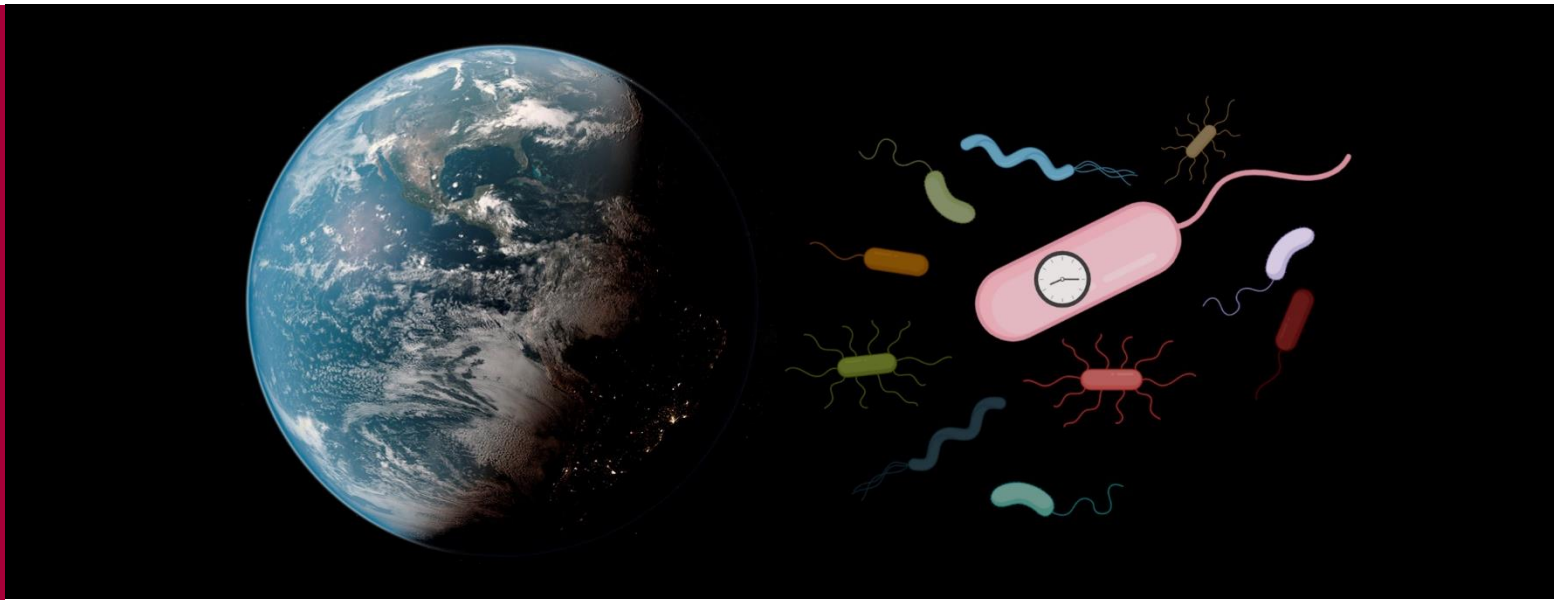
PhD Thesis
August 2022

Diel cycle and circadian clocks in Prokaryotes

Augustin Géron



UMONS
Université de Mons

UNIVERSITY of STIRLING

“E pur si muove !”

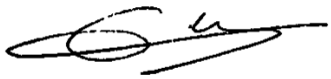
Galileo Galilei (1564-1642)

Certificate of originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at the University of Mons and the University of Stirling or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at the University of Mons and the University of Stirling or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Augustin GERON



Abstract

Diel cycle imposes daily oscillations in environmental conditions, which temporally structures most ecosystems. One of the most outstanding adaptations to this recurrent phenomenon is probably the emergence of time-keeping mechanisms – circadian clocks – that allow the organisms to synchronize their biological activities with daily variations. While circadian clocks are ubiquitous in Eukaryotes, they are so far only characterized in *Cyanobacteria* within Prokaryotes. However, growing evidence suggests that circadian clocks are widespread in the bacterial and archaeal domains. Studying the impact of diel cycle on Prokaryotes and deciphering their time-keeping mechanisms is of enormous biological importance as bacteria and archaea are crucial in all ecosystems and are essential to human health. Understanding their daily rhythms will undoubtedly provide numerous valuable insights in medical research, environmental sciences, and biotechnology.

In this thesis, we assessed for the first time the impact of diel cycle on the functionality of marine picoplanktonic communities using a metaproteomic approach. Metaproteomics is a powerful tool that allows pertinent establishment of phenotype-genotype linkages, but despite its rapid development, this technology still faces many technical challenges that hamper its potential power. In this context, we developed *mPies*, a novel bioinformatic tool that addresses critical bioinformatic issues in the environmental metaproteomics, and we showed the importance of diversifying the experimental workflow for comprehensive metaproteomic studies.

This methodological optimization helped us to compare the *in-situ* diel metaproteomic patterns of picoplanktonic communities sampled from the surface north-west Mediterranean Sea. Our data showed a taxa-specific response to diel cycle in the photoautotroph *Synechococcales* and in (photo)-heterotrophic bacteria such as *Flavobacteriales*,

Pelagibacterales and *Rhodobacterales*. Next, we investigated the taxonomic distribution of circadian clock proteins in Prokaryotes, and we updated the list of bacteria and archaea of great ecological and industrial relevance that could potentially be clock-controlled.

Based on this review, we choose to study the diel cycle impact and the circadian genes expression of the biotech promising *Rhodospirillum rubrum*. Our results showed that light and dark cycles induced broad regulation of its proteome and affected the regulation of *kai* genes at both transcriptional and translational levels. Overall, this work evidenced that diel cycle impacts the functionality of prokaryotes and reinforced the hypothesis of time-keeping mechanisms beyond *Cyanobacteria*.

Acknowledgements

I am first very grateful to my supervisors Sabine Matallana-Surget and Ruddy Wattiez who trusted me, supported me, and provided me guidance and all the resources that I needed during my PhD under a cotutelle agreement between the University of Mons (Belgium) and the University of Stirling (Scotland). I am grateful to Ruddy Wattiez who first helped me to do an internship in China during my master's degree in Biological Sciences. It was my first experience abroad and the beginning of a trusting relationship between us. Then, it was he who convinced me to further pursue a PhD. Thank you for your support and guidance. I am grateful to Sabine Matallana-Surget, who welcomed me in Scotland and made sure I was in the best working conditions while I was abroad. She provided me with regular feedback, which allowed me to get better every day. I thank her for her constant support and guidance at all the stages of my PhD's journey.

I thank the examining committee members Pr. David Gillan (President), Pr. Tim Whalley (Independent Chair), A/Prof. Joseph Christie-Oleza (External reviewer), Dr. Nico Jehmlich (External reviewer), and Dr. Jérôme Delroisse (Secretary) for evaluating my thesis and improving it with pertinent and enriching comments, suggestions, and corrections.

I am also thankful to Johannes Werner with who I collaborated during my thesis. I learned a lot while working with him on the development of *mPies* and gained precious knowledge in bioinformatic. I also thank Corentin Decroo for helping me with the proteomic analyses of *Rhodospirillum rubrum*.

I thank all present and former lab's members with who I worked in Mons and Stirling. I am especially thankful to the Proteomic and Microbiology laboratory technicians Catherine 's Heeren, Cyril Mascollo and Giuseppe Giambarresi who were always available to help me when I needed it. I thank Valentine Cyriaque, who supervised me during my master's thesis.

I am grateful to my parents, my sister and brother who always believed in me. I also thank Anna Piperi who provided me with unconditional love and support. Finally, I thank all my friends inside and outside the academic field who always encouraged me during this adventure.

Publications and international conferences

1. Publications arising from this work

Géron, **A**., Werner, J., Wattiez, R., & Matallana-Surget, S. Towards the Discovery of Novel Molecular Clocks in Prokaryotes. Critical Review in Microbiology, *Under review*.

Géron, **A**., Werner, J., Lebaron, P., Wattiez, R., & Matallana-Surget, S. (2021). Diel Protein Regulation of Marine Picoplanktonic Communities Assessed by Metaproteomics. *Microorganisms*, 9(12), 2621.

Géron, **A**., Werner, J., Wattiez, R., Lebaron, P., & Matallana-Surget, S. (2019). Deciphering the functioning of microbial communities: shedding light on the critical steps in metaproteomics. *Frontiers in microbiology*, 10, 2395.

Werner, J., **Géron**, **A**., Kerssemakers, J., & Matallana-Surget, S. (2019). *mPies*: a novel metaproteomics tool for the creation of relevant protein databases and automatized protein annotation. *Biology Direct*, 14(1), 1-5.

2. Publication in preparation

Géron, **A**., Decroo C., Matallana-Surget, S & Wattiez, R. Diel regulation of proteome and circadian genes in *Rhodospirillum rubrum*.

3. International conferences

Géron, **A**., Werner, J., Wattiez, R., Lebaron, P., & Matallana-Surget, S. Bioinformatic workflow in environmental metaproteomics. Oral communication at 3rd International Metaproteome Symposium (ISM3), Leipzig, Germany, 2018.

Géron, **A**., Werner, J., Wattiez, R., Lebaron, P., & Matallana-Surget, S. Diel Cycle in Picoplankton Communities: Structure and Function Assessed by Metaproteomic. Poster presentation at 17th International Symposium on Microbial Ecology (ISME17), Leipzig, Germany, 2018.

Géron, **A**., Werner, J., Wattiez, R., Lebaron, P., & Matallana-Surget, S. Deciphering the functioning of microbial communities: shedding light on the critical steps in metaproteomics. Poster presentation at Federation of European Microbiological Societies (FEMS19), Glasgow, Scotland, 2019.

Géron, **A**., Wattiez, R., Matallana-Surget, S. Day and Night Cycle of Marine Picoplankton Communities. National final of Thesis in 180 seconds, Belgium, Liege, Belgium, 2021.

Table of content

Table of figures

Table of tables

List of appendices

List of abbreviations

1D: One dimensional

2D: Two dimensional

AAPB: Aerobic Anoxygenic Photoheterotrophic Bacteria

ABC: ATP Binding Cassette

ADP: Adenosine diphosphate

AM-DB: Assembled Metagenome-derived Database

ATP: Adenosine Triphosphate

BLAST: Basic Local Alignment Search Tool

COG: Clusters of Orthologous Groups

Comb-DB: Combined database

CP: Crossing point

Cys: Cysteine

D: Day

DB: Database

DDA: Data Dependent Acquisition

DIC: Dissolved Inorganic Carbon

DNA: Deoxyribonucleic Acid

DOC: Dissolved organic carbon

DOM: Dissolved organic matter

DTE: Dithioerythritol

FDR: False discovery rate

GADPH: Glyceraldehyde 3 Phosphate Dehydrogenase

Gln: Glutamine

Glu: Glutamic acid

HPLC: High Performance Liquid Chromatography

IAA: Iodoacetamide

ICC: Integrated ion-current

ID: Identity

IR: Infrared radiation

kDa: kilodalton

KEGG: Kyoto Encyclopedia of Genes and Genomes

LCA: Last common ancestor

LD: Light and Dark

LL: Continuous Light

Mbp: Millions of base pairs

MRM-MS: Multiple Reaction Monitoring (MRM-MS).

mRNA: messenger Ribonucleic Acid

MS: Mass spectra

Mya: Million years ago

N: Night

NADP: Nicotinamide Adenine Dinucleotide Phosphate

NAM-DB: Non-Assembled Metagenome-derived Database

NCBI: National Center for Biotechnology Information

ncRNA: Non-coding Ribonucleic Acid

NS: Non-Sulfur

NW: North-West

OD: Optical Density

OSD14: Ocean Sampling Day 2014

OTU: Operational Taxonomic Unit

P: Phosphorous

PAR: Photosynthetically available radiation

PBS: Phosphate-buffered saline

PCA: Principal component analysis

Pcb: Prochlorophyte chlorophyll-binding protein

PCR: Polymerase Chain Reaction

PER: Photo-enzymatic repair

PES: Polyethersulfone

POC: Particulate organic carbon

POM: Particulate organic matter

PSII: Photosystem II

Psu: Practical salinity unit

RNA: Ribonucleic Acid

ROS: Reactive oxygen species

rRNA: Ribosomal Ribonucleic Acid

RT: Retention time

RT-qPCR: Reverse Transcription quantitative Polymerase Chain Reaction

RuBisCO: Ribulose bisphosphate carboxylase/oxygenase

S: Sulfur

SDS-PAGE: Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

SMN: Supplemented malate-ammonium medium

SOMLIT: Service d'Observation en Milieu Littoral

sp.: species

SRB: Sulfate-reducing bacteria

SWATH-MS: Sequential Window Acquisition of All Theoretical Mass Spectra

TAX-DB: Taxonomy-derived database

TCA cycle: Tricarboxylic Acid cycle

TFA: Trifluoroacetic Acid

UVA: Ultraviolet A

UVB: Ultraviolet B

UVR: Ultraviolet radiation

XICs: Extracted ion chromatograms

Chapter 1: General introduction

1. Background

Chronobiology examines the effects of time on biological systems. On a daily basis, organisms experiment recurrent phenomena caused by the Earth's rotation around its axis, known as the diel cycle. Diel cycle is of enormous biological importance as it temporally structures the ecosystems. The most outstanding adaptation to the diel cycle is probably the emergence of circadian clocks (Bhadra et al., 2017; Saini et al., 2019). Circadian clocks are endogenous molecular program that synchronize the biological activities of the organisms according to the diel cycle (Saini et al., 2019). Despite their central role in controlling biological functions, circadian rhythms were only characterized recently. In fact, the description of the first circadian clock mechanism in the fruit fly *Drosophila* was made in the early 1980s by Hall and Rosbash (Brandeis University), and Young (Rockefeller University) and colleagues who received the 2017 Nobel Prize in Physiology or Medicine to honor their discovery (Huang, 2018). Nowadays, circadian clocks have been documented in all type of organisms. However, while they are ubiquitous in Eukaryotes, they were so far only characterized in *Cyanobacteria* in Prokaryotes (Johnson & Rust, 2021).

Exploring the impact of diel cycle on Prokaryotes is of high relevance as they are the most abundant ($>10^{30}$ cells) and diversify organisms on Earth and are key players in all ecosystems (Whitman et al., 1998). In this thesis, we focused on the Prokaryotes inhabiting the surface of the oceans, where they carry out about half of primary production and the bulk of life-sustaining nutrient cycling (Azam et al., 1983, Field et al., 1998). While the interaction of light with photoautotrophs (e.g., *Cyanobacteria*) is well documented, little is known about the diel cycle of heterotrophs and its impact on the whole marine microbial communities

(Morimoto et al., 2020). Nowadays, with the development of omics techniques, we are expanding our understanding of microorganisms in their natural environment and at multiple temporal scales (Faust & Raes, 2012). From all the molecular tool available to study complex microbial assemblages, metaproteomics stands out as a valuable approach that allows the characterization of the active members and infer on their functions (Wilmes & Bond, 2004). Yet, this technology still faces many technical challenges that hamper its potential power and need to be addressed (Heyer et al., 2017).

Besides the study of diel wide-community rhythmicity, more targeted approaches focusing on potential clock-controlled prokaryotes will undoubtedly provide great opportunities of ecological and industrial relevance (Sartor et al., 2019). In this context, the purple non-sulphur bacteria are promising as several species possess circadian clock gene homologs (Dvornyk et al., 2003) and have remarkable use in biotechnology (De Meur et al., 2020). However, little is known about their diel rhythmicity.

This thesis focusses on three key aspects of chronobiology in Prokaryotes: i) the optimization of the metaproteomic workflow, which is essential to better understand the microbial community functional response to environmental changes; ii) the diel protein regulation of the surface picoplanktonic community of the North-West Mediterranean Sea; and iii) the exploration of the diel cycle response in new potential clock-controlled bacteria.

Therefore, this general introduction will provide the state-of the art on i) the circadian rhythms in Prokaryotes, ii) the diel cycle in the marine photic zone, and iii) the metaproteomic approach to study complex microbial assemblages.

2. Diel cycle and circadian clock in Prokaryotes

Environmental parameters vary over different timescales such as seasons, weeks, or day and organisms adapt subsequently (Fuhrman et al., 2015). While long temporal scale (monthly to weekly) has been the most common study scheme in chronobiology, diel biological responses is receiving more attention since the discovery of biological clocks (Bhadra et al., 2017). The Earth's rotation around its axis induces a 24h-periodic oscillation of solar radiation reaching a given point on the planet. This phenomenon is known as the diel cycle (or diurnal cycle) and causes light and temperature fluctuation over the course of the day and consequently, diel rhythmic changes in all type of organisms (Saini et al., 2019).

These rhythms are called circadian rhythms (from the Latin *circa*, meaning "around" and *diem*, meaning "day") only if they refer to an endogenous process entrained by the environment. Circadian rhythms, in contrast with diel rhythms, are controlled by a circadian clock, an endogenous molecular mechanism embedded in an input and an output pathway that regulates biological processes (Kondo & Ishiura, 2000). Circadian rhythms are characterized by three main criteria: i) they are free-running in constant conditions, which means that they persist in the absence of environmental signals; ii) they are entrained by environmental clues and can be reset with the appropriate stimuli (zeitgeber or time-giver); and iii) they display temperature compensation, which means that their rhythm is not altered by temperature modification (Golden & Canales, 2003). Circadian clocks allow the organisms to anticipate daily environmental changes and regulate with accuracy key functions (e.g., photosynthesis), which provides them significant fitness advantage over competitors (Sharma, 2003; Salmela & Weinig, 2019).

As prokaryotes are key players in all ecosystems, a better understanding of their adaptation to diel environmental changes will not only expand our fundamental knowledge in microbiology but will provide translational opportunities in the medical, industrial, and ecological field (**Figure 1.1**). Prokaryotes are often embedded in symbiotic or parasitic association with clock-controlled Eukaryotes such as humans or plants (Lodewyckx et al., 2002; Madsen, 2011; McFall-Ngai et al., 2013). For instance, the human's clock control circadian rhythms that were found to impact its microbiome (Deaver et al., 2018; Wu et al., 2018; Godinho-Silva et al., 2019; Nobs et al., 2019; Saran et al., 2020) as well as being interlinked to bacterial and virus infection or diseases (Thaiss et al., 2014; Kaczmarek et al., 2017; Heinemann et al., 2021; Choi et al., 2021). Similarly, plants have circadian molecular programs that regulate their metabolism and generates a rhythmic environment for the root-associated microorganisms (Asif et al., 2019). In this context, deciphering the interactions between diel rhythms of the host and microbial symbiotes or pathogens could lead to more efficient and sustainable healthcare and agricultural strategies (Maddur Puttaswamy, 2019; Asif et al., 2019; Pearson et al., 2021).

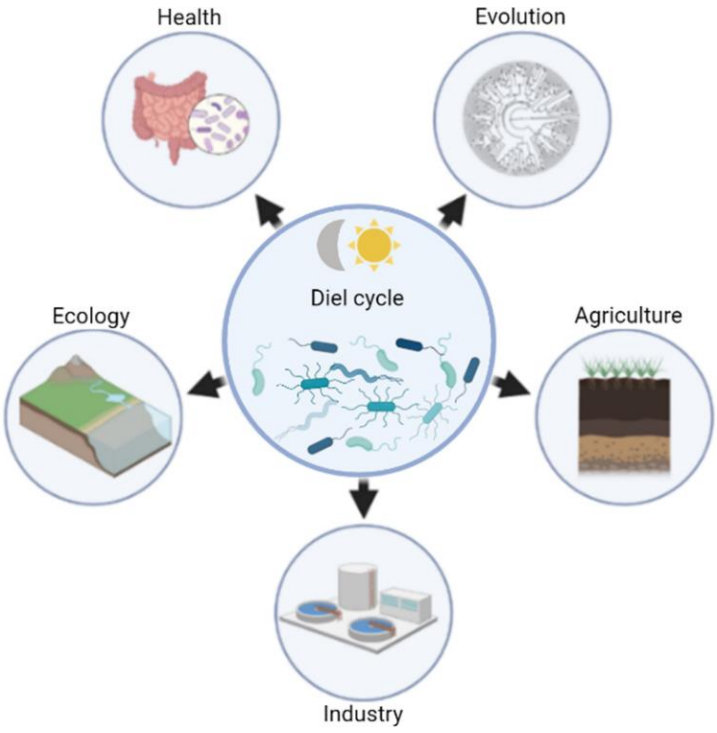


Figure 1.1: Schematic view on the rationales and impacts of studying diel cycle in Prokaryotes.

The research focus on diel microbial community dynamics will also provide valuable insights to improve the current ecological models and to better understand their impact on the environments (Morimoto et al., 2020). Indeed, microbiological studies were mainly performed during the day and consequently, the biological activity of microorganisms during the night is not well documented. Deciphering the diel changes of microbial communities in the context of ecological processes (e.g., biogeochemical cycles) or environmental stress (e.g., pollution) will offer a more complete view on the functioning of ecosystems. In addition, the integration of temporal programs into industrial processes could also offer new opportunities and lead to improved biotechnological use of Prokaryotes (Sartor et al., 2019).

2.1. Circadian clocks in *Cyanobacteria*

Until the 1980s, the existence of circadian rhythms in prokaryotes seemed improbable because they were thoughts to have neither the need nor the resource to have biological clock (Edmunds, 1983). Around the mid-1980s, diel rhythmicity in nitrogen fixation was observed in *Cyanobacteria* (Mitsui et *al*., 1987; Stal & Krumbein, 1987) (**Figure 1.2**). Huang and colleagues were the firsts to demonstrate that these rhythms matched with all the criteria of a true circadian rhythm in the cyanobacterial strain *Synechococcus* sp. RF-1 (Huang et *al*., 1990; Chen et *al*., 1991). This discovery reversed the eukaryotic exclusive circadian clock dogma and paved the way to intense investigations on prokaryotic circadian clock mechanisms (Cohen & Golden, 2015) (**Figure 1.2**). Rapidly, the strain *Synechococcus elongatus* PCC 7942 emerged as the model organism for studying the mechanism of the cyanobacterial clock (Kondo et al., 1993). Besides, *S. elongatus*, clock mechanisms were described in other *Cyanobacteria* and display functional diversity. In the next three sections, we will describe the functioning of the circadian clocks of *Synechococcus*, *Prochlorococcus*, and *Synechocystis*.

