnight incubation, tubes were centrifuged during 10 min at max. speed and 4ºC. The resulting pellets were dryed to avoid EtOH contamination of the samples. 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 eluded RNA concentration and integrity were measured using a bioanalyzer 2100 (Agilent RNA 6000 Nano Kit).

Library was prepared in accordance with the manufacturer's instructions and the sequencing was carried out on the Illumina NextSeq500 sequencer. Approximately, 10 million 75nt long single end reads were generated for each sample. The *Ostreococcus tauri* genome sequence and annotation v3.0 were used as reference genome (Blanc-Mathieu et al., 2014). Futher computational analysis were carried out using MARACAS from ALGAEFUN with MARACAS (Romero-Losada et al., 2022).

SWATH proteomics Data Generation and Processing

For each time point, a volume of 50mL of cell suspension were pelleted (4 min centrifugation at 5000 RCF and 4ºC). After a washing step with Phosphate-buffered saline solution (PBS), pellets were flash frozen with liquid Nitrogen and stored at −80 ºC. For cell disruption, 1 L of Trizol, 100µL of acid washed glass beads (0.25–0.3 mm diameter; Braun, Melsungen, Germany) and 40µL of PIC (25x) were directly applied onto frozen pellets followed by 3 disruption cycles (60s agitation-60s incubation on ice) using a Mini-Beadbeater (BioSpec Products).

Proteins were extracted using TRI Reagent (Sigma-Aldrich), according to the manufacturer's instructions. The resulting proteins pellets were washed with 2mL of guanidine solution 0.3M in 95% EtOH. They were resuspended by 10min of sonication (falta modelo) cycles (30s sonication-30s of incubating at 4ºC) and then centrifugated at 4ºC during 5min at

8000 x g. This washing process was repeated twice, followed by two additional repetitions using 90% EtOH instead of the guanidine solution. The final pellets were resuspended in ammonium bicarbonate 50 mM/0.2% Rapidgest (Waters) and total proteins were quantified using Qubit system. Of those samples, 50µg of proteins were incubated with DTT (final concentration 4.5 mM) for 30 min at 60 ºC. Then, iodoacetamide to a final concentration of 10 mM were added to continue the incubation for 30 min more under total darkness at room temperature. The treatment with trypsin was done "overnight" at 37ºC in a 1:40 ratio (trypsin:protein). The following day, formic acid was added and incubated at 37ºC for 1h. Finally, acetonitrile and water were added, so that the concentration of the digested sample was around 0.5 µg/µl and 2% acetonitrile, injecting 2ml (1 mg of protein in column) in LC-MS. A mixture of synthetic peptides from Sciex was also added to each sample at a final concentration of 20 phentomoles/ml, for subsequent alignment of the chromatograms necessary for SWATH processing.

The analysis were performed on a TOF triple quadrupole hybrid mass spectrometer (5600 plus, Sciex) equipped with a nano electrospray source coupled to an nanoHPLC Eksigent model 425. The Sciex software Analyst TF 1.7 was used for equipment control, data acquisition and processing. 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% (v/v) acetonitrile at a flow rate of 3 µL/min for 10 min. Subsequently, they were eluted on a reversed-phase analytical column, with the built-in emitter (New Objective PicoFrit column, 75 µm id × 250 mm, packed with Reprosil-PUR 3 µm). In the case of the samples corresponding to the short day conditions, the analytical column was 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). 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. Formic acid 0.1 % (v/v) and acetonitrile with formic 0.1 % (v/v) were used as solvents A and B, respectively. The source voltage was selected at 2600 V and the heater temperature was maintained at 100 ºC. Gas 1 was selected at 20 PSI, gas 2 at zero, and curtain gas at 25 PSI.

The spectral library was constructed by making 1 run with a mixture of the biological replicates corresponding to each time point (ZTO, ZT4, ZT8, ZT12, ZT16, ZT20) with the DDA (data dependent acquisition) method. It consists 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.

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, iodoacetamide as cysteine alkylating agent. The Ostreococcus tauri database, provided by the investigator (de donde la sacamos al final? No era una anotación que refinamos nosotros?), linked to a Sciex Contaminants database was used as a database. A false positive analysis (FDR) was performed and those with FDR < or equal to 1 were considered.

The SWATH method consisted of a TOF MS with a 400-1250 m/z scan window (with 50 ms accumulation time) followed by 60 m/z windows, of varying size (230-1250 m/z) overlapping 1Da at the ends (with 60 ms accumulation time). The cycle time was 3.7 s. The equivalent of 1 µg of digested protein was injected into each run. Prior to each sample, a standard (MS synthetic peptide calibration kit from Sciex) was injected to self-calibrate the equipment, control the sensitivity and chromatographic conditions.