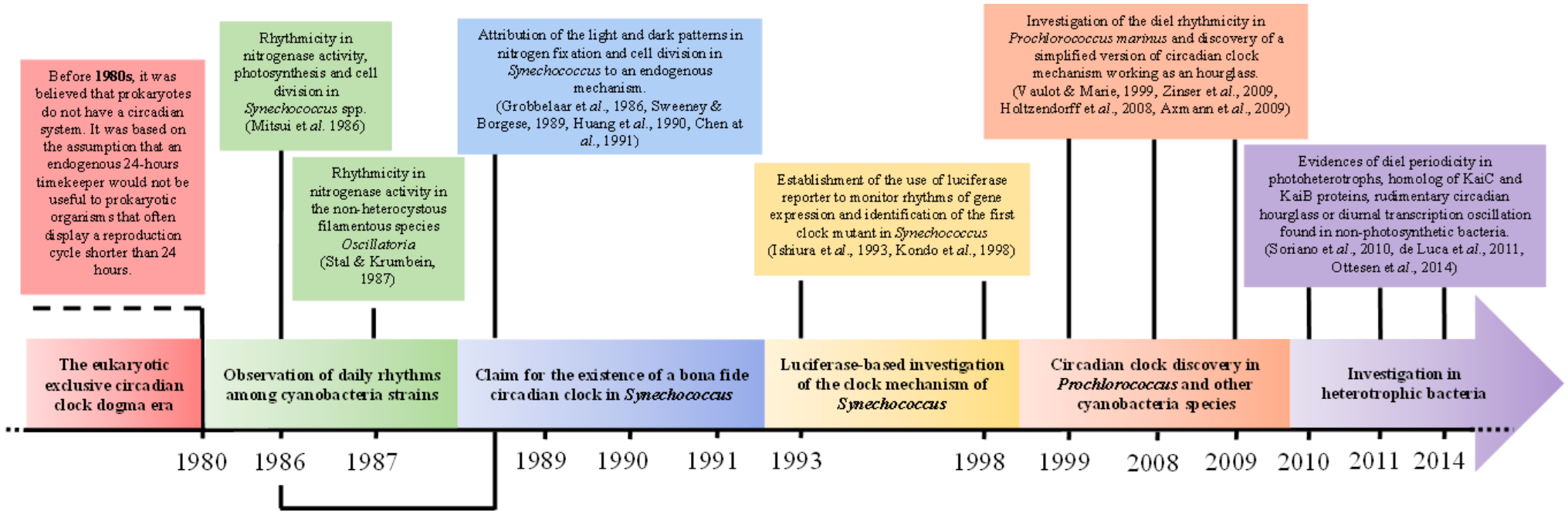


Figure 1.2: Timeline of the discovery of circadian clocks in Prokaryotes.

2.1.1. The circadian clock of *Synechococcus elongatus* PCC 7942

The circadian clock of *Synechococcus elongatus* PCC 7942 is based on a core oscillator that consists in the protein complex KaiA, KaiB, and KaiC (**Figure 1.3**). KaiC is the key player of the clock and undergoes an autophosphorylation (promoted by KaiA) and dephosphorylation (promoted by KaiB) cycle over the course of the day (Nishiwaki et al., 2000; Nishiwaki et al., 2004; Nishiwaki et al., 2012). The Kai core oscillator interacts with an input and an output pathway and determines diel rhythmicity in gene expression, metabolic switches, or timing of cell division (Dong et al., 2010; Mori et al., 1996; Pattanayek et al., 2014; Puszynska and O'Shea 2017) (**Figure 1.3**).

Over the course of the day, unphosphorylated KaiC has a loose structure and expose A-Loops where KaiA can bind and promote KaiC autophosphorylation (Kim et al., 2008) (**Figure 1.3**). Once fully phosphorylated, KaiC becomes stiff and stacked and exposes B-Loops where KaiB can bind and sequester KaiA, subsequently activating the KaiC autophosphatase (Kitayama et al., 2003; Chang et al., 2012; Tseng et al., 2014). By the end of the night, KaiC returned to its unphosphorylated state with exposed A-loops and can bin KaiA again (Nishiwaki et al., 2012). The mechanisms involved in the entrainment of the Kai oscillator are still not entirely understood (Kim et al., 2020). However, studies have shown that environmental signals such as light and temperature were transmitted to the Kai oscillator by an input pathway (**Figure 1.3**). In *S. elongatus* this pathway is composed of CikA, LdpA, KaiA and Pex. CikA, LdpA, and KaiA are sensitive to the oxidation state of the quinone pool which changes over the diel cycle, and affect the kinase and phosphatase activity of KaiC, while the *pex* gene is light suppressed and encodes a transcription repressor of the *kaiA* gene (Axman et al., 2014). Moreover, the Kai oscillator itself sense the [ATP/ADP] ratio in the cell, which changes over the day-night cycle (Axman et al., 2014). During the phosphorylation/dephosphorylation cycle

of the KaiC, the clock oscillator transmits signals through an output pathway that subsequently control gene expression. In *S. elongatus*, these output components include SasA, RpaA, and CikA, which as a dual role in the input and the output pathways. During the day, SasA interacts with KaiC and phosphorylated RpaA (Iwasaki et al., 2000; TaKai et al., 2006). At night, the KaiBC complex binds CikA and dephosphorylate RpaA (Chang et al., 2015). As phosphorylated RpaA acts as a transcription factor that induces the expression of Class I genes and repress Class II genes, the phosphorylation/dephosphorylation cycle of RpaA is responsible for the expression/repression of patterns observed in *S. elongatus* (Markson et al., 2013).

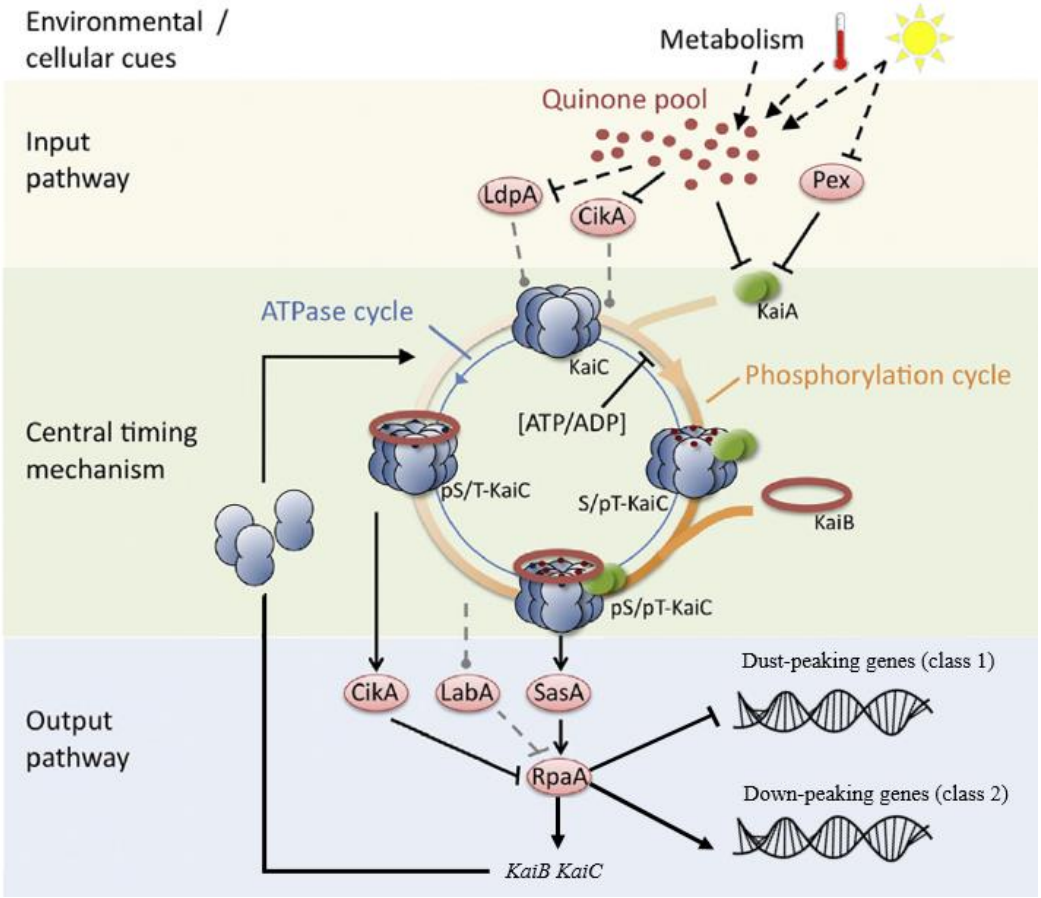


Figure 1.3: Diagram of the circadian clock mechanism in *S. elongatus* PCC 7942. KaiC undergoes a phosphorylation and dephosphorylation cycle over the course of the day, which is promoted by KaiA and KaiB, respectively. The input pathway is composed of LdpA, CikA and Pex which interact with the Kai oscillator in response to environmental signals like light and temperature. In addition, the ATP/ADP ratio, which changes according to diel cellular activity, is sensed by the Kai oscillator itself. Subsequently, the Kai oscillator interacts with the output pathway that is composed of CikA and SasA which modifies the phosphorylation state of RpaA. Phosphorylated RpaA activates Class I genes and represses Class II genes. Modified from Axmann et al., 2014.

The circadian clock of *S. elongatus* has a widespread control of gene expression as 30 to 60% of the transcripts are rhythmically regulated (Ito et al., 2009; Vijayan et al., 2009). For example, the rhythm of nitrogenase activity is controlled at the transcriptional level and the expression of the *nif* and *nif*-associated genes occurs mainly during the dark period (Huang & Lin, 2009). In addition, the clock also regulates compaction and superhelical status of the chromosomes (Smith & Williams, 2006; Woelfle et al., 2007) and the timing of cell division by gating it through the course of the day (Mori & Johnson, 2001). Interestingly, it was shown that the circadian clock disallows the cell division in the onset of darkness even if cells have attained a sufficient size to divide (Mori & Johnson, 2001). It was also demonstrated that the circadian clock maintains its periodicity even if cells are dividing several times during the day (Mori et al., 1996; Mori & Johnson, 2001).

2.1.2. The simplified clock of *Prochlorococcus*

Prochlorococcus marinus also shown diel rhythmic gene expression (Zinser et al., 2009; Waldbauer et al., 2012), cell cycling (Vaulot et al., 1995; Vaulot & Marie, 1999; Holtzendorff et al., 2008), protein accumulation (Waldbauer et al., 2012), amino acid uptake (Mary et al., 2008) and photosynthesis (Zinser et al., 2009) in natural populations and laboratory cultures. In contrast to *Synechococcus*, the rhythmicity of *Prochlorococcus* does not persist in continuous conditions and requires daily environmental input to reset (Holtzendorff et al., 2008; Cohen & Golden, 2015). In fact, the clock mechanism of *Prochlorococcus* lacks several molecular components such as KaiA or CikA and KaiC compensates with an enhanced autophosphorylation activity (Axmann et al., 2009) (**Figure 1.4**).

The functioning of the clock is based on the KaiBC complex formation, which allows SasA to bind with KaiB and induces the phosphorylation of RpaA (Axmann et al., 2009). Phosphorylated RpaA regulates the transcription of KaiBC which controls the clock expression (**Figure 1.4**). The simplified clock of *Prochlorococcus* still provides genome-wide expression rhythms (Zinser et al., 2009) and control cell division and other biological processes such as photosynthesis. It was suggested that *Prochlorococcus* evolved towards a reduced KaiBC-system because of the high regularity of its habitat (the near-equatorial oceans), where a free-running clock system may not be essential because of the high regularity of diel conditions (Axmann et al., 2009; Mullineaux & Stanewsky, 2009).

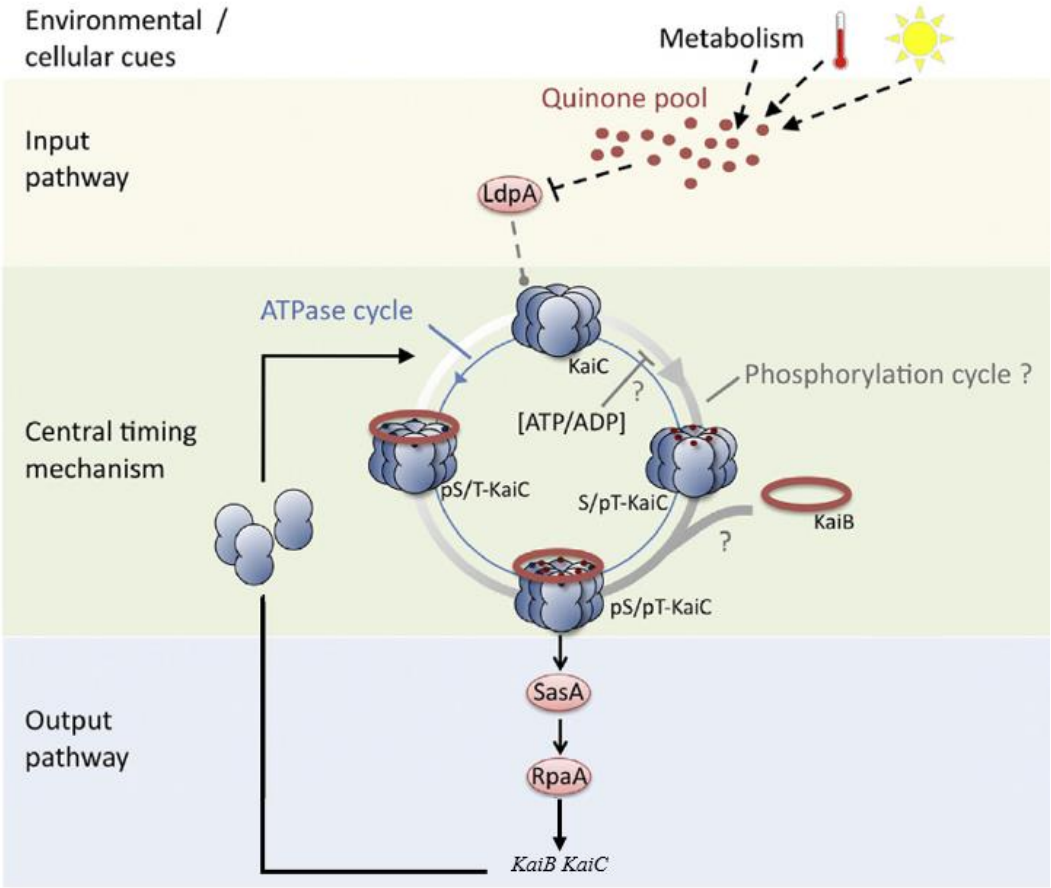


Figure 1.4: Simplified diagram of the circadian clock mechanism in *P. marinus* MED4. The time keeping mechanism of *P. marinus* MED4 lacks the central components from the *S. elongatus* clock KaiA and CikA. The input pathway is composed of LdpA only. In the absence of KaiA, KaiC compensates with an enhanced autophosphorylation activity. The KaiBC complex interacts with SasA that regulates RpaA phosphorylation level, which further controsl KaiBC expression. Modified from Axmann et al., 2014.

2.1.3. The multiple clocks of *Synechocystis* sp. PCC 6803

Despite the observation of transcriptional circadian rhythms (Beck et al. 2014) and circadian rhythmicity of photosynthesis (Yen et al., 2004), the circadian oscillations in *Synechocystis* sp. PCC 6803 are not as robust as in *Synechococcus elongatus* PCC 7942 and are still inconclusive (Schmelling et al., 2021). In contrast with *Prochlorococcus*, *Synechocystis* sp. PCC 6803 encodes a total of seven kai genes: *kaiAB1C1* and *kaiB2C2* located in clusters, and *kaiB3* and *kaiC3* as orphan genes (Wiegard et al. 2013). KaiA, KaiB1, and KaiC1 proteins are thought to constitute the core oscillator as it was demonstrated that the autophosphorylation of KaiC1 depends on KaiA (Wiegard et al. 2013). In addition, the deletion mutant of *kaiAB1C1* which shows decreased viability in light and dark cycles, which suggests a linkage between metabolic pathways and circadian oscillation (Dörrich et al. 2014) The role of KaiB2C2, KaiB3 and KaiC3 is less clear it was hypnotized that they could be imply in the fine-tuning of the core oscillator or in other unknown functions (Aoki and Onai 2009; Dörrich et al. 2014; Wiegard et al. 2020).

2.2. Circadian rhythms beyond *Cyanobacteria*

Computational analyses revealed that Kai proteins homologs were widespread within Prokaryotes (Dvornyk et al., 2003; Loza-Correa et al., 2010; Schmelling et al., 2017) (**Figure 1.5**). Although input and output factors are often missing in non-cyanobacterial species, it suggests that KaiC-based timing systems, embedded into different networks, could exist beyond *Cyanobacteria* (Schmelling et al., 2021). KaiC homologs were observed with diel rhythmicity in several extremophile archaeal species. The hyperthermophiles *Thermococcus litoralis* and *Pyrococcus horikoshii* displayed a kinase activity of KaiC homologs independently on temperature (Schmelling et al. 2017). Rhythmic transcriptional expression KaiC homologs was also observed in *Haloferax volcanii* (Maniscalo et al., 2014). In addition, the halophilic genus

Halobacterium salinarum NRC-1 exhibited diel transcriptional oscillation in up to 12% of all genes (Whitehead et al., 2009). On the other hand, KaiC and KaiB homologs were more frequently observed in Bacteria (Dvornyk et al., 2003). Purple bacteria, which are capable of photosynthesis, are particularly good candidates for potential new clock mechanisms. *Rhodobacter sphaeroides* presents a reduced KaiBC-based system that was suggested to drive circadian oscillations (Min et al. 2005). *Rhodopseudomonas palustris* show autophosphorylation activity of KaiC homologs and displayed adaptive growth under rhythmic conditions (Ma et al., 2016). Finally, rhythmic activity at the level of the uptake hydrogenase (Hup) was demonstrated in *Rhodospirillum rubrum* and was suggested to be part of a mechanism that coordinates the energy metabolism (Van Praag et al., 2000).

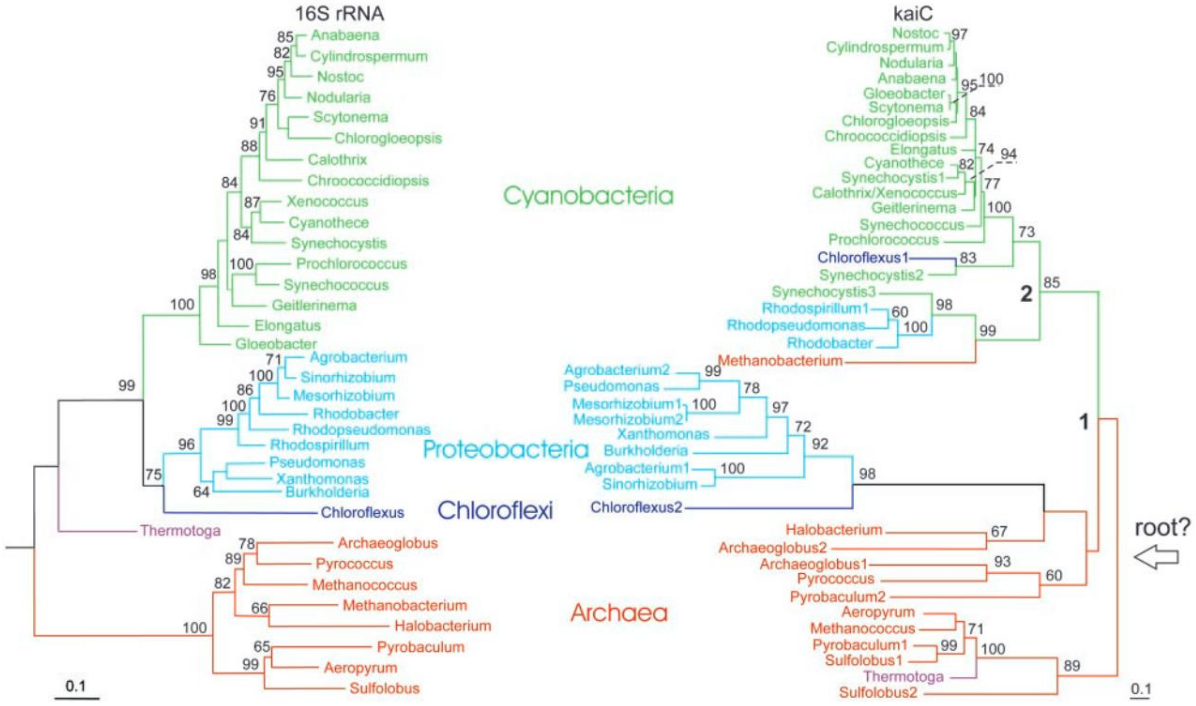


Figure 1.5: Phylogenetic trees of the *kaiC* genes compared to *16S rRNA* genes in Prokaryotes. From Dvornyk et al., 2003.

Interestingly, rhythmic activities that partly validate the criteria of a true circadian rhythms (i.e., free-running, entrainment, and temperature compensation) were also reported in bacteria that do not possess Kai homologs (Sartor et al., 2019). For instance, the soil bacterium *Bacillus subtilis* shows temperature-compensated free-running ~ 24h oscillations upon release to constant dark and temperature conditions (Eelderink-Chen et al., 2021). Similarly, the gastrointestinal bacterium *Klebsiella aerogenes* displayed an endogenously generated, temperature-compensated circadian rhythm in swarming motility (Paulose et al., 2019).

3. The diel cycle in the surface marine environment

The marine environment covers 70% of the Earth's surface and is dominated by microorganisms that represent about 60% of the ocean biomass (Bar-On et al., 2018). In the surface marine environment, light is the primary source of energy and induces diel rhythmic activity of the phytoplankton (i.e., *Cyanobacteria* and microeukaryotes), which is responsible for about 50% of the global primary production (Field et al., 1998). Besides, heterotrophic bacteria are extremely abundant and play key role in nutrient remineralization, impacting deeply the biogeochemical cycles (Azam et al., 1983). Understanding the impact of the diel cycle surface marine environment is of high biological importance because most of the biogeochemical cycles take place in this zone. While the impact of light on the physiological response of *Cyanobacteria* in laboratory condition is well documented, the broad diel dynamics of natural microbial communities still remains to be elucidated (Morimoto et al., 2020).

In this section, we will present the main actors of this ecosystem and describe the direct and indirect effects of light in the water column.

3.1. The marine picoplanktonic communities

Plankton refers to the organisms found in the water column that are incapable of independent movements. Marine planktonic organisms can be discriminated based on their size (Makoto & Tsutomu, 1984). The picoplankton is the fraction of plankton of a size ranging between 0.2–2 µm. Picoplankton is composed of picoeukaryotes and prokaryotes and accounts for the main biomass compartment in the oceans with an estimated 10^4 to 10^6 cells per milliliter (Whitman et al., 1998). It is notable that over half of these prokaryotes inhabit the 50-150 m marine surface layer (Karner et *al*., 2001). They are great contributor to primary production and carry out the bulk of life-sustaining nutrient cycling (Azam et al., 1983, Field et al., 1998).

3.1.1. Diversity of the picoplanktonic communities

Prokaryotic picoplankton form complex communities and can either be free-living, particle-attached, or alternate between both lifestyle (Grossart, 2010). The prokaryotic compartment of the marine picoplankton represents an immense pool of genetic diversity and is adapted to diverse ecological niches in the ocean (Pace, 1997, Venter et al., 2004). In fact, the marine environment is the perfect context where the adage "Everything is everywhere but the environment selects", introduced by Baas-Becking in 1934, is confirmed. Major sampling efforts across the world oceans such as the Global Ocean Sampling expedition, the Ocean Sampling Day, or the Tara expedition, have generated substantial amount of data among depth, latitudinal or longitudinal gradients (Venter et al., 2004; Bork et al., 2015; Sunagawa et al., 2015). In the past, the dominant bacterial species of the oceans were thought to be members of groups such as *Pseudomonas*, *Alteromonas* or *Vibrio*, based on culture-dependent analysis (Stal & Cretoiu, 2016). However, worldwide oceanic screening studies revealed unforeseen levels of microbial diversity in the water column. Among all microbial groups present in the oceans, the

Alphaproteobacteria, the *Gammaproteobacteria* and the *Cyanobacteria*, clearly dominate the surface waters in terms of abundance (Morris et al., 2002; Sunagawa et al., 2015) (**Figure 1.6**).

Marine microorganisms display highly diverse metabolisms depending on the sources of carbon and energy. Autotrophic bacteria obtain carbon from CO_2, heterotrophic bacteria obtain carbon from organic compounds, and mixotrophic bacteria can use both sources. Phototrophic bacteria use light as energy source and chemotrophic bacteria rely on the energy contained in organic compounds (Urry et al., 2017). In marine picoplanktonic communities, the photoautotrophs are mainly represented by *Cyanobacteria* (Kirchman, 2008). The heterotrophic compartment is phylogenetically diverse and dominated by photoheterotrophs (Koblížek, 2015). The photoheterotrophic bacteria include the proteorhodopsin-containing bacteria (e.g., "Candidatus *Pelagibacter*") and the bacteriochlorophyll-containing bacteria (e.g., aerobic anoxygenic photoheterotrophic bacteria). In the next sections, more information will be presented on the numerically most abundant bacteria inhabiting the pelagic environment.

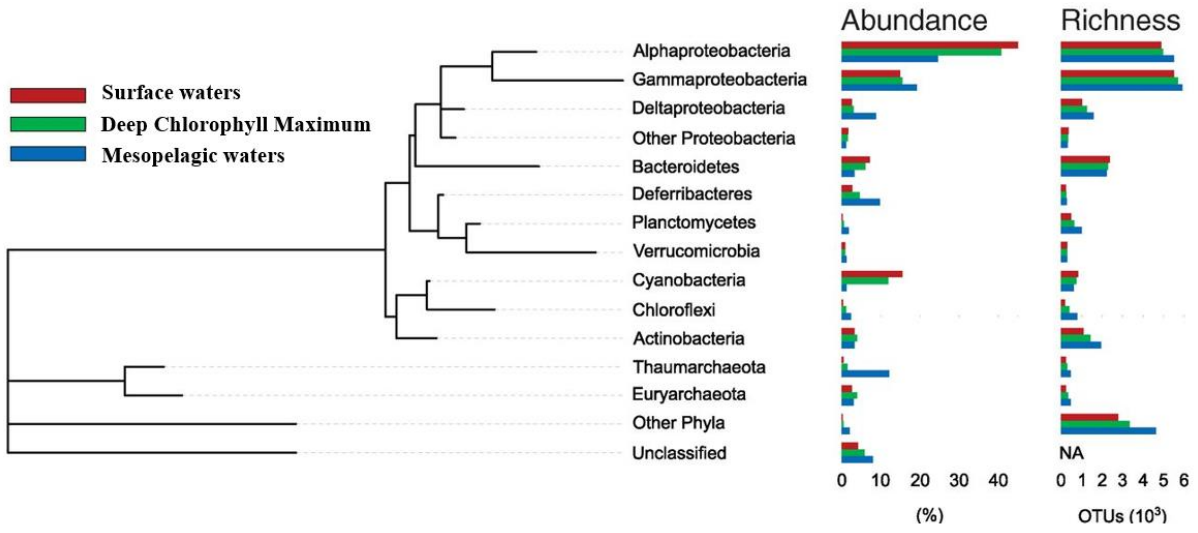


Figure 1.6: Comparison of phyla abundance and richness in the world's ocean. From Sunagawa et al., 2015.

3.1.2. *Cyanobacteria*

Cyanobacteria emerged 3.5 to 3.7 billion years ago, and they have fundamentally transformed our planet by oxidizing the atmosphere, fixing large quantities of atmospheric CO_2 into the oceans, reducing their acidity (Sánchez-Baracaldo & Cardona, 2020). *Cyanobacteria* may be unicellular, colonial, or filamentous. They are the sole bacteria able to perform oxygenic photosynthesis, which convert light energy into chemical energy by using water as electron donor and producing oxygen. They contribute for a high fraction (nearly 90 %) of the primary production, especially in oligotrophic regions (Kirchman, 2008). *Cyanobacteria* also play a major role in nitrogen fixation in the open ocean. Genetic approach targeting the gene that encodes nitrogenase enzyme, which catalyzes N_2 fixation reaction, identified the three main nitrogen-fixing cyanobacterial groups: the filamentous non-heterocyst-forming *Trichodesmium*, the filamentous heterocyst-forming symbionts with unicellular eukaryotic algae (e.g., *Richelia* and *Calothrix*) and single-celled or unicellular *Cyanobacteria* (e.g., *Crocosphaera*) (Zehr et al., 1998).

In the world oceans, the two cyanobacterial genera *Prochlorococcus* and *Synechococcus* are especially large contributors to marine primary production. *Prochlorococcus* and marine *Synechococcus* species are thought to have diverged from a common ancestor about 150 million years ago (Dufresne et al., 2005). While 16S rRNA sequences analysis would consider these two groups to be the same species, whole-genome phylogenies clearly separate *Prochlorococcus* from *Synechococcus*. This separation is justified by several physiological and ecological traits (Biller et al., 2015) (**Figures 1.7 and 1.8**). *Synechococcus* marine species are Gram-negative coccoid cells between 0.6 to 1.6 µm in size and are obligate photoautotrophs (Mitsui et al., 1987; Morel et al., 1993) (**Figure 1.7**). *Synechococcus* can be found in almost all marine environments up to the polar circle (Partensky et al., 1999).

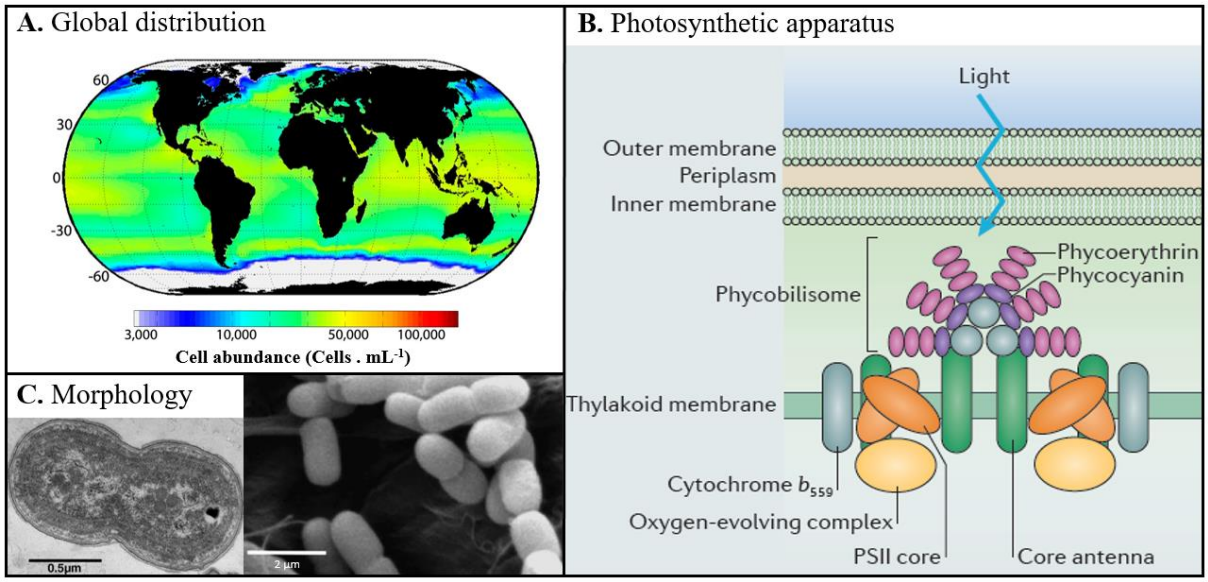


Figure 1.7: Description of *Synechococcus* sp. (a): Geographical distribution of *Synechococcus* sp. in the world ocean. (b): Photosynthetic apparatus in *Synechococcus* sp. (Redraw from Biller et al., 2015) (c): Transmission electronic microscope view (left) and scanning electronic microscope view (right) of *Synechococcus* sp. (Zhao et al., 2015).

Besides, *Prochlorococcus* forms tiny coccoid cells of about 0.5 - 0.7 µm of diameter (Morel et al., 1993) (**Figure 1.8**) and is mainly distributed in the open oceans between 40°N and 40°S, from the surface to a depth of about 200 meters and in a temperature range of 10 to 33°C, but it is absent from colder, nutrient-rich waters at high latitude as well as in most nutrient-rich coastal waters. It is one of the most abundant organisms in the oceans with an estimated total population of 2.8 to 3.0 $x10^{27}$ cells (Johnson et al., 2006; Flombaum et al., 2013). These bacteria are known as to carry the smallest genome of any free-living phototroph (Dufresne et al., 2003). Some strain such as *Prochlorococcus* MED4 display a genome of 1.65 Mbp encoding for only ~ 1 700 genes (Rocap et al., 2003). This small genome results probably from the removal of genes that provide only a small fitness benefit compared to the associated costs they represent (Sun & Blanchard, 2014).

Synechococcus performs photosynthesis via a light-harvesting antenna called the phycobilisome (**Figure 1.7**). The phycobilisome contains phycobiliproteins, such as phycocyanin and phycoerythrin that binds one or several light-harvesting chromophores

(phycocyanobilin and phycoerythrobilin). The light energy collected by the phycobilisome is transferred to the photosystem II (PSII) core antenna proteins (CP43 and CP47) and then into the PSII reaction center which comprises different proteins such as the cytochrome b559 (Ting et al., 2002) (**Figure 1.7**).

Prochlorococcus is one of the few *Cyanobacteria* that display a light-harvesting antenna complex made up of prochlorophyte chlorophyll-binding protein (Pcb) instead phycobilisomes (**Figure 1.8**). This light-harvesting antenna binds divinyl chlorophyll a and b pigments (Goericke & Repeta, 1992). In addition, *Prochlorococcus* also use monovinyl chlorophyll b as an accessory pigment (Chisholm et al., 1992). The high surface-to-volume ratio of *Prochlorococcus*, coupled with its unique pigmentation, makes it the most efficient light absorber of any photosynthetic cells (Biller et al., 2015). This explains why *Prochlorococcus* populations extend deeper in the water column than almost any other phototroph (Zwirglmaier et al., 2008).

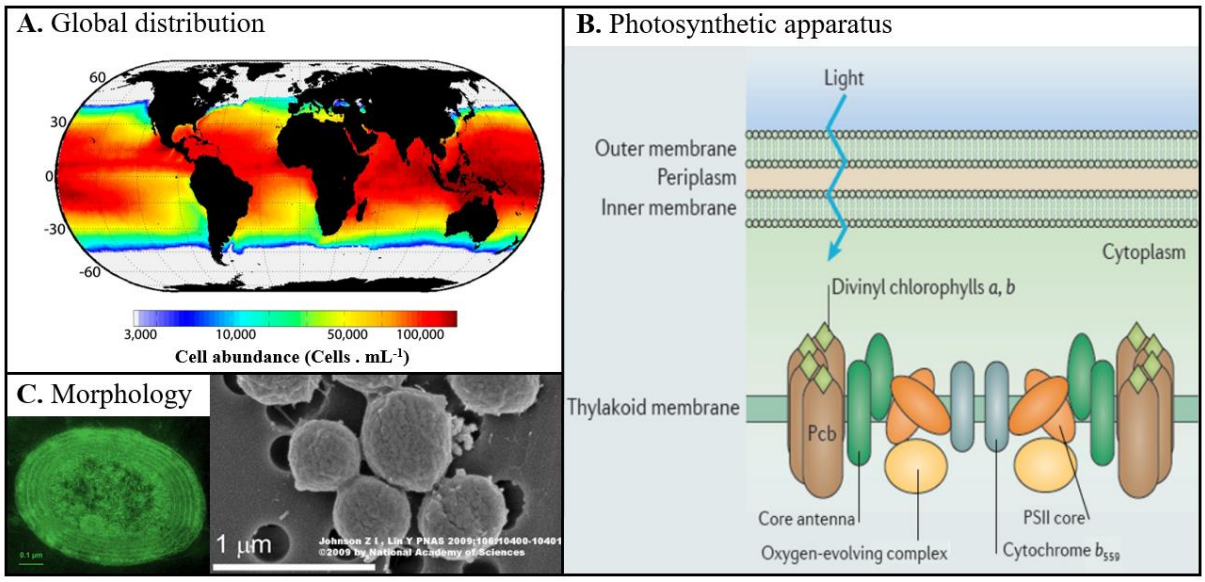


Figure 1.8: Description of *Prochlorococcus* sp. (a): Geographical distribution of *Prochlorococcus* sp. in the world ocean. (b): Photosynthetic apparatus in *Prochlorococcus* sp. (Redraw from Biller et al., 2015) (c): Transmission electronic microscope view (left) (Structural Biology Labs, Biomedical Cebtre, Uppsala, Sweden) and scanning electronic microscope view (right) of *Prochlorococcus* sp. (Johnson & Lin, 2009).

3.1.3. Candidatus *Pelagibacter*

Candidatus *Pelagibacter* are small (diameter < 0.7 µm) Gram-negative, rod-shaped bacteria classified in the class of *Alphaproteobacteria* and in the order of *Pelagibacterales* (SAR11 clade) (Rappé et al., 2002). Discovered in 1990 by metagenomic in the Sargasso Sea, these bacteria accounted for 35% of all prokaryotes in surface waters of the Sargasso Sea (Moris et al., 2002). In some oceanic regions, *Pelagibacterales* represented as much as 50 % of the total surface microbial community (Morris et al., 2002). *Pelagibacterales* could be the most abundant microorganism on earth with an estimated number of 2.4 $x10^{28}$ cells and half of it would be in the euphotic zone (Morris et al, 2002; Giovannoni et al., 2005). *Pelagibacterales* are oligotrophic and feed on dissolved organic carbon and nitrogen. In addition, they possess the proteorhodopsin, a light-driven proton pump producing ATP present in 20% of their internal membrane surface (Giovannoni et al., 2005).

3.1.4. Aerobic Anoxygenic photosynthetic bacteria

Aerobic anoxygenic photosynthetic bacteria (AAPB) are phylogenetically widespread and can be found in *Proteobacteria* (purple bacteria), *Chlorobi* (green sulfur bacteria), *Chloroflexi* (green non-sulfur bacteria), *Firmicutes*, *Acidobacteria* or *Gemmatimonadetes* (Koblížek, 2015). In the ocean, AAPB are mainly *Proteobacteria* and represent between 1 to 10% of the total pelagic microbial community (Yutin et al., 2007). They use bacteriochlorophyll-containing reaction centers to harvest light energy and complete their energy demand through the anoxygenic photosynthesis process (Koblížek, 2015). Anoxygenic photosynthesis differs from the oxygenic photosynthesis in that i) it uses hydrogen, sulfide or organic substrates instead of water as an electron donor; ii) it does not produce oxygen; iii) it involves different photosynthetic apparatus varying in the light-harvesting complex architecture, pigment composition and function of their reaction center (Béja et al., 2000;

Imhoff, 2008; Hohmann-Marriott & Blankenship 2011). Marine AAPB include the *Roseobacter*, a genus of *Rhodobacteraceae* from the *Alphaproteobacteria* class. They are one of the most abundant and versatile bacteria in the marine environment and are extremely abundant during algal bloom (Buchan et al., 2005; DeLong et al., 2005).

3.1.5. *Rhodospirillum rubrum*

Although *Rhodospirillum rubrum* is not abundant in the marine environment, this bacterium has characteristics that are interesting to investigate in the context of this thesis.

Rhodospirillum rubrum is a is a Gram-negative purple non-sulfur bacterium of a size of 0.8 to 1 µm that belongs to the *Rhodospirillaceae* (*Alphaproteobacteria*). Purple bacteria use bacteriochlorophyll and carotenoids to capture light energy and cyclic electron transport to generates a proton motive force used to produce ATP (Klamt et al., 2008). While purple sulfur bacteria use sulfide, sulfur, thiosulphate or hydrogen as electron donor, purple non-sulfur bacteria use generally hydrogen (Basak and Das, 2007). As presented previously in the section "Circadian clock beyond *Cyanobacteria*", *Rhodospirillum rubrum* displayed rhythmic activities (Van Praag et al., 2000) and possess Kai protein homologs (Dvornyk et al., 2003). Moreover, purple non-sulfur bacteria such as *Rhodospirillum rubrum*, are particularly remarkable for their extreme metabolic versatility: they can grow by photoautotrophy in anoxic conditions and by heterotrophy in oxic conditions and are able to fix nitrogen (McEwan, 1994, Madigan, 1995). For these reasons, purple non-sulfur bacteria are extremely valuable from a biotechnological perspective (De Meur, 2020). *Rhodospirillum rubrum* is particularly studied for its use in bioplastic, and biological hydrogen fuel (Bayon-Vincente et al., 2021, De Meur et al., 2020). This species was even chosen to be part of the Melissa loop, a micro-ecological system to support the metabolic needs of a crew during long-term space mission (Gòdia et al., 2002).

3.2. The role of picoplankton in biogeochemical cycles

Picoplanktonic communities play key roles in the biogeochemical cycles of the oceans (Azam et al., 1983). For instance, the pico-cyanobacteria such as *Synechococcus* and *Prochlorococcus* carry out up to half of the primary production, while heterotrophic bacteria are essential for the life-sustaining nutrient cycles (Cole et al., 1988, Field et al., 1998) (**Figure 1.9**).

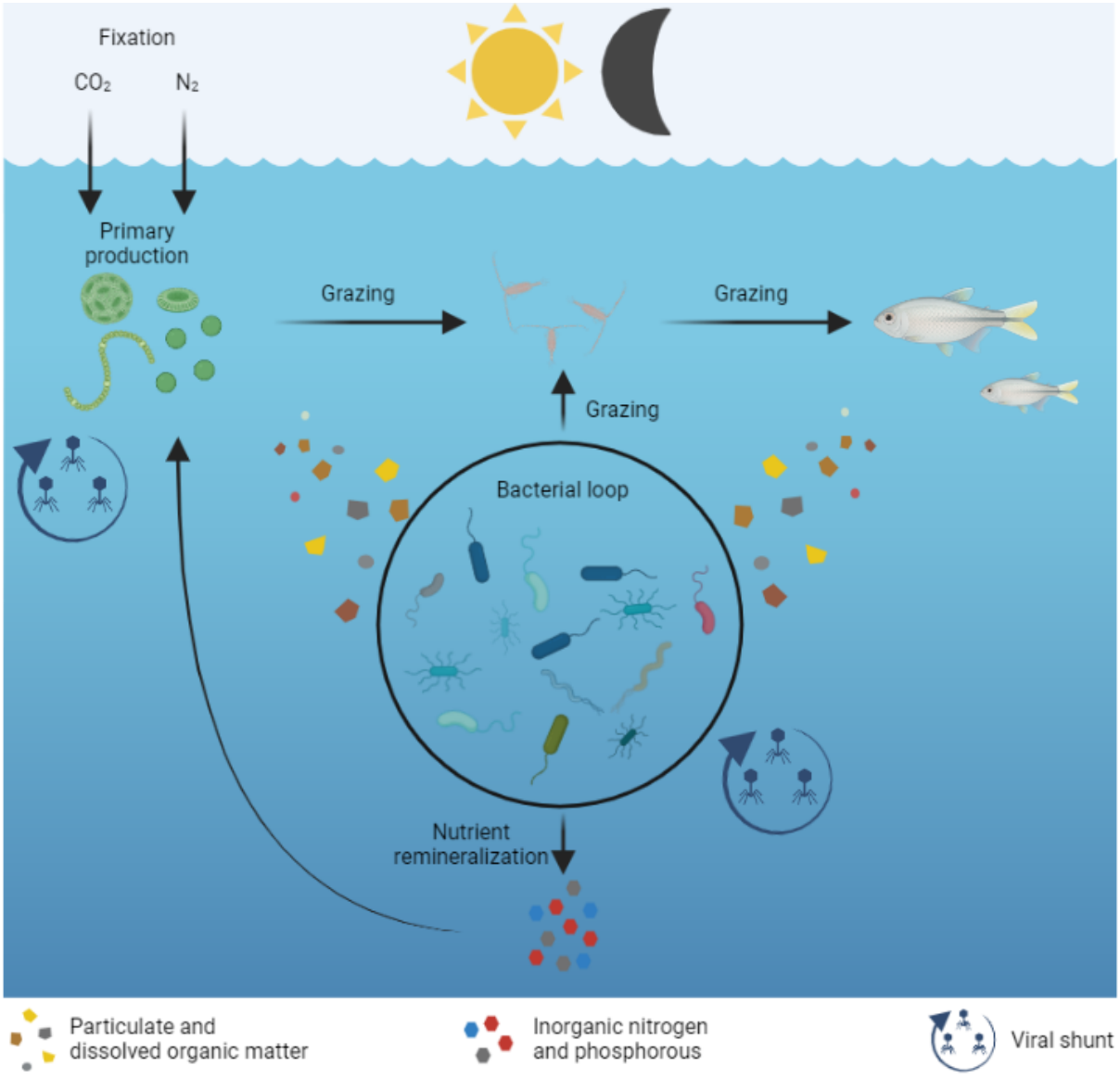


Figure 1.9: Schematization of the roles of picoplanktonic communities in the marine food web and biogeochemical cycles.

In the oceans, inorganic carbon (CO_2) is fixed by the phytoplankton (*Cyanobacteria*, algae, and plants) through the photosynthesis process that convert light energy into chemical energy (**Figure 1.9**). The carbon integrated into the prokaryotic compartment travels in the food web through predation by microeukaryotes (e.g., ciliates and heterotrophic dinoflagellates), which are further eaten by higher order consumers (e.g., fish, aquatic mammals) (Azam et al., 1983; Kirchman, 2008; Worden et al., 2015) (**Figure 1.9**). Across the food web, particulate and dissolved organic matter (POM and DOM) is released in the water column from bacterial lysis, exudation, cell death, sloppy feeding by zooplankton, or excretion. Viruses play an important role in the decay of microorganisms which contribute to the release of POM and DOM (i.e., "viral shunt") (Breitbart et al., 2018). A substantial part of POM and DOM is incorporated by the heterotrophic bacteria and reintegrated to higher trophic levels by predation (i.e., grazing). This process is called the microbial loop and is an essential part of the marine biogeochemical cycles. In addition, the microbial loop accelerates the mineralization of nutrient such as nitrogen or phosphorous and increase their bioavailability for the primary producers (Azam et al., 1983) (**Figure 1.9**). The fraction of primary production that is not respired is exported in depths in the form of particulate and dissolved organic carbon (POC, DOC) (Volk & Hoffer, 1985; Emerson and Hedges, 2008; Urry et al., 2017).

In the oceans, the atmospheric nitrogen (N_2) is the main source of nitrogen, and it is incorporated in the biota via the nitrogen fixating microorganisms (i.e., the diazotrophs) (Urry et al., 2017) (**Figure 1.9**). The nitrogen fixation process involves the nitrogenase, an enzyme that breaks the strong triple bond between nitrogen atoms to produce ammonium (Urry et al., 2017). Once fixated by diazotrophs, nitrogen is released in its bio-available forms (ammonium, nitrite, nitrate) as nitrogen organic matter, which can be used by non-diazotrophs organisms (Fowler et al., 2013).