The library with the proteins generated by DDA (1 % FDR) was used in the analysis performed using the Sciex software PeaKView 2.2 with the microapp SWATH 2.0, together with the data obtained with the SWATH method. Using this program, the chromatographic traces of the ions were extracted and dumped into the Marker view 1.2.1.1 program where the list of identified proteins with their corresponding areas was generated. The parameters for extraction of ions and obtaining the areas were: 10 peptides per protein, 7 transitions of each peptide, threshold of confidence of the peptides set at 90 and FDR 1%. These data was later used to perform the statistical treatment (cual ?) and obtain the relative quantifications of the proteins. The parameters for extraction of the ions and obtaining the areas were: 10 peptides per protein, 7 transitions of each peptide, threshold of confidence of the peptides set at 90 and FDR 1%.

(protocolo de analisis en R?)

Cell cytometry data Generation and Processing

A volume of 1,5mL of cell suspension were harvested for each time point. This samples were diluted 1:10 in PBS to be sure the cell concentration is suitable to the assay. 2ml of these dilutions were pelleted and cells were fixed with 10 mL of 100% EtOH before stored at -20ºC for at least 24h. After fixation, cell suspensions were centrifuged for 5 min at 3500 x g (room temperature) and resuspended in 1 mL of PBS. They were washed once with PBS and sonicated for 3 minutes in a Ultrasonic Cleaner (JSP, US21, ultrasonic power 50W), in order to eliminate cell clumps and aggregates before staining.

In the staining process, 2µL of the Vibrant DyeCycle Green (V35004, ThermoFisher) were added to each sample and incubated 30 min (37°C) for DNA labeling. Treatment with RNase was not necessary as Vibrant DyeCycle Green is a DNA-selective stain. Final stain concentration was 10 µM. After incubation, cells were washed and transferred to flow cytometry tubes for cell cycle analysis.

Flow Cytometry acquisition were performed with a BD FACS Canto II (BD Biosciences) where stained DNA were excited by a 488nm 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 platform that FlowJo provides, specifically the Watson pragmatic algorithm (Watson et al., 1987) to adjust the data to the model.

Pulse-Amplitude-Modulation Fluorometry

Fresh culture was harvested at the different specific times of the day. The samples were diluted 1:1 with growing medium. Pulse-Amplitude-Modulation fluorometry measurements were performed using a Waltz DUAL-PAM-100.

Debería meter los demás calculos o solo el que voy a representar?

After 10 min of relaxation at 20 ºC in total darkness, the non-actinic modulated light (450nm, 2.8 µE µE m^{-2} s^{-1}) was turned on in order to measure F_o (fluorescence basal level). To determine F_M (the maximum fluorescence level) in the sample adapted to darkness, a saturating red light pulse of 655nm and 5000 µE m^{-2} s^{-1} is applied to the sample during 400ms. The F_V/F_M, that corresponds to the maximum potential quantum efficiency of Photosystem II if all reaction centers were open, was calculated using this formula:

$$F_V / F_M = (F_M - F_O) / F_M$$

Starch Content Data Generation and Processing

At specific times of the day, fresh culture was harvested and lyophilized after a washing step with ammonium formiate to eliminate salts from the growing medium.

For starch extraction, approximately 2-3 mg of lyophilized biomass was added to an hermetic tube containing 1 mL of glass beads (0.25–0.3 mm diameter; Braun, Melsungen, Germany) and 2 mL of chloroform:methanol (2:1). To ensure cell disruption, three Mini-Beadbeater (BioSpec Products) cycles (60s agitation-60s incubation on ice) were applied to the samples. Then the liquid was separated from the beads and it was saved in a new tube. For the starch extraction, the disrupted cells were centrifuged for 4 min. at max. speed and the supernatant was discarded. The centrifugation was repeated until the pellets appeared white. Before continuing with the following steps, the pellets must be dry. 1ML of KOH 0.2M was added to the dry pellets and heated at 100 ºC. After 30 min, the tubes were gradually cooled and centrifuged at 13 rpm during 10 min. The supernatant is collected and mixed with 300 µL of acetic acid 1M (in order to adjust pH 5.0) and 200µL of destile water (in order to adjust concentration).

From that mix, 200 µL of each sample were placed in a new tubes with 7.4 U of α-amilase diluted in 35 µL of sodium acetate 0.1M pH 4.5 and incubated at 37 ºC for 30 min. After

that, 5U of amyloglucosidase diluted in 165 µL sodium acetate 0.1M pH 4.5 were added to the tubes and incubated at 55 ºC for 1-2 h. In order to inactivate the enzymes, the tubes were incubated at 100 ºC for 2 min and centrifuged at 13 rpm for 10 min.