Oceanic nitrogen fixation is mainly performed by *Cyanobacteria* but *Proteobacteria*, *Actinobacteria*, *Firmicutes* or *Euryarchaeota* also contribute to this process (Urry et al., 2017). As the nitrogenase is inhibited by oxygen, aerobic diazotrophs (i.e., *Cyanobacteria*) developed strategies to protect this enzyme. For instance, filamentous *Cyanobacteria* developed cells specialized in nitrogen fixation (i.e., heterocyst) where oxygenic photosynthesis is not performed. In unicellular and non-heterocystous diazotrophic *Cyanobacteria*, nitrogen fixation was found to be temporally segregated and occurs at night, when photosynthesis is shut down (Thiel, 2005). These diel patterns were extensively studied in the context of circadian clocks (Mitsui et *al*., 1987; Stal & Krumbein, 1987).

In contrast with carbon and nitrogen cycle, the phosphorus cycle does not practically affect the atmosphere. Principal phosphate inputs in the oceans are continental and come from wind and water erosion of phosphate rocks, soils leaching, wastewater or submarine volcanism (Urry et al., 2017). In the water column, phosphorous is the main limiting factor in the phytoplanktonic growth and is efficiently scavenged by microorganism. The orthophosphoric acid (HPO_4^{2-}) is the major source of inorganic phosphate and represents the only form of phosphate that autotrophs can assimilate (Urry et al., 2017). The organic forms of phosphate (e.g., cell debris containing nucleic acids, phospholipids, or nucleotides) are oxidized by bacteria, which produces inorganic phosphate that are directly assimilable by phytoplankton and bacteria (Azam et al., 1983) (**Figure 1.9**). To cope with low phosphate concentration in the oceans, organisms have evolved adaptation mechanisms. For example, Williams and Cavicchioli showed that bacteria overexpress nutrient transporters to efficiently scavenge nutrients in oligotrophic environment (Williams & Cavicchioli, 2014). Van Mooy and colleagues also evidenced that planktonic microbes decrease their phosphate cellular demand by switching metabolism. For instance, *Prochlorococcus* and *Synechococcus* use from S-lipids

rather than P-lipids in their membrane when they inhabit in low-P waters, which could partially explain their success in oligotrophic oceans (Van Mooy et al., 2006).

3.3. Diel cycle in marine picoplanktonic communities

3.3.1. Penetration of light in the water column

The sun emits a large spectrum of radiation including cosmic, gamma rays, x-rays, ultraviolet, visible, and infrared radiations (Thekaekara, 1976). Most of solar radiation is absorbed or reflected by the atmosphere (Iqbal, 2013) (**Figure 1.10**). The fraction of the solar radiation reaching the earth's surface is divisible in three main wavelength ranges: the ultraviolet radiation (UVR, 100-400 nm), the visible light or photosynthetically available radiation (PAR, 400-700) and the infrared radiation (IR, >700 nm). In average, the sunlight incident on the ocean surface consists of 50% of IR, 45% of visible radiation and only 5% of UVR (Tedetti & Sempéré, 2006).

IR radiations are absorbed by the seawater after a depth of one meter, causing surface waters to be warmer than deep waters. The temperature remains constant in the mixed layer because water is homogenized by winds, waves, and convection movements, then drops quickly (Tedetti & Sempéré, 2006). The red-orange-yellow (λ = 550 to 700 nm) visible range of sunlight are absorbed after few meters. In optically clear oceanic waters, the blue-green light (λ = 450 nm) penetrates the deepest, whereas in seawaters containing large amounts of organic matters, the greenish yellow (λ = 550 nm) is the most penetrating. Finally, UVB penetrate less deeply (30 to 40 meters) than UVA (60 to 70 meters) (Obernosterer et al., 2001). In average, in clear water condition, the half of the photic layer is usually affected by UVR (Tedetti & Sempéré, 2006; Wozniak & Dera, 2007).

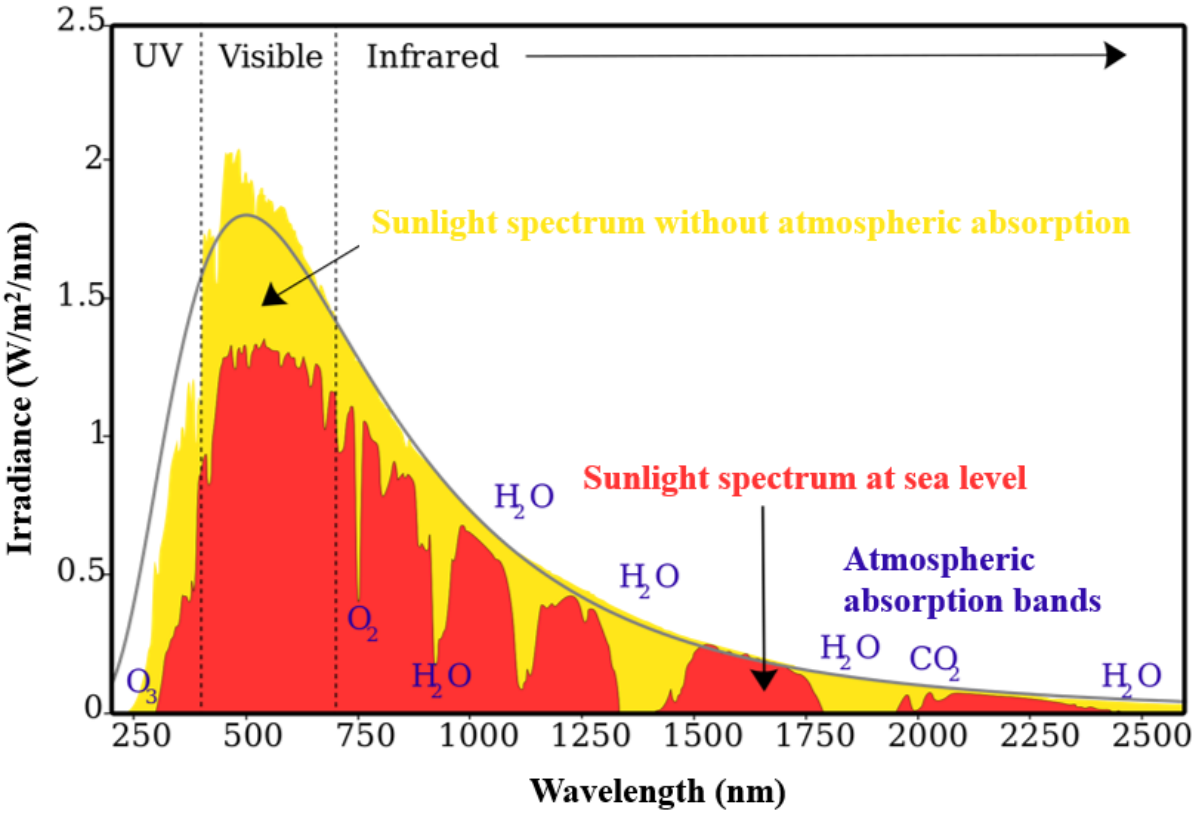


Figure 1.10: Evolution of solar radiation spectrum from space to the sea level. Modified from Beckmann & Spizzichino, 1987.

3.3.2. Effect of solar radiations in the marine environment

Besides being the primary source of energy in the oceans, solar radiations affect the marine life by inducing interactions between aquatic organisms such as competitive, mutualistic, or parasitic interaction (Sommaruga, 2005). Solar radiations also interact with organic matter and can increase the nutrient bio-lability through the process of photodissolution (Mayer et al., 2006). Depending on the quality of DOM, sunlight can photolyse recalcitrant DOM into more utilizable forms or instead, transform bio-labile DOM into recalcitrant matter (Bastidas-Navarro et al., 2009). Sunlight also causes direct DNA lesions or increase the formation of reactive oxygen species (ROS) inducing oxidative stress on DNA, proteins, or lipids. UV-induced DNA lesions and ROS toxicity can be limited via different mechanisms such as photo-enzymatic repair (PER), base and nucleotide excision repair, and post-replication repair (Matallana-Surget et al., 2009), or ROS detoxification processes (Zhao & Drlica, 2014).

3.3.3. Diel cycle of marine picoplanktonic communities

The primary producers present in marine picoplanktonic communities, such as *Cyanobacteria*, display diel cycles in response to solar radiations. Before sunrise, *Cyanobacteria* start the transcription of genes encoding for their photosynthesis apparatus to be ready to generate energy when light is available (Hellweger et al., 2020). Marine metatranscriptomic studies reported distinct day and night metabolic activities of marine microorganisms from oligotrophic marine environments (Poretsky et al., 2009; Gilbert et al., 2010; Ottesen et al., 2014). North Pacific subtropical gyre microbial communities showed that transcripts for photosynthesis, C1 metabolism and oxidative phosphorylation were overabundant at day, while housekeeping activities, such as amino acid or vitamin biosynthesis, were overrepresented at night (Poretsky et al., 2009). Similarly, transcripts involved in light-driven processes were also found to be more abundant during the day in surface marine picoplankton sampled from the Western English Channel (Gilbert et al., 2010).

Although less information is available for the heterotrophic bacteria inhabiting the photic zone, studies showed that the diel dynamics of photoautotrophs is propagated to heterotrophs and generate diel cycles. For instance, changes in transcriptional activity in response to rhythmic release of photosynthetic products was observed in heterotrophs with an upregulation of diverse metabolic pathways involved in the uptake and the transformation of organic matter (Ottesen et al., 2014; Tsai et al., 2012; Aylward et al., 2015; Frischkorn, 2018; McCarren et al., 2010; Straub et al., 2011).

The diel cycles originating from phototrophic microorganisms also propagated indirectly toward the heterotrophic compartment for those that cannot use photosynthetic products as they are. The two drivers of these indirect rhythm are the virus and the zooplankton grazing within the community. Viruses induce cell lyse in *Cyanobacteria* at night, which results

in rhythmic release of organic matter that can further propagate to heterotrophs (Aylward et al., 2017; Kimura et al., 2012; Liu, et al., 2019; Welkie et al., 2019; Yoshida et al., 2018). Grazers have shown temporal adaptations and preferentially feed on *Cyanobacteria* at night when their nitrogen content is higher (Deng et al., 2020).

4. Metaproteomics to study complex microbial assemblages

4.1. The great plate count anomaly

In 1985, Staley and Konopka introduced the term "*the great plate count anomaly*" to describe the difference between the numbers of environmental microorganisms visible by microscopy and those that were forming viable colonies on agar plate (Staley & Knopka, 1985). In fact, numerous microorganisms present in an environmental sample are in dormancy or require specific physico-chemical parameters and that are not provided in laboratory-growth conditions (Deming & Baross, 2000). In addition, microorganisms growing on agar plate can compete for nutrients and inhibit each other's growth by producing bacteriocins (Tamaki et al., 2005). Therefore, many microorganisms that were abundant and playing key roles in the environment remained ignored because of culture plate-based analysis biases (Staley & Konopka, 1985). The studies focusing on laboratory cultivable single microorganisms provided valuable information on their physiology and behavior, the specific response in artificial condition. These single-species molecular mechanisms largely contributed to the understanding of environmental microbial community functioning (i.e., bottom-up approaches).

4.2. The metaproteomic approach

The study of microbial communities was greatly facilitated by the development of the meta-omics technologies (i.e., metagenomics, metatranscriptomics, metaproteomics), which offer a "panoramic" view on the microbial molecular diversity present in a sample (Franzosa et al., 2015) (**Figure 1.11**). These advances expended our understanding of the structure and the functioning of microbial communities. According to the central dogma of molecular biology enunciated by Crick in 1958, the biological information flows from genes to transcript to proteins (Crick, 1970). In this context, the metaproteomic approach provides the best insights into the phenotypes of microorganisms, allowing pertinent characterization of the active members of a complex microbial assemblage as well as their functionality (Wilmes & Bond, 2004; Kleiner, 2019).

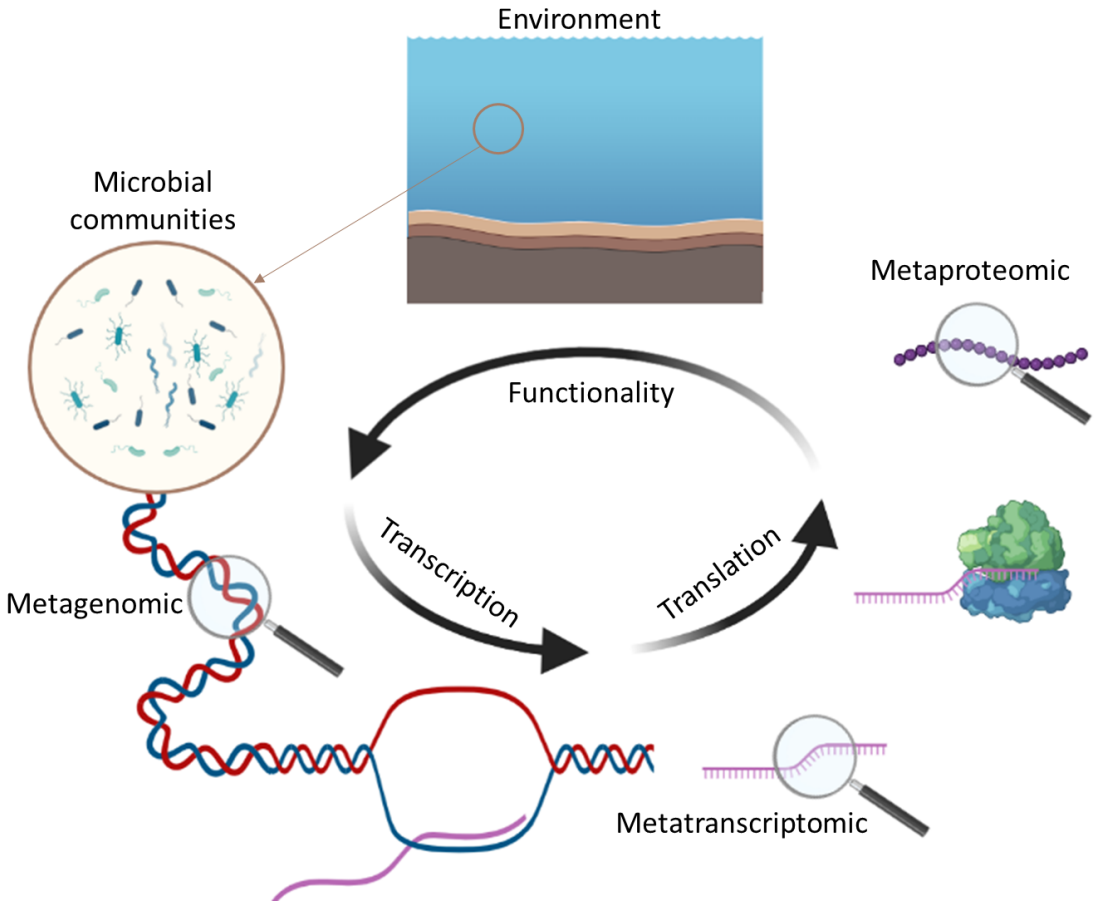


Figure 1.11: Schematic representation of the omics approaches.

Since the first environmental metaproteomic study performed in the Chesapeake Bay (Kan et al., 2005), numerous investigations were carried out in a variety of environments using descriptive, comparative, and/or quantitative approaches (Matallana-Surget et al., 2018). Comparative metaproteomics was often used to describe spatial and seasonal changes in aquatic ecosystems using (i) *in situ* (Morris et al., 2010; Teeling et al., 2012; Williams et al., 2012; Georges et al., 2014), (ii) mesocosms (Lacerda and Reardon, 2009; Bryson et al., 2016), or (iii) microcosms (Russo et al., 2016) approaches. Marine metaproteomics rapidly expanded and broadened our knowledge of marine ecosystems (Saito et al., 2019). For example, metaproteomic provided evidence for complete nitrification pathways (Morris et al., 2010; Williams et al., 2012) as well as the main synthetized kind of bacterial transporters can inform on which substrates are limiting in the environment (Sowell et al., 2011). Moreover, it revealed insights into the dynamics in organic matter transformation by microorganisms (Hawley et al., 2014; Bergauer et al., 2018) and showed the spatiotemporal variation in metabolic activities in oceanic plankton communities (Georges et al., 2014). Despite its rapid development, metaproteomics still encounters several challenges that hamper it potential power (Wilmes et al., 2015; Heyer et al., 2017; Matallana-Surget et al., 2018; Saito et al., 2019).

4.3. The workflow in metaproteomics

The metaproteomic workflow consists in four main steps: (i) sampling and protein extraction, (ii) protein separation, (iii) mass spectrometry, and (iv) protein identification/annotation (Wöhlbrand et al., 2013) (**Figure 1.12**). Metaproteomics studies on marine ecosystems are extremely heterogenous from a methodological perspective. As there is no standardized methodology, the metaproteomic workflows often differ from sampling, mass spectrometry analysis and data processing (Matallana-Surget et al., 2018; Saito et al., 2019). After sampling, proteins are extracted from the biomass using chemical (e.g., detergent) and

mechanical (e.g., sonication) disruption in order to lyse the cells and release the proteins (Matallana-Surget et al., 2018). After the extraction, the proteins are quantified via Bradford assay and/or SDS PAGE. The sample complexity can be reduced using 1D– or 2D- gels as well as chromatography steps (Matallana-Surget et al., 2018). Finally, the proteins are fragmented into peptides (e.g., trypsinolyse), and peptides are subsequently analyzed via high-throughput mass spectrometry (Matallana-Surget et al., 2018).

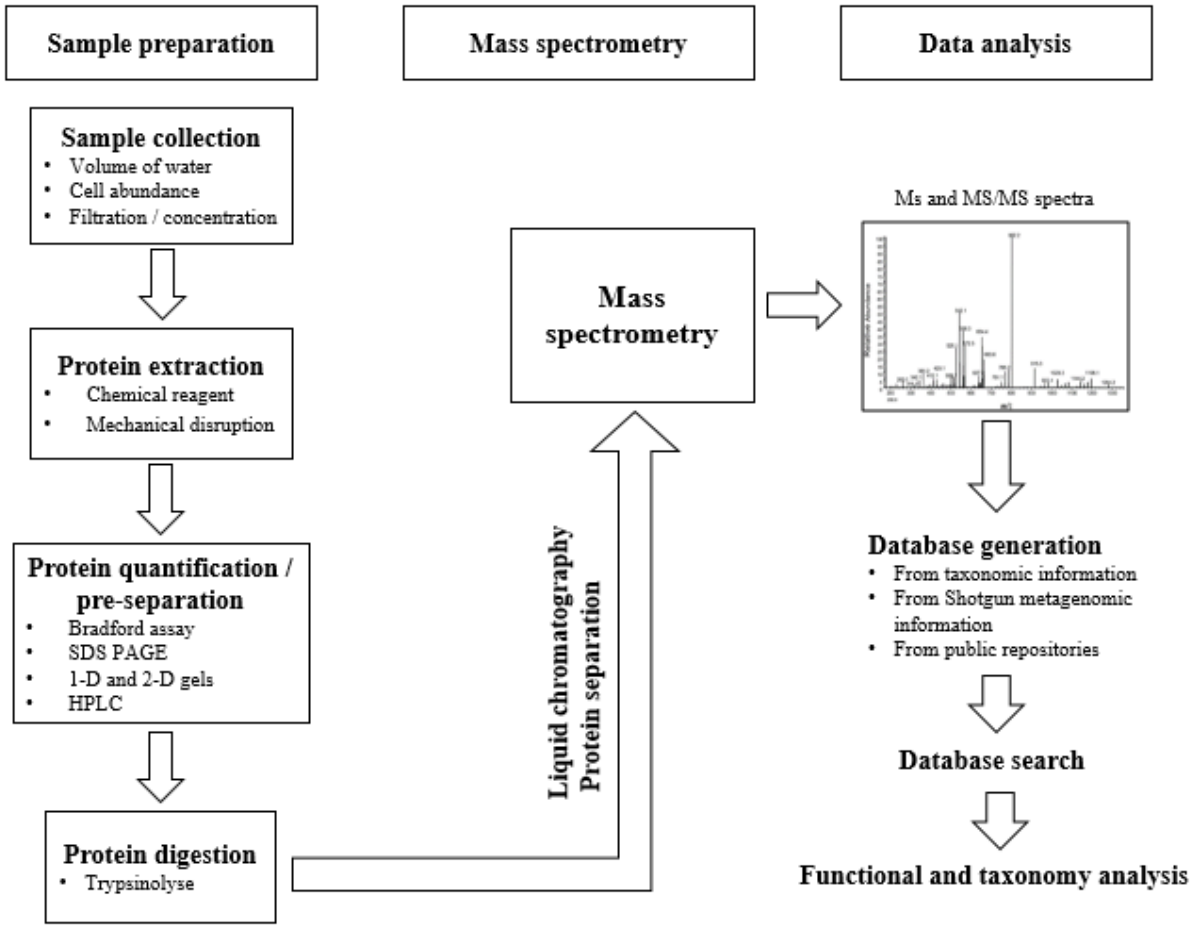


Figure 1.12: The metaproteomic workflow.

Mass spectrometer provides peptide mass spectra that correspond to the footprints of the digested peptides. Identifying these spectra and associating them to relevant peptides is challenging because of the inherent complexity of environmental microbial communities, which contain thousands of proteins from diverse taxonomic groups. The common approach is to match experimental spectra with theoretical spectra derived from a protein search database. Although the creation of protein search database can vary dramatically, the two main approaches consist in using either available public protein repository such as UniProtKB or to sequence and assemble a metagenome that correspond to the studied microbial community (Heyer et al., 2017). Identifying proteins by using large size public protein repositories such as UniProtKB can negatively affect the protein identification yield due to the presence of irrelevant sequences, which increases search space and lead to high false discovery rate (FDR) (Nesvizhskii, 2010; Jagtap et al., 2013; Tanca et al., 2013; Timmins-Schiffman et al., 2017). Building relevant protein search database that efficiently covers the metaproteome diversity consists in one of the main bioinformatical challenges in the metaproteomic workflow.

The second bioinformatical challenge encountered in metaproteomic is known as the protein inference issue, which occurs when the same peptide sequence is found in multiple proteins, thus leading to inaccurate taxonomic and functional interpretations (Herbst et al., 2016). Protein identification software such as ProteinPilot (Pro Group algorithm) (AbSciex, 2019), Prophane (Schneider et al., 2011) or MetaProteomeAnalyzer (Muth et al., 2015) addressed this issue by performing protein groups that share similar peptide-matches. However, in environmental metaproteomics, interpreting protein groups can be challenging because homologous proteins can be found in a broad taxonomic range.

5. Thesis aims

The aim of this thesis is to provide insights into the impact diel cycle on Prokaryotes and in their time-keeping mechanisms. A better understanding of the daily rhythmic activity of Archaea and Bacteria at both community and species levels will undoubtedly provide valuable translational information in the health, ecological, and industrial fields.

Deciphering the impact of diel cycle on surface marine picoplanktonic communities is essential because they are key players in this ecosystem, where they carry out about half of photosynthesis and the bulk of life-sustaining nutrient cycling (Cole et al., 1988; Whitman et al., 1998). Diel cycle induces rhythmic activity of *Cyanobacteria*, which possess a well described circadian clock (Cohen & Golden, 2015). Previous studies revealed that the diel cycle impacted picoplanktonic communities at the transcript level (Ottesen et al. 2014). However, to what extent picoplanktonic communities are rhythmically regulated at the protein level was never studied. In this purpose, the metaproteomic approach was used in this work to measure the picoplanktonic proteome fluctuations between the day and night periods during two consecutive days. Metaproteomics stands out from all the molecular tools available to decipher the real functions expressed in complex microbial assemblages (Matallana-Surget et al., 2018). Despite a rapid development, this technology still faces numerous methodological challenges (Heyer et al., 2014). Therefore, this work also aimed at providing new tools and guidance for further metaproteomic studies.

Circadian clocks were so far only characterized in *Cyanobacteria*, which represents only a small fraction of the diversity of prokaryotes on Earth (Whitman et al., 1998). However, evidence of rhythmic activity in other prokaryotes is growing. First, the *kai* genes, which encode for the core circadian mechanism in *Cyanobacteria*, are widespread among prokaryotes (Dvornyk et al., 2003). Second, numerous studies reported diel rhythmic activities in non-

cyanobacterial species (Sartor et al., 2019). Therefore, this thesis intended to provide an up-to-date taxonomic distribution of the potential clock-controlled Archaeal and Bacterial species. A clearer view on the presence of clock components in Prokaryotes will offer new opportunities for further studies in microbiology. To do so, a bioinformatical analysis was performed to identify the homologs of the Kai proteins, which encode for the core clock mechanism in *Cyanobacteria*, among the whole sequencing data available nowadays.

Studying time-keeping mechanisms and diel rhythmic activity beyond *Cyanobacteria* is crucial to better understand the evolution and the role of circadian clocks in Prokaryotes. Among the prokaryotes that could potentially be controlled by a clock, the purple non-sulfur bacteria stand out for their ecological importance and their remarkable use in biotechnology (e.g., bioplastic and biofuel production) (De Meur et al., 2020). Among them, *R. rubrum* drawn the interest of the scientific community and was chosen to be part of the Melissa loop, a micro-ecological system to support the metabolic needs of a crew during long-term space mission (Gòdia et al., 2002). Interestingly, *R. rubrum* possesses multiples copies of the *kai* homologs which makes it an excellent candidate for circadian rhythms investigations, but little is known about its diel rhythmic activity. Therefore, this work investigated for the first time the diel cycle response of this bacterium at the proteomic level and measured the regulation of its *kai* gene homologs at both transcriptional and translational levels.

6. Thesis outline

This thesis is composed of two main parts which consist in deciphering the impact of diel cycle on microbial communities and in exploring new time-keeping mechanisms in Prokaryotes. While **Chapters 2**, **3**, and **4** focus on the impact of diel cycle on marine picoplanktonic communities using the metaproteomic approach and on the development of this

technique, **Chapters 5** and **6** focus on the circadian clocks in Prokaryotes and on the case of potential new clock-controlled bacterium, respectively.

Chapter 2 presents *mPies*, a novel metaproteomics tool designed during this thesis to strengthen the metaproteomic toolkit. *mPies* overcomes two of the main challenges encountered in metaproteomic: i) the creation of relevant protein search database and ii) the automatization of reliable taxonomic and functional consensus annotations. **Chapter 3** shed light on the critical steps in the metaproteomic workflow and highlight the impact of the choice of protein search database in both gel-free and gel-based metaproteomics. Based on the optimization of the metaproteomic workflow, **Chapter 4** deciphers the diel variation in the protein expression of marine picoplanktonic communities by comparing two consecutive day and night metaproteomes of both the free-living and the particle-attached bacterial fractions sampled from the northwest Mediterranean Sea surface.

Chapter 5 provides a bioinformatical analysis of the Prokaryotes harboring homologs of the Kai proteins and elaborates on the new potential clock-controlled microorganisms of high ecological and industrial relevance. Based on the data obtained from the diel metaproteomic study and the inventory of Kai protein homologs in Prokaryotes, we decided to further investigate the diel cycle and the expression of circadian clock genes in the purple non-sulfur bacterium *Rhodospirillum rubrum*. **Chapter 6** describes an overview on the diel cycle impact on the whole proteome expression and the transcriptional and translational regulation of *kai* genes in *R. rubrum* by using shotgun and targeted proteomics coupled with RT-qPCR.

Chapter 2: *mPies*

mPies: a novel metaproteomics tool for the creation of relevant protein databases and automatized protein annotation

Werner, J., **Géron**, **A**., Kerssemakers, J., & Matallana-Surget, S.
Biology Direct, 14, (2019), 21

Contributions

JW, SMS, and AG designed mPies. JW developed and implemented mPies, JK contributed valuable discussions with regards to the software design of mPies and performed code review. JW, AG, and SMS wrote substantial parts of the manuscript, all authors performed proofreading and approved the final version of the manuscript.

Abstract

Metaproteomics allows to decipher the structure and functionality of microbial communities. Despite its rapid development, crucial steps such as the creation of standardized protein search databases and reliable protein annotation remain challenging. To overcome those critical steps, we developed a new program named *mPies* (metaProteomics in environmental sciences). *mPies* allows the creation of protein databases derived from assembled or unassembled metagenomes, and/or public repositories based on taxon IDs, gene or protein names. For the first time, *mPies* facilitates the automatization of reliable taxonomic and functional consensus annotations at the protein group level, minimizing the well-known protein inference issue, which is commonly encountered in metaproteomics. *mPies'* workflow is highly customizable with regards to input data, workflow steps, and parameter adjustment. *mPies* is implemented in Python 3/Snakemake and freely available on GitHub: https://github.com/johanneswerner/mPies/.

1. Introduction

Metaproteomics is a valuable method to link the taxonomic diversity and functions of microbial communities (Wilmes & Bond, 2004). However, the use of metaproteomics still faces methodological challenges and lacks standardisation (Matallana-Surget et al., 2018). The creation of relevant protein search databases and protein annotation remain hampered by the inherent complexity of microbial communities (Heyer et al., 2017). Protein search databases can be created based on reads or contigs derived from metagenomic and/or metatranscriptomic data (Tanca et al., 2016; Timmins-Schiffman et al., 2017). Public repositories such as Ensembl (Zerbino et al., 2018), NCBI (NCBI Resource Coordinators, 2018) or UniProtKB (The UniProt Consortium, 2016) can also be used as search databases but it is necessary to apply relevant filters (e.g., based on the habitat or the taxonomic composition) in order to decrease computation time and false discovery rate (Tanca et al., 2016). Until now, no tool exists that either creates taxonomic or functional subsets of public repositories or combines different protein databases in order to optimize the total number of identified proteins.

The so-called protein inference issue occurs when the same peptide sequence is found in multiple proteins, thus leading to inaccurate taxonomic and functional interpretation (Herbst et al., 2016). To address this issue, protein identification software tools such as ProteinPilot (Pro Group algorithm) (AbSciex, 2019), Prophane (Schneider et al., 2011) or MetaProteomeAnalyzer (Muth et al., 2015) perform automatic grouping of homologous protein sequences. Interpreting protein groups can be challenging especially in complex microbial community where redundant proteins can be found in a broad taxonomic range. A well-known strategy to deal with homologous protein sequences is to calculate the lowest common ancestor (LCA). For instance, MEGAN performs taxonomic binning by assigning sequences on the nodes of the NCBI taxonomy and calculates the LCA on the best alignment hit (Huson et al.,

2016). However, another crucial challenge related to protein annotation still remains: protein sequences annotation often relies on alignment programs automatically retrieving the first hit only (Pible & Armengaud, 2015). The reliability of this approach is hampered by the existence of taxonomic and functional discrepancies among the top alignment results with very low e-values (Timmins-Schiffman et al., 2017). Here, we present *mPies*, a new highly customizable program that allows the creation of protein search databases and performs post-search protein consensus annotation, thus facilitating biological interpretation.

2. Workflow design

mPies provides multiple options for optimizing metaproteomic analysis within a standardized and automatized workflow (**Figure 2.1**). *mPies* is written in Python 3.6, uses the workflow management system Snakemake (Köster & Rahmann, 2012) and relies on Bioconda (Grüning et al., 2018) to ensure reproducibility (**Appendix A**). *mPies* can run in up to four different modes to create databases (DBs) for protein search using amplicon/metagenomic and/or public repositories data: (i) non-assembled metagenome-derived DB, (ii) assembled metagenome-derived DB, (iii) taxonomy-derived DB, and (iv) functional-derived DB. After protein identification, *mPies* can automatically compute sequence alignment-based consensus annotation at protein group level. By considering multiple alignment hits for reliable taxonomic and functional inference, *mPies* limits the protein inference issue and allows more relevant biological interpretation of metaproteomes from diverse environments.

2.1. Mode (i): Non-assembled metagenome-derived DB

In mode (i), *mPies* trims metagenomic raw reads (fastq files) with Trimmomatic (Bolger et al., 2014), and predicts partial genes with FragGeneScan (Rho et al., 2010) which are built into the protein DB.

2.2. Mode (ii): Assembled metagenome-derived DB

In mode (ii), trimmed metagenomic reads are assembled either with MEGAHIT (Li et al., 2015) or metaSPAdes (Nurk et al., 2017). The genes are subsequently called with Prodigal (Hyatt et al., 2010). The utilization of Snakemake allows easy adjustment of the assembly and gene calling parameters.

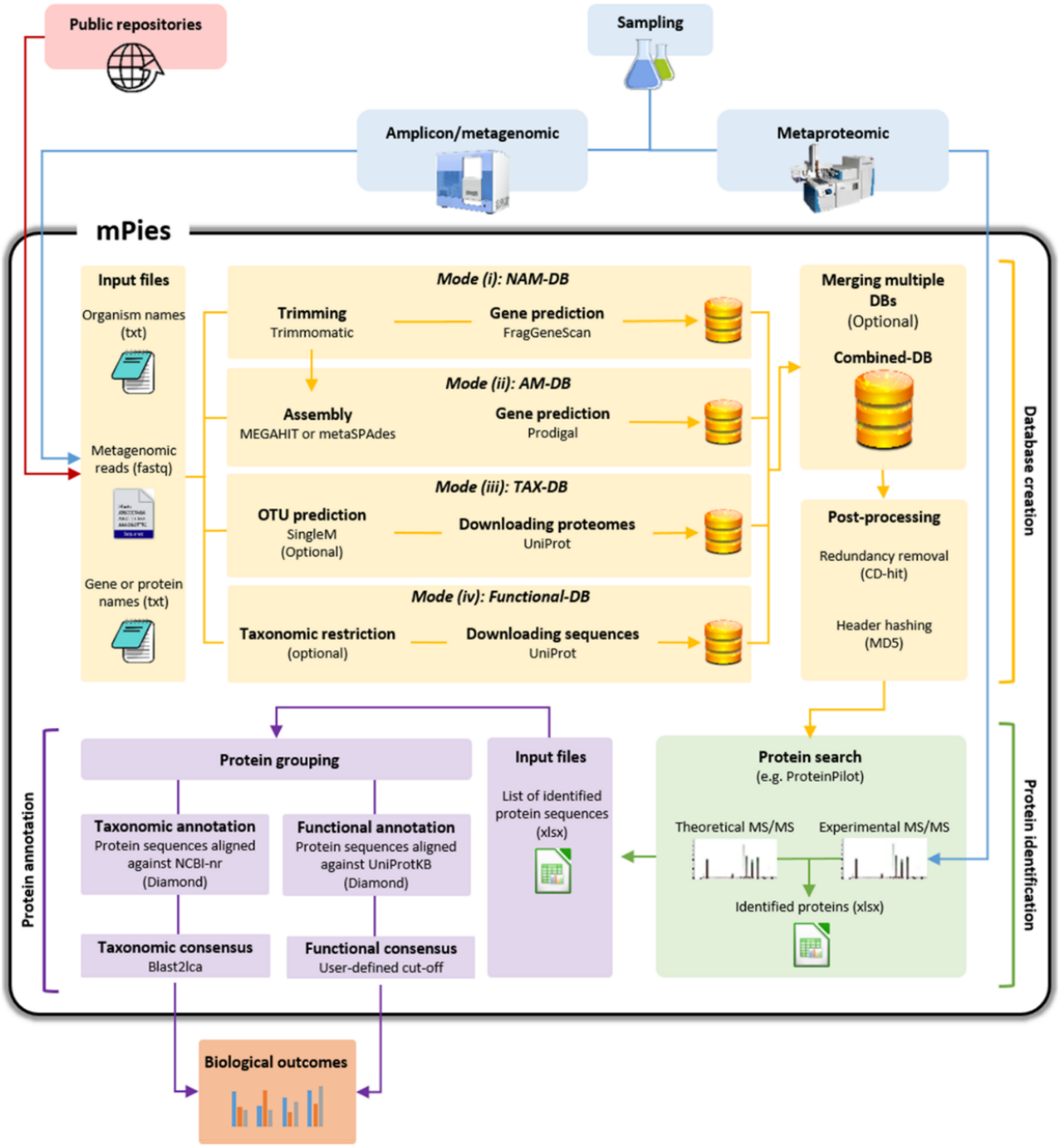


Figure 2.1: *mPies* workflow.

2.3. Mode (iii): Taxonomy-derived DB

In mode (iii), *mPies* extracts the taxonomic information derived from the metagenomic raw data and downloads the corresponding proteomes from UniProt. To do so, *mPies* uses SingleM (Woodcroft, 2018) to predict OTUs from the metagenomic reads. Subsequently, a non-redundant list of taxon IDs corresponding to the taxonomic diversity of the observed habitat is generated. Finally, *mPies* retrieves all available proteomes for each taxon ID from UniProt. It is noteworthy that the taxonomy-derived DB can be generated from 16S amplicon data or a user-defined list.

2.4. Mode (iv): Functional-derived DB

Mode (iv) is a variation of mode (iii) which allows to create DBs that target specific functional processes (e.g., carbon fixation or sulphur cycle) instead of downloading entire proteomes for taxonomic ranks. For that purpose, *mPies* requires a list of gene or protein names as input and downloads all the corresponding protein sequences from UniProt. Taxonomic restriction can be defined (e.g., *Proteobacteria*-related sequences only) for highly specific DB creation.

2.5. Post-processing

If more than one mode was selected for protein DB generation, all proteins are merged into one combined protein search DB. Duplicated protein sequences (default: sequence similarity 100%) are removed with CD-HIT (Fu et al., 2012). All protein headers are hashed (default: MD5) to obtain uniform headers and to reduce the file size for the final protein search database in order to keep the memory requirements of downstream analysis low.

2.6. Protein annotation

mPies facilitates taxonomic and functional consensus annotation at protein level. After protein identification, each protein is aligned with Diamond (Buchfink et al., 2015) against NCBI-nr (NCBI Resource Coordinators, 2018) for the taxonomic annotation. For the functional prediction, proteins are aligned against UniProt (Swiss-Prot or TrEMBL) (The UniProt Consortium, 2016) and COG (Galperin et al., 2015). The alignment hits (default: retained aligned sequences = 20, bitscore ≥80) are automatically retrieved for consensus taxonomic and functional annotation, for which the detailed strategies are provided below.

The taxonomic consensus annotation uses the alignment hits against NCBI-nr and applies the LCA algorithm to retrieve a taxonomic annotation for each protein group (protein grouping comprises the assignment of multiple peptides to the same protein and is facilitated by proteomics software) as described by Huson et al. (Huson et al., 2016). For the functional consensus, the alignment hits against UniProt and/or COG are used to extract the most frequent functional annotation per protein group within their systematic recommended names. This is the first time that a metaproteomics tool includes this critical step, as previously only the first alignment hit was kept. In order to ensure the most accurate annotation, a minimum of 20 best alignment hits should be kept for consensus annotation. Nevertheless, this parameter is customizable and and this number could be modified.

3. Conclusion

The field of metaproteomics has rapidly expanded in recent years and has led to valuable insights in the understanding of microbial community structure and functioning. In order to cope with metaproteomic limitations, new tools development and workflow standardization are of urgent needs. With regards to the diversity of the technical approaches found in the literature

which are responsible for methodological inconsistencies and interpretation biases across metaproteomic studies, we developed the open-source program *mPies*. It proposes a standardized and reproducible workflow that allows customized protein search DB creation and reliable taxonomic and functional protein annotations. *mPies* facilitates biological interpretation of metaproteomics data and allows unravelling microbial community complexity.

Chapter 3: Optimization in metaproteomics

Deciphering the Functioning of Microbial Communities: Shedding Light on the Critical Steps in Metaproteomics

Géron, **A**., Werner, J., Wattiez, R., Lebaron, P., & Matallana-Surget, S.
Frontiers in microbiology, 10, (2019), 2395

Abstract

Unravelling the complex structure and functioning of microbial communities is essential to accurately predict the impact of perturbations and/or environmental changes. From all molecular tools available today to resolve the dynamics of microbial communities, metaproteomics stands out, allowing the establishment of phenotype–genotype linkages. Despite its rapid development, this technology has faced many technical challenges that still hamper its potential power. How to maximize the number of protein identification, improve quality of protein annotation, and provide reliable ecological interpretation are questions of immediate urgency. In our study, we used a robust metaproteomic workflow combining two protein fractionation approaches (gel-based versus gel-free) and four protein search databases derived from the same metagenome to analyze the same seawater sample. The resulting eight metaproteomes provided different outcomes in terms of (i) total protein numbers, (ii) taxonomic structures, and (iii) protein functions. The characterization and/or representativeness of numerous proteins from ecologically relevant taxa such as *Pelagibacterales*, *Rhodobacterales*, and *Synechococcales*, as well as crucial environmental processes, such as nutrient uptake, nitrogen assimilation, light harvesting, and oxidative stress response, were found to be particularly affected by the methodology. Our results provide clear evidence that the use of different protein search databases significantly alters the biological conclusions in both gel-free and gel-based approaches. Our findings emphasize the importance of diversifying the experimental workflow for a comprehensive metaproteomic study.

1. Introduction

Metaproteomics aims at characterizing the total proteins obtained from microbial communities (Wilmes & Bond, 2004) and, in association with metagenomics, unraveling the functional complexity of a given ecosystem (Franzosa et al., 2015). Since the first environmental metaproteomic study performed in the Chesapeake Bay (Kan et al., 2005), numerous investigations were carried out in a variety of environments using descriptive, comparative, and/or quantitative approaches (Matallana-Surget et al., 2018). Comparative metaproteomics was often used to describe spatial and seasonal changes in aquatic ecosystems using (i) *in situ* (Morris et al., 2010; Teeling et al., 2012; Williams et al., 2012; Georges et al., 2014), (ii) mesocosms (Lacerda & Reardon, 2009; Bryson et al., 2016), or (iii) microcosms (Russo et al., 2016) approaches.

Metaproteomics on marine ecosystems is a rapidly expanding field that involves a series of challenging steps and critical decisions in its workflow (Wilmes et al., 2015; Heyer et al., 2017; Matallana-Surget et al., 2018; Saito et al., 2019). The marine metaproteomic workflow consists mainly of four steps: (i) sampling and protein extraction, (ii) protein separation, (iii) mass spectrometry, and (iv) protein identification/annotation (Wöhlbrand et al., 2013). Until now, standardized experimental protocols are still missing, leading to methodological inconsistencies and data interpretation biases across metaproteomic studies (Leary et al., 2013; Tanca et al., 2013; Timmins-Schiffman et al., 2017).

Protein identification strongly relies on both the quality of experimental mass spectra (MS) and the comprehensiveness of the protein search database (DB) (Wöhlbrand et al., 2013). Both gel-based (Wang et al., 2014) and shotgun gel-free (Morris et al., 2010; Saito et al., 2015) approaches have been used in metaproteomic analyses and both were found to be complementary (Matallana-Surget et al., 2018). Two main data sources are commonly used to

construct protein search DB: public protein repositories, and/or metagenomic data (Heyer et al., 2017). Identifying proteins by searching against public protein repositories such as UniProtKB/SwissProt, UniProtKB/TrEMBL, UniRef, NCBI, or Ensembl is challenging because of the large size of these DBs, which increase search space and overestimate false discovery rate (FDR), thus decreasing the total number of identified proteins (Nesvizhskii, 2010; Jagtap et al., 2013; Tanca et al., 2013; Timmins-Schiffman et al., 2017). To address the issue of large size DB, different strategies were developed such as pseudo-metagenome approach (Heyer et al., 2017), partial searches against smaller sub-DB (Muth et al., 2015a; Tanca et al., 2016), or the two-round DB searching method (Jagtap et al., 2013). The two-round DB searching method consists in searching experimental MS against a refined database composed of the protein sequences identified in a preliminary error tolerant search, allowing significant increase in the total number of identified proteins. This strategy was extensively used in recent metaproteomics studies (Russo et al., 2016; Serrano-Villar et al., 2016; Deusch et al., 2017; Gallois et al., 2018). Regarding metagenomic data, both assembled (Teeling et al., 2012) and non-assembled (Herbst et al., 2016; Tanca et al., 2016) sequencing reads were used in metaproteomics for protein search DB creation. Skipping read assembly was shown to prevent information loss and potential noise introduction and led to higher protein identification yield (May et al., 2016).

Metaproteomic data analysis also involves taxonomic and functional annotation. Due to the protein inference issue (i.e., a same peptide can be found in homologous proteins), inaccurate protein annotations are commonly encountered in metaproteomics (Herbst et al., 2016). To overcome this issue, protein identification tools such as Pro Group algorithm (Absciex, 2014), Prophane (Schneider et al., 2011), or MetaProteomeAnalyzer (Muth et al., 2015b) automatically group homologous protein sequences. In our study, we used the *mPies* tool (Werner et al., 2019), which uses sequence-based alignment to compute taxonomic

consensus annotation on protein groups using last common ancestor (LCA) (Huson et al., 2016; Heyer et al., 2017). *mPies* also provides a novel consensus functional annotation using UniProt, that gives more accurate insights into the diversity of protein functions compared to former strategies mapping proteins on broader functional categories, such as KEGG (Kanehisa et al., 2018) or COGs (Galperin et al., 2015).

To what extend the methodology affects the metaproteome interpretation has already been studied in artificial microbial communities (Tanca et al., 2013) and gut microbiomes (Tanca et al., 2016; Rechenberger et al., 2019) but its impact on marine samples still remains poorly documented (Timmins-Schiffman et al., 2017). In this study, we used a robust experimental design comparing the combined effect of protein search DB choice and protein fractionation approach on the same sea surface sample. For this purpose, two sets of peptide spectra resulting from gel-based and gel-free approaches were searched against four DBs derived from the same raw metagenomic data. The resulting eight metaproteomes were quantitatively and qualitatively compared, demonstrating to which extent diversifying metaproteomic workflow allows the most comprehensive understanding of microbial communities' dynamics.

2. Materials and methods

2.1. Sampling

Seawater samples (n = 4) were collected in summer (June 2014) at the SOLA station, located 500 m offshore of Banyuls-sur-Mer, in the Northwestern Mediterranean Sea (42° 49′N, 3° 15′W). Each sample consisted of 60 L of sea surface water, pre-filtered at 5 μm and subsequently sequentially filtered through 0.8 and 0.2 μm pore-sized filters (polyethersulfone

membrane filters, PES, 142 mm, Millipore). Four independent sets of filters were obtained and flash frozen into liquid nitrogen before storage at −80°C.