Finally, for the spectrophotometric quantification, the following mix is added to quartz spectrophotometer cuvettes: 100 µL of the sample, 500 µL of hexoquinase buffer (?), 100 µL of ATP 10mM (diluted in Hepes 100 mM ph 7.7), 100 µL of NAD+ 4 mM (diluted in Hepes 100 mM ph 7.7), 5 U of glucose-6-phosphate dehydrogenase and 200 µL of destile water. The absorbance was determined twice at 340 nm. The first measure is followed by the addition of 5 U of hexoquinase and the incubation at 27 ºC for 20 min to ensure the efficiency of the reaction. The second measure was used to calculate the amount of NADH produced during the reaction, that is related to the initial amount of starch in the sample.

Como se calcula el almidon? Relacion 1:1 con el NADH?

Carotenoid Content Data Generation and Processing

For carotenoids extraction, total darkness is needed. Approximately 4 mg of lyophilized biomass was added to an hermetic tube containing 1 mL of glass beads (0.25–0.3 mm diameter; Braun, Melsungen, Germany) and 1 mL of acetone (% o pureza?). To ensure cell disruption, three Mini-Beadbeater (BioSpec Products) cycles (60s agitation-60s incubation on ice) were applied to the samples. Then they were centrifuged for 4 min. at max. speed and the supernant was saved in a new tube. This proccess was repeated six times or until the supernant turns colorless. Once all the supernants of each sample were collected in the same tube, acetone were evaporated using a stream of nitrogen gas which is directly blown in the solutions. Finally, 350 µL of acetone were added to each tube, and if was distributed in HPLC tubes following manufacturer's instructions.

A Gynotech HPLC, equipped with a photodiode-array detector was used. (Me falta la columna usada)Separation was performed on a Waters Nova-Pak C18 column (3.9×150 mm, 4 µm particle size, 60 Å pore size) containing dimethyloctadecylsilyl-bonded amorphous silica, protected with a guard cartridge (4 µm particle size, sealed with 2-µm filters). The eluents used to create a gradient through the mobile phase were: eluent A (ammo-

nium acetate 0.1 M and H_2O-methanol 15:85 v/v) and eluent B (methanol-acetonitrile-acetone 44:43:13 v/v). Temperature was maintained constant (20 ºC) during the whole process and eluents flow at 800 µL min^{-1}. Mamiellales carotenoids (cita q no encuentro).

Different carotenoids were identified following retention times and absorption profile (esto se puede decir o hay q citar?) and quantified as a percentage of the total peak area.

Rhythmic Patterns Detection and Characterization

The R package RAIN (Thaben & Westermark, 2014) from Bioconductor was used to statistically identify rhythmic patterns in the different data collected. A 0.05 p-value threshold was used in all the cases under study. Rhythmic patterns with a single peak over a complete diurnal cycle were detected by setting the period parameter from RAIN to 24 hours. A similar process was used to detect more complex rhythmic patterns (the ones with two or even three peaks), but changing the period parameter from RAIN to 12 and 8 hours respectively.

To ensure the accuracy of our study, three complete diurnal cycles from both photoperiods 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, amount of cells in the different cell cycle phases and amount of different carotenoids and starch.

Moreover, in order to differentiate genes and proteins that respond to light from the ones that are actually regulated by the circadian clock, the last two diurnal cycles and two consecutive days of continuous light were considered for the RAIN analysis described above. In that way, RAIN can statistically test if a similar pattern is maintained after changing the cycling light regime to a continuous light input and it prevents a bias towards any of the two conditions.

Once that rhythmic patterns in our data were detected by RAIN, they were fitted to a co-sinusoidal wave in order to be characterized and to enable comparison between them. Rhythmic patterns are mathematically treated as waves, so they can be characterized according to the classic parameters of mesor (the mean values around which this wave oscil-

lates), amplitude (half the difference between the highest and the lowest value in the wave) and phase (when does the wave reach its highest). 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”.

Hypothesis and Objetives

(texto de hipotesis)

Results

Chapter 1. ALGAEFUN with MARACAS: user-friendly tool for analysing and integrating omic data generated from microalgae.

In order to contribute to the characterization of the molecular systems regulating microalgae physiology, high throughput sequencing technologies have been recently applied to obtain the genome of a wide range of microalgae (cita todas las microalgas 7-19 en paper). This has promoted the emergence of molecular systems biology studies and the use of different omics like transcriptomics based on RNA-seq data (cita klebs y haematococcus) and cistromics based on ChIP-seq data (paper 23, 24) in microalgae. Nonetheless, the progress of this type of studies on microalgae are limited by the lack of freely available and easy-to-use online tools to analyze, extract relevant information 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 more 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.