2.2. Protein isolation for gel-based and gel-free approaches

A combination of different mechanical (sonication/freeze–thaw) and chemical (urea/thiourea containing buffers, acetone precipitation) extraction techniques were used on the filtered seawater samples to maximize the recovery of protein extracts from the filters. The 0.2 μm filters were removed from their storage buffer and cut into quarters using aseptic procedures. Protein isolation was performed on four 0.2 μm filters. The same protein isolation protocol was used for both gel-based and gel-free approaches. The filters were suspended in a lysis buffer containing 8 M urea/2 M thiourea, 10 mM HEPES, and 10 mM dithioerythritol (DTE). Filters were subjected to five freeze–thaw cycles in liquid N_2 to release cells from the membrane. Cells were mechanically broken by sonication on ice (five cycles of 1 min with tubes on ice, amplitude 40%, 0.5 pulse rate) and subsequently centrifuged at 16,000 × g at 4°C for 15 min. To remove particles that did not pellet during the centrifugation step, we filtered the protein suspension through a 0.22 μm syringe filter and transferred into a 3 kDa cutoff Amicon Ultra-15 filter unit (Millipore) for protein concentration. Proteins were precipitated with cold acetone overnight at −80°C, with an acetone/aqueous protein solution ratio of 4:1. Total protein concentration was determined by a Bradford assay, according to the Bio-Rad Protein Assay kit (Bio-Rad, Hertfordshire, United Kingdom) according to the manufacturer's instructions, with bovine γ-globulin as a protein standard. Protein samples were reduced with 25 mM DTE at 56°C for 30 min and alkylated with 50 mM iodoacetamide at room temperature for 30 min. For gel-free liquid chromatography tandem mass spectrometry analysis, a tryptic digestion (sequencing grade modified trypsin, Promega) was performed overnight at 37°C, with an enzyme/substrate ratio of 1:25.

2.3. Gel-based proteomics approach

Protein isolates diluted in Laemmli buffer (2% SDS, 10% glycerol, 5% β-mercaptoethanol, 0.002% bromophenol blue, and 0.125 M Tris–HCl, pH 6.8) and sonicated in a water bath six times for 1 min at room temperature. After 1 min incubation at 90°C, the protein solutions were centrifuged at 13,000 rpm at room temperature for 15 min. The SDS-PAGE of the protein mixtures was conducted using 4–20% precast polyacrylamide mini-gels (Pierce). The protein bands were visualized with staining using the Imperial Protein Stain (Thermo) according to the manufacturer's instructions. The corresponding gel lane containing proteins was cut in 17 pieces of 1 mm each. In-gel proteins were reduced with 25 mM DTE at 56°C for 30 min and alkylated with 50 mM iodoacetamide at room temperature for 30 min. In-gel enzymatic digestion was performed by the addition of 10 µl modified sequencing grade trypsin (0.02 mg/ml) in 25 mM NH_4HCO_3. The samples were placed for 15 min at 4°C and incubated overnight at 37°C. The reaction was stopped with 1 µl 5% (v/v) formic acid. Tryptic peptides were collected by centrifugation analyzed by liquid chromatography tandem mass spectrometry.

2.4. Liquid chromatography tandem mass spectrometry analysis

Purified peptides from digested protein samples from gel-free and gel-based proteomics were identified using a label-free strategy on an UHPLC-HRMS platform composed of an Eksigent 2D liquid chromatograph and an AB SCIEX Triple TOF 5600. Peptides were separated on a 25 cm C18 column (Acclaim pepmap 100, 3 µm, Dionex) by a linear acetonitrile (ACN) gradient [5–35% (v/v), in 15 or 120 min] in water containing 0.1% (v/v) formic acid at a flow rate of 300 nL min–1. MS were acquired across 400–1,500 m/z in high-resolution mode (resolution >35,000) with 500 ms accumulation time. Six microliters of each fraction were loaded onto a pre-column (C18 Trap, 300 µm i.d. × 5 mm, Dionex) using the Ultimate 3000

system delivering a flow rate of 20 µl/min loading solvent [5% (v/v) ACN, 0.025% (v/v) TFA]. After a 10 min desalting step, the pre-column was switched online with the analytical column (75 µm i.d. × 15 cm PepMap C18, Dionex) equilibrated in 96% solvent A [0.1% (v/v) formic acid in HPLC-grade water] and 4% solvent B [80% (v/v) ACN, 0.1% (v/v) formic acid in HPLC-grade water]. Peptides were eluted from the pre-column to the analytical column and then to the mass spectrometer with a gradient from 4 to 57% solvent B for 50 min and 57 to 90% solvent B for 10 min at a flow rate of 0.2 µL min–1 delivered by the Ultimate pump. Positive ions were generated by electrospray and the instrument was operated in a data-dependent acquisition mode described as follows: MS scan range: 300–1,500 m/z, maximum accumulation time: 200 ms, ICC target: 200,000. The top four most intense ions in the MS scan were selected for MS/MS in dynamic exclusion mode: ultrascan, absolute threshold: 75,000, relative threshold: 1%, excluded after spectrum count: 1, exclusion duration: 0.3 min, averaged spectra: 5, and ICC target: 200,000. Gel-based and gel-free metaproteomic data were submitted to iProx (Ma et al., 2018) (Project ID: IPX0001684000/PXD014582).

2.5. Databases creation and protein identification

Protein searches were performed with ProteinPilot (ProteinPilot Software 5.0.1; Revision: 4895; Paragon Algorithm: 5.0.1.0.4874; AB SCIEX, Framingham, MA, United States) (Matrix Science, London, United Kingdom; v. 2.2). Paragon searches 34 were conducted using LC MS/MS Triple TOF 5600 System instrument settings. Other parameters used for the search were as follows: Sample Type: Identification, Cys alkylation: Iodoacetamide, Digestion: Trypsin, ID Focus: Biological Modifications and Amino acid substitutions, Search effort: Thorough ID, Detected Protein Threshold [Unused ProtScore (Conf)]>: 0.05 (10.0%).

Three DBs were created using the same metagenome (EMBL-EBI Project number: ERP009703, Ocean Sampling Day 2014, sample: OSD14_2014_06_2m_NPL022, run ID: ERR771073) (MiSeq Illumina Technology) and were generated with *mPies* v 0.9, our recently in house developed *mPies* program freely available at https://github.com/johanneswerner/mPies/ (Werner et al., 2019). The three DBs were: (i) a non-assembled metagenome-derived DB (NAM-DB), (ii) an assembled metagenome-derived DB (AM-DB), and (iii) a taxonomy-derived DB (TAX-DB) (**Table 3.1**). Briefly, *mPies* first trimmed sequencing raw reads with Trimmomatic (Bolger et al., 2014). For NAM-DB, *mPies* directly predicted genes from trimmed sequencing reads with FragGeneScan (Rho et al., 2010). For AM-DB, *mPies* first assembled trimmed sequencing reads into contigs using metaSPAdes (Nurk et al., 2017) and subsequently called genes with Prodigal (Hyatt et al., 2010). For TAX-DB, *mPies* created a pseudo-metagenome using SingleM (Woodcroft, 2018) to predict operational taxonomic units from the trimmed sequencing reads and retrieved all the taxon IDs at genus level. All available proteomes for each taxon ID were subsequently downloaded from UniProtKB/TrEMBL. Duplicated protein sequences were removed with CD-HIT (Fu et al., 2012) from each DB.

Gel-based and gel-free MS/MS spectra were individually searched twice against the DBs. In the first-round search, full size NAM-DB, AM-DB, and TAX-DB were used (**Table 3.1**). In the second-round search, each DB was restricted to the protein sequences identified in the first-round search. For both gel-free and gel-based approaches, the second round NAM-DB, AM-DB, and TAX-DB were merged, and redundant protein sequences were removed, leading to two combined DBs (Comb-DBs), subsequently searched against gel-based and gel-free MS/MS spectra. Consequently, a total of eight metaproteomes obtained from four DBs: NAM-DB, AM-DB, TAX-DB, and Comb-DB were analyzed in this paper. An FDR threshold of 1%, calculated at the protein level, was used for each protein searches. Proteins identified with one

single peptide were validated by manual inspection of the MS/MS spectra, ensuring that a series of at least five consecutive sequence-specific b-and y-type ions was observed.

Table 3.1: Two-round search performances depending on the metaproteomic workflow.

	Database	Number of proteins in database	Number of peptide spectra identified[a]	Coverage of peptide spectra identified (%)	Number of distinct peptides Identified[a]	Number of proteins after validation[a,b]
First-round searches						
Gel-free	AM-DB	64,613	24,684	7.8	3,237	347
	NAM-DB	462,821	35,430	11.1	4,295	834
	TAX-DB	13,426,277	10,626	3.3	1,624	607
Gel-based	AM-DB	64,613	2,066	24.7	1,408	201
	NAM-DB	462,821	2,584	30.9	1,849	652
	TAX-DB	13,426,277	2,304	27.5	1,526	496
Second-round searches						
Gel-free	AM-DB	782	42,831	13.4	8,487	549
	NAM-DB	4,277	57,840	18.2	9,113	1,131
	TAX-DB	18,480	31,700	10.0	4,497	464
	Comb-DB	23,405	56,530	17.7	8,273	1,048
Gel-based	AM-DB	377	2,619	31.3	1,815	277
	NAM-DB	3,080	2,897	34.6	2,034	714
	TAX-DB	19,036	2,951	35.3	1,777	434
	Comb-DB	22,493	3,684	44.0	2,244	700

[a]Values comprised in 95% confidence interval. [b]Values at 1% global FDR and after manual validation for proteins identified with one peptide. Searching parameters are provided in the section "Materials and Methods."

2.6. Protein annotation

Identified proteins were annotated using *mPies*. For taxonomic and functional annotation, *mPies* used Diamond (Buchfink et al., 2015) to align each identified protein sequences against the non-redundant NCBI DB and the UniProt DB (Swiss-Prot), respectively, and retrieved up to 20 best hits based on alignment score (>80). For taxonomic annotation, *mPies* returned the LCA among the best hits via MEGAN (bit score >80) (Huson et al., 2016). For functional annotation, *mPies* returned the most frequent protein name, with a consensus tolerance threshold >80% of similarity among the 20 best blast hits. Proteins annotated with a score below this threshold were manually validated. Manual validation was straightforward as the main reasons leading to low annotation score were often explained by the characterization of protein isoforms or different sub-units of the same protein (**Appendix B**).

3. Results and Discussion

3.1. Database choice affects the total number of protein identification

The two-rounds search strategy commonly used in recent metaproteomics studies (Russo et al., 2016; Serrano-Villar et al., 2016; Deusch et al., 2017; Gallois et al., 2018) significantly reduced the size of protein search DBs, which in turn increased the total number of identified proteins with both AM-DB and NAM-DB (**Table 3.1**). Overall, the total number of identified proteins was found to be consistent with other metaproteomics studies conducted in marine oligotrophic waters (Morris et al., 2002; Sowell et al., 2009; Williams et al., 2012, 2013; Dong et al., 2014). NAM-DB led to greater protein identifications (gel-based: 714, gel-free: 1,131) than AM-DB (gel-based: 277 and gel-free: 549) and TAX-DB (gel-based: 434 and gel-free: 464) for both proteomics approaches. Comb-DB gave comparable results than NAM-DB in both approaches (gel-based: 700 and gel-free: 1,048). In AM-DB approach, the assembly process involved the removal of reads that cannot be assembled into longer contigs, leading to loss of gene fragments and consequently fewer identified proteins (Cantarel et al., 2011). As high proportions of prokaryotic genomes are protein-coding, gene fragments can directly be predicted from non-assembled sequencing reads (Koonin, 2009). TAX-DB suffered from a reduction of protein detection sensitivity due to its large size in the first-round search, which negatively influenced FDR statistics and protein identification yield (Jagtap et al., 2013).

3.2. Protein search DB affects the taxonomic structure

The proportion of proteins, for which a LCA was found, decreased with lowering taxonomic hierarchy (Domain > Phylum > Class > Order > Family > Genus), independently of the methodology (**Figure 3.1**). The proportion of annotated proteins at the domain, phylum and class levels remained constant with an average of 97.3 ± 1.0, 92.0 ± 1.1, and 80.3 ± 0.8%,

respectively (**Figure 3.1 and Appendix C**). At order level and below, TAX-DB performed the best at assigning an LCA, in both gel-free and gel-based approaches. These results can be explained by the fact that proteins were annotated using sequence-based alignment method (Werner et al., 2019). TAX-DB comprised complete protein sequences from UniProtKB, which allowed accurate annotations. This result confirmed that LCA approach performed at the protein level is affected by DB, as it was previously demonstrated at the peptide level (May et al., 2016).

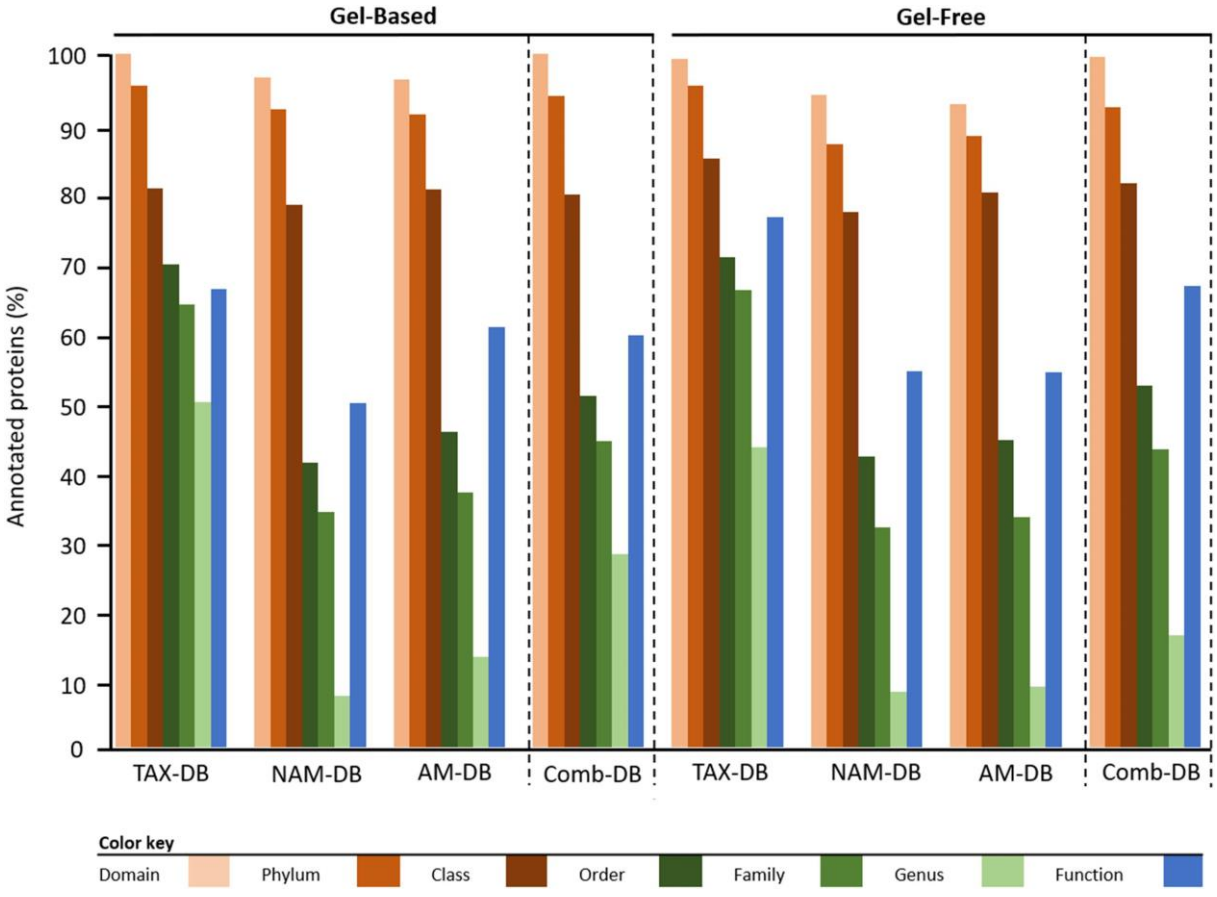


Figure 3.1: Taxonomic and functional distribution depending on database. Comparison of the proportion of proteins for which a consensus annotation was found. Bars represent the percentage of annotated proteins versus the total identified proteins depending on methodology.

At phylum level, most of the proteins identified were assigned to *Proteobacteria* and the least abundant were mainly assigned to *Bacteroidetes* and *Cyanobacteria* (**Table 3.2**). Although *Proteobacteria* showed similar proportion in all metaproteomes (90.9 ± 0.97%), the representativeness of *Bacteroidetes* and *Cyanobacteria* was found to be more variable across the different DBs. The similar distribution can be explained by the fact that the three DBs used in this study were derived from the same metagenome. Indeed, by using distinct data sources (metagenomes and different public repositories), contrasting distributions can be anticipated, as it was recently demonstrated (Timmins-Schiffman et al., 2017). In our study, *Alphaproteobacteria* were found to be the most represented class (72.9 ± 1.9%) followed by *Gammaproteobacteria* (18.2 ± 2.0%), *Flavobacteriia* (4.1 ± 0.5%), and unclassified *Cyanobacteria* (3.0 ± 0.7%) (**Table 3.2**). The dominance of *Alpha*- and *Gammaproteobacteria* was often reported in other marine metaproteomic studies (Morris et al., 2010; Williams et al., 2012; Georges et al., 2014) due to their high distribution in most marine sampling sites. Other studies focusing on sea surface sample also supported the presence of *Cyanobacteria* (Sowell et al., 2009) and *Flavobacteriia* (Williams et al., 2013).

Table 3.2: Taxonomic distribution of proteins assigned at phylum and class levels depending on metaproteomic workflow. Values represent the proportion of annotated proteins with identical taxonomic annotation versus the total identified proteins depending on methodology. The number of peptides detected for each protein was used as quantitative value. Taxa displaying a proportion <1% were gathered into "Other" category.

	Gel-based				Gel-free			
	TAX-DB	NAM-DB	AM-DB	Comb-DB	TAX-DB	NAM-DB	AM-DB	Comb-DB
Phylum								
Proteobacteria	87.2	92.7	95.2	89.8	87.6	92.9	91.5	90.0
Cyanobacteria	6.3	2.9	2.0	4.0	2.6	4.8	1.6	2.0
Bacteroidetes	4.9	4.2	2.4	5.0	7.6	1.4	6.1	5.8
Other (<1%)	1.6	0.2	0.4	1.2	2.2	0.9	0.8	2.2
Class								
Alphaproteobacteria	73.7	81.6	79.9	74.3	69.7	68.4	68.4	67.5
Gammaproteobacteria	12.9	11.5	14.8	14.4	18.0	25.7	24.4	23.5
Flavobacteriia	6.5	3.7	2.2	3.9	6.3	3.5	4.9	4.8
Unclassified Cyanobacteria	3.9	2.7	1.8	5.3	2.5	1.4	1.2	2.1
Other (<1%)	3.0	0.5	1.2	2.1	3.6	1.0	1.0	2.2

At the order level and below, the choice of DB was found to affect both qualitatively and quantitatively the taxonomic distribution, independently of the protein fractionation approach (**Figure 3.2 and Appendices D and E**). Although *Pelagibacterales* and *Rhodobacterales* were found to be the most dominant taxa independently of the methodology, *Pelagibacterales* were found to represent >50% of the total annotated proteins in both NAM-DB and AM-DB (**Figure 3.2a**). *Pelagibacterales* are comprised of the most dominant marine microorganisms in the oceans (Morris et al., 2002) and the dominance of this order in all metaproteomes was in line with prior sea surface metaproteomic studies (Sowell et al., 2009, 2011; Morris et al., 2010; Williams et al., 2012; Georges et al., 2014). The observation of high protein expression profiles assigned to *Rhodobacterales* was also previously reported (Dong et al., 2014). *Flavobacteriales* were overall more represented in the gel-free approach as well as *Cellvibrionales* but only with NAM-DB and AM-DB. *Synechococcales* were more frequently identified in the metaproteomes obtained from the gel-based approach. TAX-DB led to the characterization of many proteins from the following taxa: *Pseudomonadales*, *Rhizobiales*, and *Sphingomonadales*. These taxa were either absent or rarely represented in NAM-DB or AM-DB. As stated above, TAX-DB provided the highest number of annotated proteins, explaining the more diverse distribution obtained using this DB. Interestingly, the taxonomic distributions obtained with Comb-DB were found to be a good compromise between TAX-DB, NAM-DB, and AM-DB (**Figure 3.2a**). As shown in the Venn diagrams provided in **Figure 3.2b**, only one quarter out of the 34 and 41 unique orders observed in gel-based and gel-free approaches, respectively, was common to all DBs. Around 40 and 30% of unique orders were exclusively characterized in TAX-DB and Comb-DB in gel-based and gel-free approaches, respectively, demonstrating the performance of those DBs at extracting the broadest diversity.

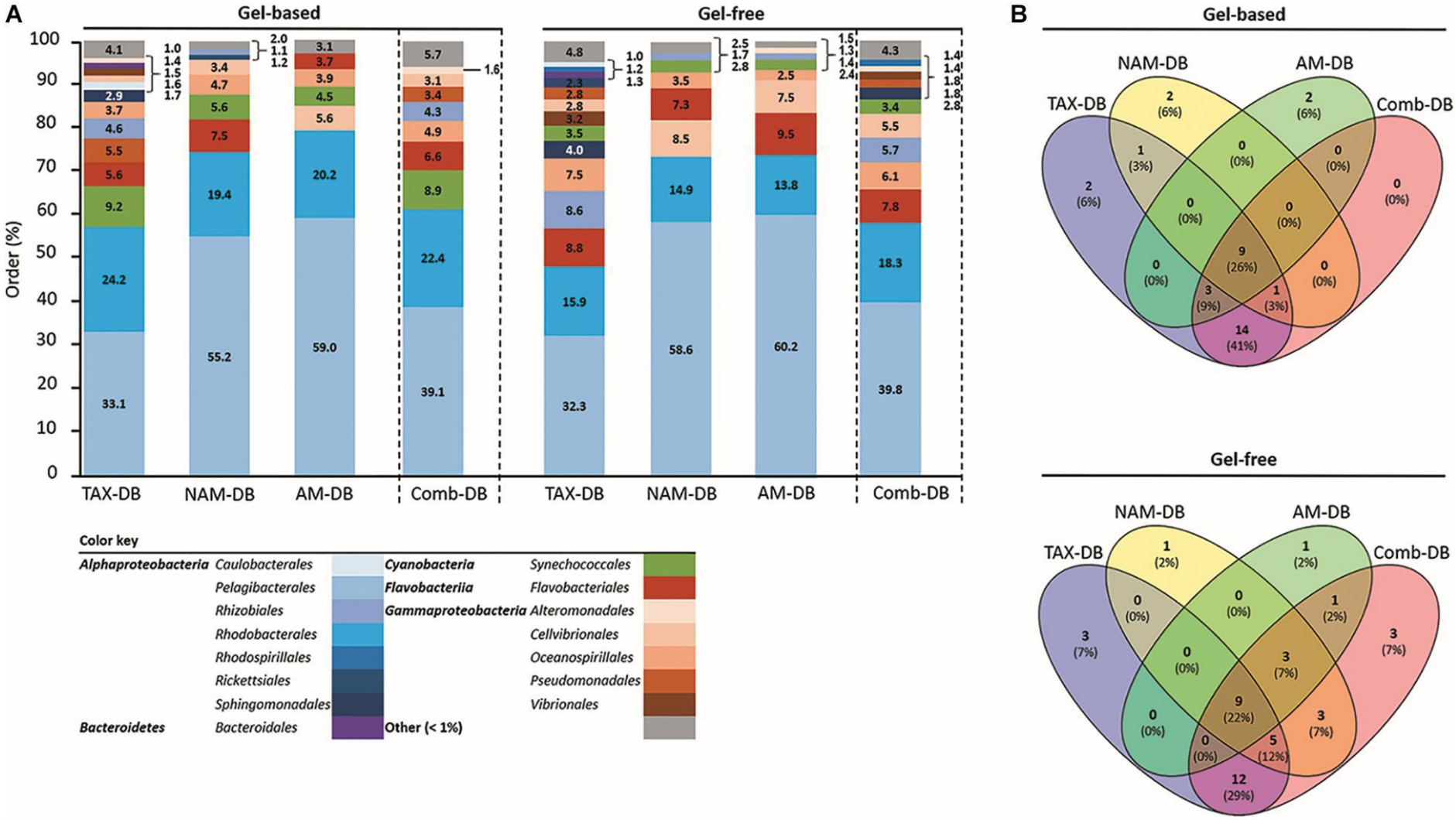


Figure 3.2: Relative taxonomic composition at order level depending on the methodology. (a) Values represent the proportion of proteins with identical taxonomy on total identified protein using TAX-DB, NAM-DB, AM-DB, or Comb-DB in both gel-free and gel-based approaches. The number of peptides detected for each protein was used as quantitative value. Taxa displaying a proportion <1% were gathered into "Other" category. (b) Venn diagrams showing the number of common and unique taxa identified at order level.

3.3. Proteomics workflow and protein search DB affect functional identification

The total number of proteins, for which a functional consensus annotation was found, decreased with the following order: TAX-DB (gel-based: 66%, gel-free: 77%) > AM-DB (gel-based: 61%, gel-free: 54%) > NAM-DB (gel-based: 50%, gel-free: 54%) (**Figure 3.1**). Using Comb-DBs, 59 and 67% of functional annotation were observed in gel-based and gel-free approach, respectively. TAX-DB probably performed better than metagenome-derived DBs at assigning functional annotation because it comprised complete protein sequences from UniProtKB, which facilitated consensus functional annotations using the alignment-based annotation method of *mPies* (Werner et al., 2019). Similarly, the AM-DB contained more complete sequences than NAM-DB because of assembly. Alignment-based functional annotation might be sub-optimal when protein sequences are incomplete. In that case, domain prediction using InterProScan (Jones et al., 2014) would be a complementary approach that would confirm an alignment-based functional consensus and perhaps increase the annotation rate in metagenome-derived databases.

In all metaproteomes, the 60 kDa chaperonin was found to be the most abundant protein (**Figure 3.3a**). The prevalence of chaperonin proteins was previously observed in other marine metaproteomic studies (Sowell et al., 2009, 2011; Williams et al., 2012). The 60 kDa chaperonin is an essential protein involved in large range of protein folding and could potentially act as signaling molecule (Maguire et al., 2002). Moreover, this protein is found in nearly all bacteria. Some taxa, such as *Alphaproteobacteria* or *Cyanobacteria*, often contain several 60 kDa chaperonin homologs (Lund, 2009). On top of its ubiquity and its vital role, the abundance of the 60 kDa chaperonin could be interpreted as a response to environmental stresses exposure (Sowell et al., 2009, 2011; Williams et al., 2012).

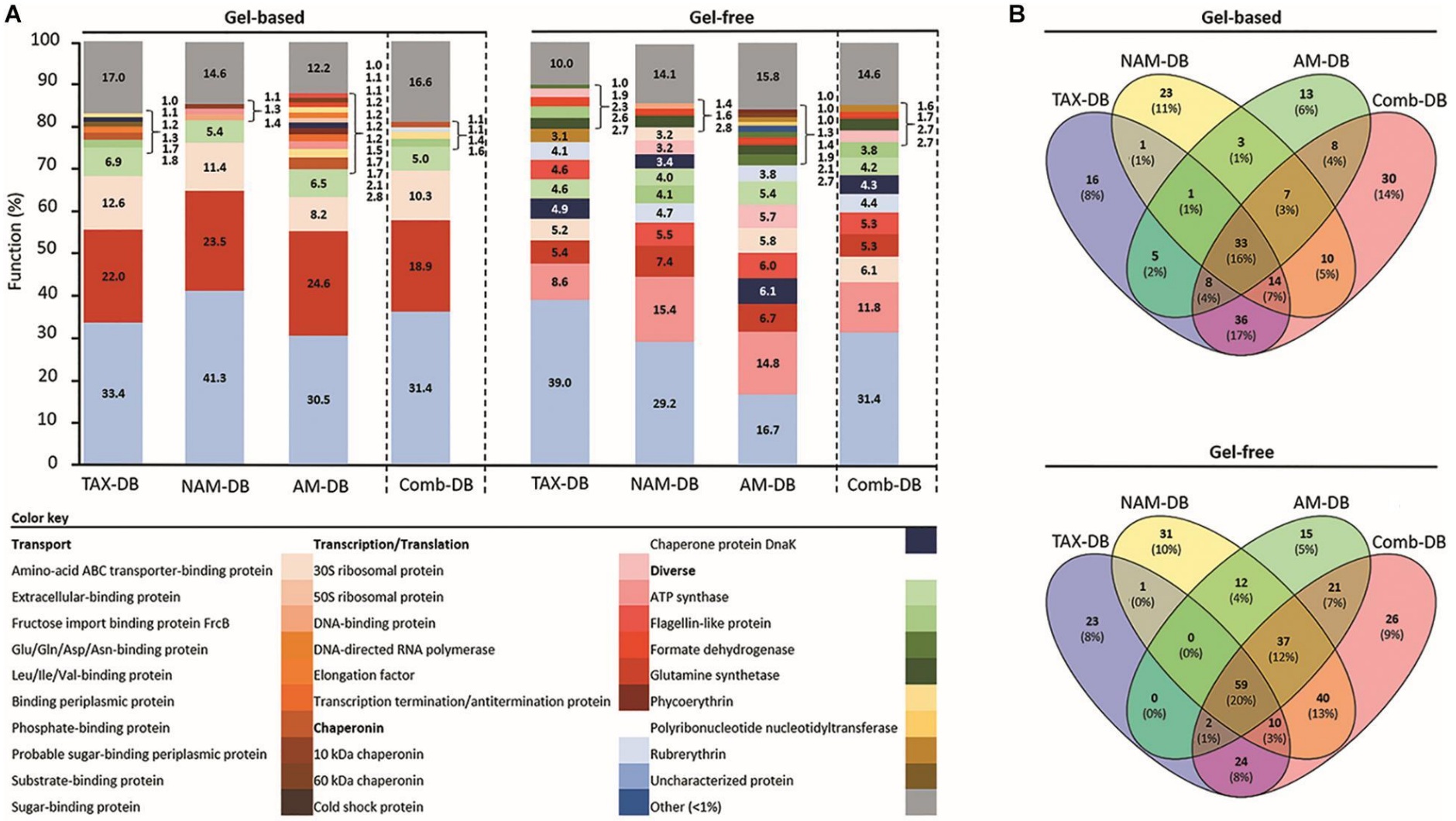


Figure 3.3: Relative functional composition depending on the methodology. Values represent the proportion of proteins with identical functional name on total identified protein using TAX-DB, NAM-DB, AM-DB, or Comb-DB in both gel-free and gel-based approaches. The number of peptides detected for each protein was used as quantitative value. Protein isoforms and/or sub-units were grouped under the same function. Functions displaying a proportion <1% were gathered into "Other" category. (b) Venn diagrams showing the number of common and unique protein functions.

Protein fractionation (gel-based versus gel-free) was found to affect both qualitatively and quantitatively the functional distribution as shown in **Figure 3.3**. The gel-free approach provided the greatest diversity of protein functions in comparison to the gel-based approach (**Figure 3.3a**). Only 16 and 20% of the protein functions were found to be common in all DBs from the gel-based and gel-free approaches, respectively (**Figure 3.3b**). In the gel-based approach, three main functions namely the elongation factor protein, the amino-acid ABC transporter-binding protein, and the ATP synthase were observed in all DBs (**Figure 3.3a**).

In contrast, in the gel-free approach, a higher number of abundant proteins was observed, including: 50S ribosomal proteins, elongation factor protein, ATP synthase, DNA-binding protein, amino-acid ABC transporter-binding protein, 10 kDa chaperonin, and the chaperone protein DnaK (**Figure 3.3a**). In both proteomics approaches, each individual DB allowed the characterization of a significant number of unique protein functions (**Figure 3.3b**). Comb-DB proved to be effective at merging the results obtained from each individual DB, leading to the highest number of identified functions.

3.4. Metaproteomic workflow alters biological interpretation

All proteins annotated at both taxonomic (order rank) and functional levels were clustered and visualized into heatmaps for each DB (**Figure 3.4**). Interestingly, in five out of six heatmaps derived from NAM-DB, AM-DB, and TAX-DB, *Pelagibacterales* was found to be a taxonomic cluster that stood out from all other taxa comprising of *Rhodobacterales*, *Rhizobiales*, *Pseudomonadales*, *Oceanospirillales*, *Cellvibrionales*, *Flavobacteriales*, or *Synechococcales*. An exception was observed for TAX-DB in the gel-based approach where *Rhodobacterales* formed a distinct cluster instead of *Pelagibacterales*. Both *Pelagibacterales* and *Rhodobacterales* clustered apart together from all other taxa when using the Comb-DB. Regarding the functional clustering, the 60 kDa chaperonin was found to stand out all other

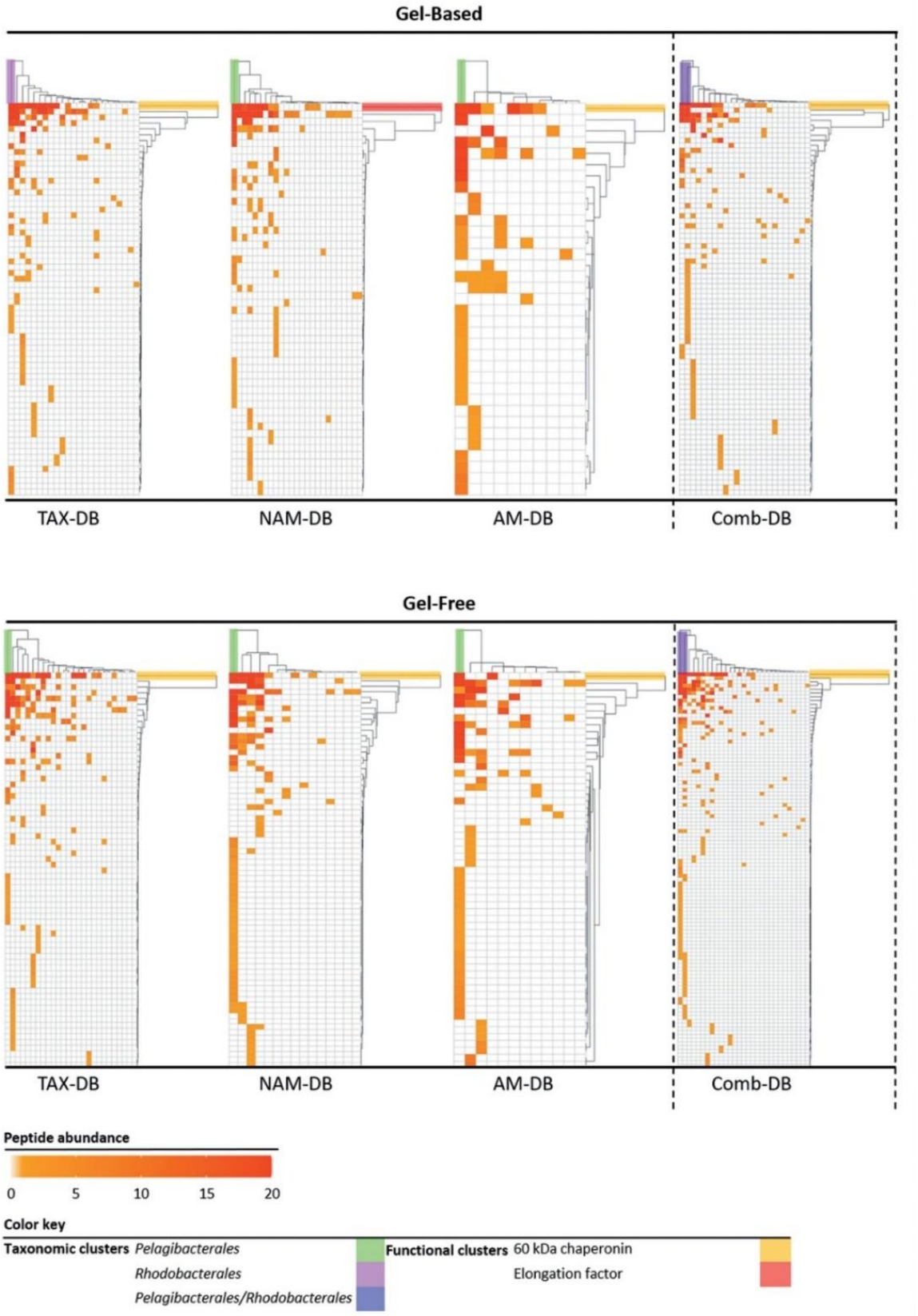


Figure 3.4: Taxonomic and functional linkages depending on the methodology. Proteins annotated at both order (top-cluster) and functional (right-cluster) levels were ranked according to the number of identified peptides. Protein isoforms and/or sub-unit were grouped under the same function. Clusters were determined using complete linkage hierarchical clustering and Euclidean distance metric.

functions apart from NAM-DB in the gel-based approach. Despite the similar trend observed for the most abundant taxa and most represented protein functions for all metaproteomes, **Figure 3.4** clearly shows that the methodology was found to significantly alter the structure/function network.

Interestingly, the detection in all metaproteomes of numerous transporters, mainly for amino-acid/peptide and carbohydrate substrates, across different taxa demonstrated the strategy evolved by bacteria to survive under nutrient-limited environments (**Figure 3.5**) (Button and Robertson, 2000; Zimmer et al., 2000; Hoch et al., 2006; Williams & Cavicchioli, 2014). In contrast, key proteins involved in iron, nitrogen, phosphorous, or vitamin metabolisms were characterized in only few metaproteomes. These results emphasized the risk of misinterpretation on the bacterial response to oligotrophic conditions. The detection of proteins involved in light-harvesting, photosynthesis, and oxidative stress response was found to be particularly dependent of the workflow (**Figure 3.5**). A total of 16 of the 26 proteins were characterized in only one metaproteome, showing that a robust experimental design using multiple methodologies will improve the understanding of the microbial light response. Indeed, combining the information found in all metaproteomes helped at depicting the variety of pigments belonging to photoautotrophs or photoheterotrophs (Giovannoni et al., 2005).

The characterization of the carbon dioxide-concentrating mechanism protein Ccmk together with the ribulose bisphosphate carboxylase (RuBisCO) informed on phototrophs, such as *Synechococcales*, overcome inorganic carbon limitation (Woodger et al., 2003; Sowell et al., 2009). Overall, several oxidative stress-related proteins and numerous chaperonin proteins were identified in all metaproteomes, suggesting the adaptability of the microbial community to cope with oxidative stress. As a reminder, surface water samples were collected in summer at the surface of the Mediterranean Sea, where high solar irradiance was encountered. Chaperones are essential for coping with UV-induced protein damage and maintaining proper protein function

(Matallana-Surget et al., 2013). Consequently, those metaproteomics results suggest that strategies used by microorganisms to cope with high solar radiation could be similar to the ones extensively described in axenic cultures using microcosms experiments (Matallana-Surget & Wattiez, 2013).

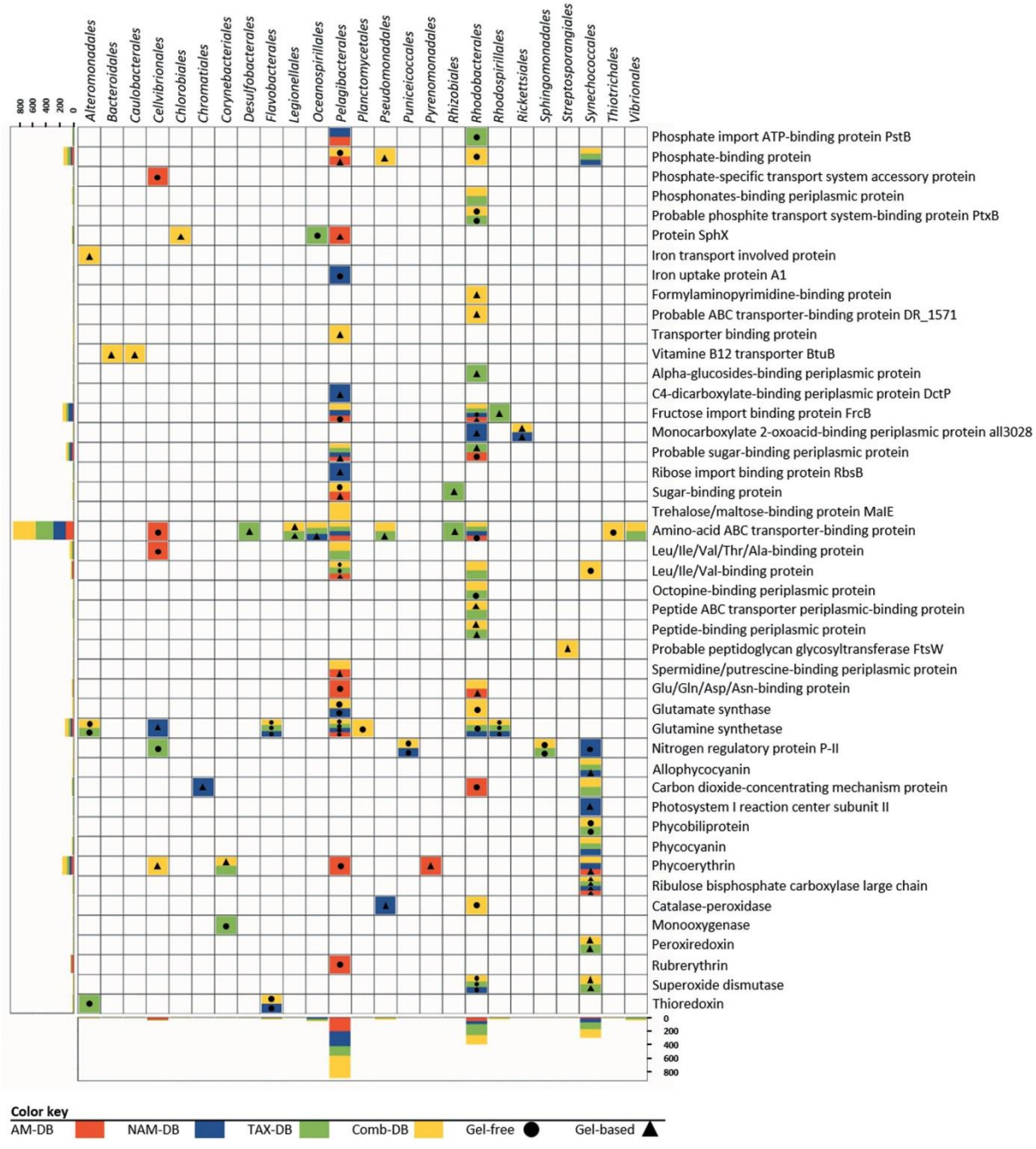


Figure 3.5: Impact of metaproteomic workflow on the taxonomic and functional diversity. This figure focuses on the proteins involved in nutrient transport, nitrogen assimilation, light harvesting, and oxidative stress response for each methodology. Horizontal and vertical bar charts correspond to the total number of peptides detected for a given function (*y*-axis) or order (*x*-axis) in all metaproteomes. Protein isoforms and/or sub-unit were grouped under the same function. The lack of symbol in colored boxes means that the protein was observed in both gel-free and gel-based approaches.

4. Conclusion

Metaproteomics enables to progress beyond a mere descriptive analysis of microbial community diversity and structure, providing specific details on which bacteria, and which pathways of those key players, are impacted by possible perturbations. Nevertheless, using this powerful tool without fully apprehending the limitations could lead to significant misinterpretations, especially in the case of comparative metaproteomic studies. This study clearly evidenced the implications of critical decisions in metaproteomic workflow. Our findings lead to the general recommendation of diversifying, when possible, the protein search database as well as protein fractionation, especially if only one condition/ecosystem was studied. A robust diversified workflow allows crossing information from multiple metaproteomes in order to accurately describe the functioning of microbial communities. In a comparative metaproteomic study however, the best compromise relies on the creation of a Comb-DB. Our findings will undoubtedly serve future studies aiming at reliably capturing how microorganisms operate in their environment.

Chapter 4: Diel cycle in marine picoplankton

Diel Protein Regulation of Marine Picoplanktonic Communities Assessed by Metaproteomics

Géron, **A**., Werner, J., Lebaron, P., Wattiez, R., & Matallana-Surget, S. *Microorganisms,* 9(12), (2021), 2621.

Abstract

The diel cycle is of enormous biological importance in that it imposes temporal structure on ecosystem productivity. In the world's oceans, microorganisms form complex communities that carry out about half of photosynthesis and the bulk of life-sustaining nutrient cycling. How the functioning of microbial communities is impacted by day and night periods in surface seawater remains to be elucidated. In this study, we compared the day and night metaproteomes of the free-living and the particle-attached bacterial fractions from picoplanktonic communities sampled from the northwest Mediterranean Sea surface. Our results showed similar taxonomic distribution of free-living and particle-attached bacterial populations, with *Alphaproteobacteria, Gammaproteobacteria* and *Cyanobacteria* being the most active members. Comparison of the day and night metaproteomes revealed that free-living and particle-attached bacteria were more active during the day and the night, respectively. Interestingly, protein diel variations were observed in the photoautotroph *Synechococcales* and in (photo)-heterotrophic bacteria such as *Flavobacteriales*, *Pelagibacterales* and *Rhodobacterales*. Moreover, our data demonstrated that diel cycle impacts light-dependent processes such as photosynthesis and UV-stress response in *Synechococcales* and *Rhodobacterales*, respectively, while the protein regulation from the ubiquitous *Pelagibacterales* remained stable over time. This study unravels, for the first time, the diel variation in the protein expression of major free-living and particle-attached microbial players at the sea surface, totaling an analysis of eight metaproteomes.

1. Introduction

Microorganisms in marine ecosystems are extremely diverse, dominate biomass and play key roles in biogeochemical processes (Hedges & Oades, 1997; Kujawinski, 2011). Picoplankton (i.e., the microorganisms of a size ranging between 0.2–2 µm) carry out up to half of the world ocean's primary production and the bulk of life-sustaining nutrient cycling (Cole et al., 1988). Marine picoplanktonic communities are composed of both free-living and particle-attached microorganisms. A comparison of these bacterial fractions in coastal environments showed differences in cell abundance (Turley & Stutt, 2000), morphology and metabolic activity (Simon et al., 2002). In terms of phylogenetic diversity, studies suggested that free-living and particle-attached communities were fundamentally different (Acinas et al., 1999; Crump et al., 1999; DeLong et al., 1993), while others reported high similarities between both fractions (Ghiglione et al., 2007; Hollibaugh et al., 2000; Moeseneder et al., 2001).