In order to cover these microalgae research community needs and promote studies in molecular systems biology we have developed the web portal ALGAEFUN with MARACAS using the R package Shiny (cita) and other Bioconductor (cita) packages.

Our web portal consists of two different tools. First, MARACAS (MicroAlgae RnA-seq and Chip-seq AnalysiS) implements a fully automatic computational pipeline receiving as input RNA-seq or ChIP-seq raw data from microalgae studies. MARACAS generates sets of differentially expressed genes or lists of genomic loci for RNA-seq and ChIP-seq analysis respectively. These results can be further analyzed using our second tool ALGAEFUN (microAlgae FUNctional enrichment tool). On the one hand, when receiving the results from an RNA-seq analysis, sets of genes are functionally annotated by performing GO (Gene Ontology) (cita) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (cita) enrichment analysis. On the other hand, 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 and https://github.com/fran-romero-campero/MARACAS.

Implementation

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. In *Nature* (Vol. 583, Issue 7818). https://doi.org/10.1038/s41586-020-2449-8

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

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. In *Annual Review of Plant Biology* (Vol. 73). https://doi.org/10.1146/annurev-arplant-071921-100530

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

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. In *Trends in Plant Science* (Vol. 19, Issue 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

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

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. In *Trends in Biotechnology* (Vol. 39, Issue 5, pp. 433–437). Elsevier Ltd. https://doi.org/10.1016/j.tibtech.2020.09.002

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

Coleto-Alcudia, V., & Vega-Rodríguez, M. A. (2020). Artificial Bee Colony algorithm based on Dominance (ABCD) for a hybrid gene selection method. *Knowledge-Based Systems*, *205*. https://doi.org/10.1016/j.knosys.2020.106323

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

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

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

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? In *Molecular Microbiology* (Vol. 62, Issue 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. (2017). Evolution of Daily Gene Co-expression Patterns from Algae to Plants. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2017.01217

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

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., Wuyts, 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., Grimbaldeston, 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 Open Research*, *7*(2), 00104–02021. https://doi.org/10.1183/23120541.00104-2021

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

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

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.

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

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

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

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

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

Jamers, A., Blust, R., & De Coen, W. (2009). Omics in algae: Paving the way for a systems biological understanding of algal stress phenomena? In *Aquatic Toxicology* (Vol. 92, Issue 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. In *Nature Reviews Molecular Cell Biology* (Vol. 7, Issue 3). https://doi.org/10.1038/nrm1857

Karahalil, B. (2016). Overview of Systems Biology and Omics Technologies. *Current Medicinal Chemistry*, *23*(37), 4221–4230. https://doi.org/10.2174/0929867323666160926150617

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

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

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

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

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), 2060–2070. 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

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

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

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

McClung, C. R. (2006). Plant circadian rhythms. In *Plant Cell* (Vol. 18, Issue 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

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. In *EMBO Reports* (Vol. 6, Issue 10). https://doi.org/10.1038/sj.embor.7400541

Moreau, H, Grimsley, N. Derelle, E, Ferraz, C, Escande, ML, 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.

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

Noordally, Z. B., & Millar, A. J. (2015). Clocks in algae. *Biochemistry*, *54*(2). 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

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

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

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

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

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

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

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

Rock, A., Wilcockson, D., & Last, K. S. (2022). Towards an Understanding of Circatidal Clocks. In *Frontiers in Physiology* (Vol. 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*, *April*, 1–14. https://doi.org/10.1111/jsr.13621

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

Roenneberg, T., & Merrow, M. (2016). The circadian clock and human health. In *Current Biology* (Vol. 26, Issue 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. In *Biology* (Vol. 8, Issue 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

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

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*, 35677956. 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

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*(SUPPL. 1). https://doi.org/10.1093/nar/gkm965

Takahashi, J. S. (2021). The 50th anniversary of the konopka and benzer 1971 paper in PNAS: “Clock mutants of drosophila melanogaster.” In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 118, Issue 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

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

Veenstra, T. D. (2021). Omics in Systems Biology: Current Progress and Future Outlook. In *Proteomics* (Vol. 21, Issues 3–4). https://doi.org/10.1002/pmic.202000235

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

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. In *Journal of Proteomics* (Vol. 75, Issue 1). https://doi.org/10.1016/j.jprot.2011.07.010

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

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

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

Zee, P. C., & Abbott, S. M. (2020). Circadian Rhythm Sleep-Wake Disorders. In *CONTINUUM Lifelong Learning in Neurology* (Vol. 26, Issue 4). https://doi.org/10.1212/CON.0000000000000884

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. In *Molecular Plant* (Vol. 9, Issue 12). https://doi.org/10.1016/j.molp.2016.09.014

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