The diel oscillation of solar radiation reaching the Earth's surface temporally structures biological events, activities, and physiological processes across all kingdoms of life (Ditty et al., 2009). Day/night changes were found to modulate the functioning of sea surface picoplanktonic communities on the following processes: metabolites consumption (Galí et al., 2013; Kuipers et al., 2000), viral infection (Winter et al., 2004), DNA/protein synthesis and dissolved organic carbon distribution (Gasol et al., 1998). In the northwest (NW) Mediterranean Sea, under oligotrophic conditions, free-living bacteria were found to be more abundant than particle-attached bacteria (Ghiglione et al., 2007). Moreover, bacterial activity estimated from 3H-leucine incorporation rates showed that free-living bacteria contributed the most to bacterial activity during the day and night, while higher cell-specific activity was found in particle-attached bacteria (Ghiglione et al., 2007).

The development of omics approaches has improved the understanding of marine microbial assemblages (Matallana-Surget et al., 2018). Environmental metatranscriptomic studies reported distinct day and night metabolic activities of marine microorganisms from oligotrophic marine environments (Gilbert et al., 2010; Poretsky et al., 2009). Microbial assemblages isolated from the North Pacific subtropical gyre showed an overabundance of transcripts for photosynthesis, C1 metabolism and oxidative phosphorylation during the day (Poretsky et al., 2009). However, housekeeping activities, such as amino acid or vitamin biosynthesis, were overrepresented at night. Transcripts of genes involved in light-driven processes were found in higher abundance during daytime, in surface marine picoplankton sampled from the Western English Channel (Gilbert et al., 2010). Diel transcriptional rhythms in *Cyanobacteria* were evidenced together with diel oscillations in different heterotrophic bacterial groups including photoheterotrophic and proteorhodopsin-containing bacteria (Ottesen et al., 2014). Phylogenetic analysis of gene transcripts revealed that the composition of marine microbial assemblages was stable over the day and night periods, especially for the most abundant taxa (Ottesen et al., 2014). To what extent picoplanktonic communities are collectively entrained by the day and night periods and rhythmically regulate their protein expression remains poorly documented.

Metaproteomics allows for the characterization of the total proteins within microbial communities (Wilmes & Bond, 2004) and, in association with other omics, deciphers the functional complexity of microbial ecosystems (Franzosa et al., 2015). Since the first environmental metaproteomic study (Kan et al., 2005), this method rapidly expanded and broadened our knowledge of marine ecosystems (Saito et al., 2019). For example, marine metaproteomic revealed the extreme microbial competition for nutrients in oligotrophic systems (Sowell et al., 2009; Williams & Cavicchioli, 2014), provided insights into the dynamics in organic matter transformation by microorganisms (Bergauer et al., 2018; Hawley

et al., 2014) and showed the spatiotemporal variation in metabolic activities in oceanic plankton communities (Georges et al., 2014). Environmental metaproteomics is a growing discipline, hampered by the inherent complexity of natural microbial assemblages (Géron et al., 2019). Over the past years, the development of sampling protocols, fast scanning high-resolution mass spectrometers and protein identification and annotation software significantly improved the metaproteomic workflow (Heyer et al., 2017; Matallana-Surget et al., 2018; Saito et al., 2019). Marine oligotrophic waters still present significant challenges for metaproteomic studies because of (i) the low bacterial biomass preventing high protein rate recovery, (ii) the difficulty of separating prokaryotes from microeukaryotes and (iii) the protein inference issue (Géron et al., 2019; Matallana-Surget et al., 2018; Saito et al., 2019).

In this study, we compared the day and night metaproteomes of both free-living (0.2–0.8 µm) and the particle-attached (>0.8 µm) bacterial fractions sampled at the surface of NW Mediterranean Sea in summer. A combined protein search database allowed us to maximize the number of protein identifications (Géron et al., 2019). The protein inference issue, commonly encountered in metaproteomics, was overcome using taxonomic and functional consensus protein annotation, providing an accurate assessment of the diel variation (Werner et al., 2019). To the best of our knowledge, this is the first metaproteomics study that depicts day and night metaproteomes of marine picoplankton.

2. Materials and Methods

2.1. Water sampling

Seawater sampling was performed in summer (June 2014) at the SOLA station, located 500 m offshore of Banyuls-sur-mer, in the NW Mediterranean Sea (42°49′ N, 3°15′ W). Samples were collected at sunset and sunrise during two consecutive days and consisted of 70

L of sea surface water each. Water was pre-filtered onto a 5 µm mesh and sequentially filtered through 0.8 and 0.2 µm pore-sized filters (polyethersulfone membrane filters, PES, 142 mm, Millipore, Burlington, Massachusetts, United States) to collect the particle-attached and the free-living bacterial fractions, respectively. A pre-filtration onto a 5 µm mesh was mandatory to prevent the studied bacterial fractions (0.8 µm and 0.2 µm) from being contaminated by eukaryotic organisms, which would otherwise alter the metaproteomic workflow. As a reminder, metaproteomics allows for the characterization of the most abundant proteins. The eight filters were flash frozen in liquid nitrogen before storage at −80 °C.

The physicochemical parameters were provided by the Service d'Observation en Milieu Littoral (SOMLIT). On site average temperature and salinity in June were 18.7 ± 0.7 °C and 37.8 ± 0.1 psu, respectively, as provided by SOMLIT. pH was stable, with an average of 8.26 ± 0.04. Nutrient concentrations averaged 0.03 ± 0.01 µM NH_4^+, 0.05 ± 0.03 µM NO_3^-, 0.01 ± 0.001 µM NO_2^-, 0.02 ± 0.01 µM PO_4^{3-} and 0.75 ± 0.09 µM $Si(OH)_4$.

2.2. Protein isolation

The filters were cut using aseptic procedures and suspended in a lysis buffer containing 8 M Urea/2 M Thiourea, 10 mM HEPES and 10 mM dithiothreitol. Filters were subjected to five freeze–thaw cycles in liquid N_2 to release cells from the membrane. Cells were mechanically broken by sonication on ice (5 cycles of 1 min with tubes on ice, amplitude 40%, 0.5 pulse rate) and subsequently centrifuged at 16,000 g at 4 °C for 15 min. To remove particles that did not pellet during the centrifugation step, the protein suspension was filtered through a 0.22 µm syringe filter and transferred into a 3 kDa cutoff Amicon Ultra-15 filter unit (Millipore) for protein concentration. Proteins were precipitated with cold acetone overnight at −80 °C, with an acetone/aqueous protein solution ratio of 4:1. Total protein concentration was determined by a Bradford assay, using the Bio-Rad Protein Assay kit (Bio-Rad, Hertfordshire,

UK) according to manufacturer's instructions, with bovine γ-globulin as a protein standard. Protein samples were reduced with 25 mM dithioerythritol (DTE) at 56 °C for 30 min and alkylated with 50 mM iodoacetamide at room temperature for 30 min. Gel-free liquid chromatography tandem mass spectrometry was performed utilizing a trypsic digestion (sequencing grade modified trypsin, Promega, Madison, Wisconsin, États-Unis) overnight at 37 °C, with an enzyme/substrate ratio of 1:25.

2.3. Liquid chromatography tandem mass spectrometry analysis

Purified peptides from digested protein samples were identified using a label-free shotgun approach on an UHPLC-HRMS platform composed of an eksigent 2D liquid chromatograph and an AB SCIEX Triple TOF 5600. Peptides were separated on a 25 cm C18 column (Acclaim pepmap 100, 3 µm, Dionex, Sunnyvale, Californie, États-Unis) by a linear acetonitrile (ACN) gradient (5–35% (v/v), in 15 or 120 min) in water containing 0.1% (v/v) formic acid at a flow rate of 300 nL min−1. Mass spectra (MS) were acquired across 400–1500 m/z in high-resolution mode (resolution >35,000) with 500 ms accumulation time. Six microliters of each fraction were loaded onto a pre-column (C18 Trap, 300 µm i.d. × 5 mm, Dionex) using the Ultimate 3000 system, delivering a flow rate of 20 µL/min loading solvent (5% (v/v) acetonitrile (ACN), 0.025% (v/v) TFA). After a 10 min desalting step, the pre-column was switched online with the analytical column (75 µm i.d.× 15 cm PepMap C18, Dionex) equilibrated in 96% solvent A (0.1% (v/v) formic acid in HPLC-grade water) and 4% solvent B (80% (v/v) ACN, 0.1% (v/v) formic acid in HPLC-grade water). Peptides were eluted from the pre-column to the analytical column and then to the mass spectrometer, with a gradient from 4–57% solvent B for 50 min and 57–90% solvent B for 10 min at a flow rate of 0.2 µL min−1 delivered by the Ultimate pump. Positive ions were generated by electrospray and the instrument was operated in a data-dependent acquisition mode, described as follows: MS scan

range: 300–1500 m/z, maximum accumulation time: 200 ms, ICC target: 200,000. The top 4 most intense ions in the MS scan were selected for MS/MS in dynamic exclusion mode: ultrascan, absolute threshold: 75,000, relative threshold: 1%, excluded after spectrum count: 1, exclusion duration: 0.3 min, averaged spectra: 5 and ICC target: 200,000. Metaproteomic raw data are available in the iProx public platform (Ma et al., 2019) (Project ID: IPX0002008000; Subproject IDs: IPX0002008001 (free-living fractions), IPX0002008002 (particle-attached fractions)).

2.4. Ocean Sampling Day 2014 metagenomic data set

Metagenomic data from the Ocean Sampling Day 2014 (OSD14) were downloaded from the EMBL-EBI MGnify platform (Project number: ERP009703, sample: OSD14_2014_06_2m_NPL022, run ID: ERR771073). Briefly, water was sampled at the same location (42°49′ N, 3°15′ W) and month (June 2014) as our metaproteomic study, using a CTD rosette with Niskin bottles. Water was filtered on a 0.22 µm pore-sized filter and stored at −80 °C until subsequent DNA extraction and sequencing. Illumina sequencing was performed using an Illumina MiSeq instrument, and reads were processed using the OSD14 pipeline version 4.0. Briefly, paired-end overlapping reads were merged using SeqPrep (John, 2011) and low-quality sequences were trimmed using Trimmomatic (Bolger et al., 2014). Adapter sequences and sequences <100 nucleotides in length were removed using Biopython (Cock et al., 2009). Infernal (Nawrocki et al., 2009) was used for ncRNAs identification and cmsearch deoverlap script was used to remove lower scoring overlaps. Genes were called using FragGeneScan (short reads) (Rho et al., 2010) and Prodigal (Hyatt et al., 2010). InterProScan (Jones et al., 2014) was used for gene identification and MAPseq for taxonomic annotation (Matias Rodrigues et al., 2017).

2.5. Databases creation and protein identification

Protein identification was performed with ProteinPilot (ProteinPilot Software 5.0.1; Revision: 4895; Paragon Algorithm: 5.0.1.0.4874; AB SCIEX, Framingham, MA, USA) (Matrix Science, London, UK; v. 2.2). Paragon searches were conducted using LC MS/MS Triple TOF 5600 System instrument settings. Other parameters used for the search were as follows: Sample Type: Identification, Cys alkylation: Iodoacetamide, Digestion: Trypsin, ID Focus: Biological Modifications and Amino acid substitutions, Search effort: Thorough ID, Detected Protein Threshold [Unused ProtScore (Conf)] > 0.05 (10.0%).

Three protein search databases (DBs) were created with *mPies* v. 0.9 (Werner et al., 2019), using the OSD14 metagenome as a template. The three DBs were: (i) a non-assembled metagenome-derived DB (NAM-DB), (ii) an assembled metagenome-derived DB (AM-DB) and (iii) a taxonomy-derived DB (TAX-DB) (Géron et al., 2019). An initial protein search was performed for each sample against the three DBs individually. Subsequently, each DB was restricted to the protein sequences identified in the first-round search. The resulting DBs were merged, and redundant protein sequences were removed, leading to a unique combined DB per sample. Finally, a second protein search was performed for each sample against their respective combined DB, except for the 0.8 µm samples, where all combined DB were merged to increase the identification yield. The identified proteins were selected based on a FDR threshold of 1%, calculated at the protein level was used for each protein searches. Proteins identified with one single peptide spectrum were validated by manual inspection of the MS/MS spectra, ensuring that a series of at least five consecutive sequence-specific b-and y-type ions was observed.

2.6. Protein annotation and downstream analyses

Identified proteins were annotated using *mPies* (Werner et al., 2019). The *mPies* tool used Diamonds (Buchfink et al., 2015) to align each identified protein sequences against NCBI

nr and UniProt DBs, respectively, and retrieved up to 20 best hits based on alignment score. For taxonomic annotation, *mPies* returned the last common ancestor (LCA) among the best NCBI hits via MEGAN (bit score >80) (Huson et al., 2016). For functional annotation, *mPies* returned the most frequent protein name, with a consensus tolerance threshold above 80% similarity amongst the 20 best UniProt hits. Proteins annotated with a score below this threshold were manually validated.

3. Results

3.1. Diel structure of the microbial communities

The reads encoding for the 16S rRNA were extracted from the OSD14 metagenome and reflected the abundance of each operational taxonomic unit (OTU) in the studied bacterial communities. The metagenome taxonomic structure showed that *Proteobacteria* was the most abundant phylum, with 66.89% of the total detected 16S rRNA bacterial reads, followed by *Bacteroidetes* (15.51%) and *Cyanobacteria* (12.22%) (**Table 4.1**). *Alphaproteobacteria* was the class with the highest representation (47.35%), followed by *Gammaproteobacteria* (17.77%), *Flavobacteriia* (14.32%) and unclassified *Cyanobacteria* (12.33%) (**Table 4.1**). At order level, *Pelagibacterales* reads were dominant (28.85%), followed by *Flavobacteriales* (16.48%) and, to a lesser extent, *Rickettsiales* (10.99%), *Oceanospirillales* (8.85%), *Rhodobacterales* (7.17%) and *Cellvibrionales* (6.10%) (**Figure 4.1**).

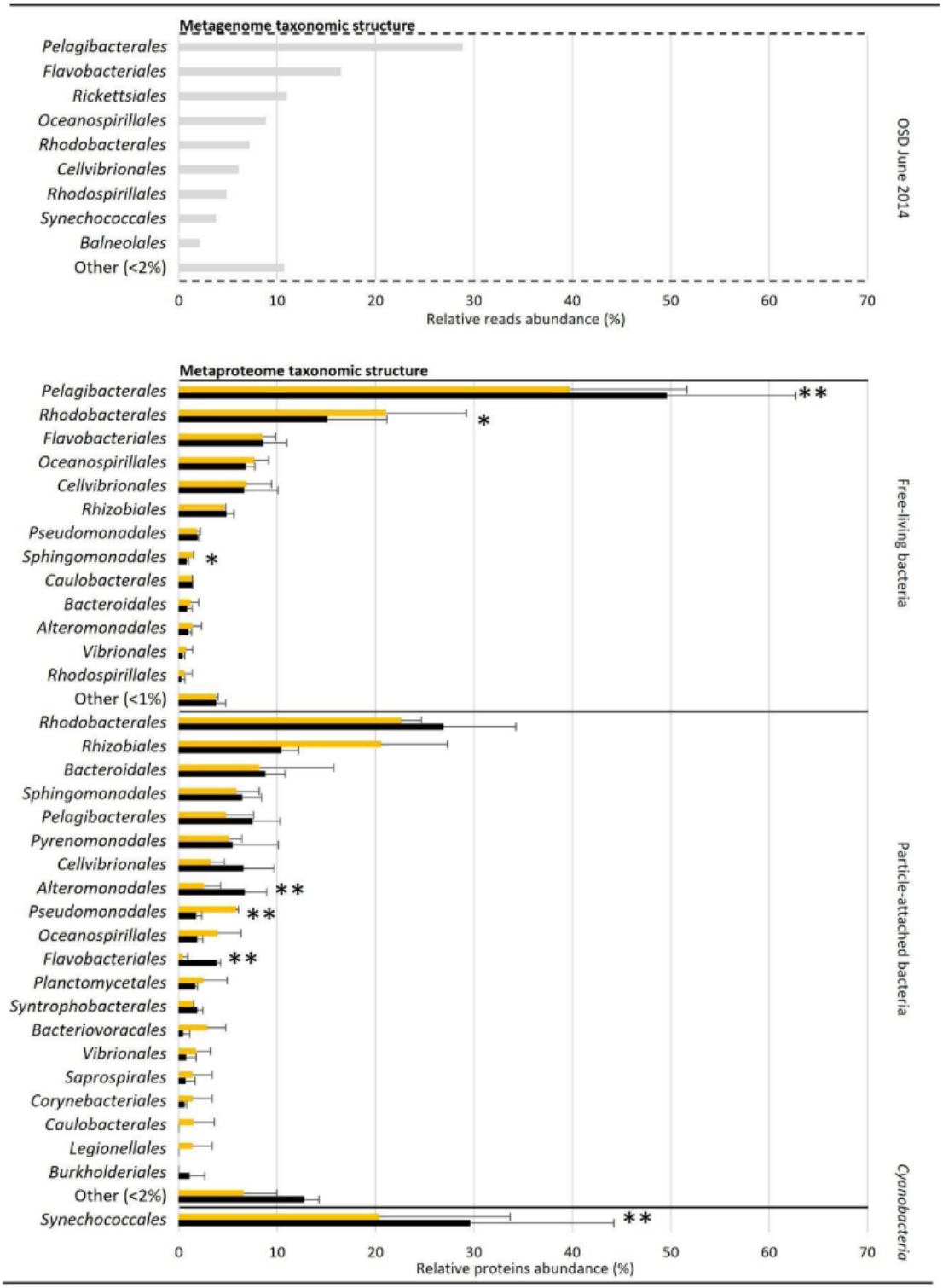


Figure 4.1: Diel variability of active bacterial community orders. Metagenomic data consisted of the percentage of total 16S rRNA reads observed over the OSD14 sampling effort (day for 0.2 µm pore-sized fraction). Metaproteomic data consisted of the average percentage of total unique peptide spectra detected per order for each metaproteome (day (yellow) and night (black) for both 0.2 and 0.8 µm size-fractions, n = 2). The least abundant taxa (<2% of reads and <1 or 2% of peptide spectra) were classified in "Other" category. Significative differences between day and night samples are shown with a * (p-value ≤ 0.1) or ** (p-value ≤ 0.05) and were calculated with a paired t-test.

Table 4.1: Diel variability of major bacterial phyla and classes. Metagenomic data consisted of the percentage of total 16S rRNA bacterial reads observed over the OSD14 sampling effort (day for 0.2 µm pore-sized fraction). Metaproteomic data consisted of the average percentage of total unique bacterial peptide spectra detected per phylum or class for each metaproteome (day (yellow) and night (black) for both 0.2 and 0.8 µm pore-sized fractions, n = 2). The least abundant taxa (<1% of total reads or peptide spectra) were classified in "Other" category. Significative differences between day and night samples are shown with a * (p-value ≤ 0.1) or ** (p-value ≤ 0.05) and were calculated with a paired t-test.

	Metagenome Taxonomic Structure	Metaproteome Taxonomic Structure							
		0.2 µm Size-Fraction				0.8 µm Size-Fraction			
	OSD June 2014	Day	SD	Night	SD	Day	SD	Night	SD
Total reads/proteins	761	550	±49	452	±4	123	±28	170	±28
Phylum									
Proteobacteria	66.89	** 89.34	±1.62	** 92.43	±1.19	34.75	±7.92	33.19	±6.02
Bacteroidetes	15.51	6.48	±0.61	5.48	±0.40	2.71	±1.35	4.47	±0.84
Cyanobacteria	12.22	2.30	±1.83	0.43	±0.15	60.52	±8.06	60.83	±6.60
Rhodothermaeota	1.84	0.69	±0.22	0.56	±0.61				
Planctomycetes	0.13			0.03	±0.05	0.75	±0.79	0.46	±0.03
Other (<1%)	3.42	1.19		1		1.26		1.05	
Class									
Alphaproteobacteria	47.35	* 67.06	±2.88	* 71.54	±4.87	20.88	±2.04	20.86	±4.04
Gammaproteobacteria	17.77	23.29	±0.69	21.57	±4.01	12.79	±4.64	11.41	±0.97
Flavobacteriia	14.32	4.77	±0.75	4.94	±0.75	** 0.10	±0.14	** 1.46	±0.10
Unclassified *Cyanobacteria*	12.33	0.54	±0.77	0.45	±0.16	60.25	±7.55	60.26	±5.63
Bacteroidia	0.13	0.68	±0.47	0.51	±0.32	2.27	±1.95	2.65	±0.86
Deltaproteobacteria		0.06	±0.08			* 0.43	±0.03	* 0.78	±0.21
Oligoflexia		0.06	±0.09	0.07	±0.09	0.87	±0.66	0.44	±0.29
Planctomycetia				0.03	±0.05	0.77	±0.80	0.49	±0.03
Other (<1%)	8.09	3.54		0.9		1.64		1.64	

The metaproteomic analyses were performed on duplicate day and night samples and showed that the average number of identified bacterial proteins on the four 0.2 µm pore-sized filters was 550 ± 49 in the day (n = 2) and 452 ± 4 at night (n = 2) (**Table 4.1**). The active taxa of this bacterial fraction were largely characterized as *Proteobacteria* (average relative protein abundance during the day (D): 89.34 ± 1.62%, during the night (N): 92.43 ± 1.19%), followed by *Bacteroidetes* (D: 6.48 ± 0.61%, N: 5.48 ± 0.40%). Few *Cyanobacteria* proteins (D: 2.30 ± 1.83%, N: 0.43 ± 0.15%) were also observed (**Table 4.1**). At class level, *Alphaproteobacteria*, *Gammaproteobacteria* and *Flavobacteriia* were found to be the most represented (**Table 4.1**). At order level, *Pelagibacterales* proteins (D: 39.8 ± 11.9%, N: 49.6 ± 13.1%) were dominant, especially during the night (p value = 0.03), followed by *Rhodobacterales* (D: 21.1 ± 8.1%, N:

15.1 ± 6.1%). *Rhodobacterales* and *Sphingomonadales* proteins were represented more during the day (p values = 0.08 and 0.07, respectively) (**Figure 4.1**).

On the four 0.8 µm pore-sized filters, an average of 123 ± 28 (n = 2) and 170 ± 28 (n = 2) bacterial proteins were identified at day and at night, respectively (**Table 4.1**). *Cyanobacteria* (D: 60.25 ± 7.55%, N: 60.83 ± 6.60%) were the most abundant active players of this fraction, followed by *Proteobacteria* (D: 34.75 ± 7.92%, N: 33.19 ± 6.02%). Classes were mainly represented by unclassified *Cyanobacteria* and, to a lesser extent, by *Alphaproteobacteria* and *Gammaproteobacteria* (**Table 4.1**). At order level, *Rhodobacterales* proteins (D: 22.6 ± 2.1%, N: 26.9 ± 7.4%) were dominant during the day and the night (**Figure 4.1**). Proteins from *Pseudomonadales* were more abundant in the day (p value = 0.05), and *Alteromonadales* and *Flavobacteriales* at night (p values = 0.03 and 0.01, respectively) (**Figure 4.1**).

Cyanobacteria were more represented on the 0.8 µm pore-sized filters (D: 60.52 ± 8.06%, N: 60.83 ± 6.60%) compared to the 0.2 µm pore-sized filters (D: 2.30 ± 1.83%, N: 0.43 ± 0.15%) (**Table 4.1**), due to their rod-shaped morphology (>0.2 µm in length) (Yu et al., 2015). Therefore, all identified cyanobacterial proteins were grouped for clarity purposes and compared with free-living and particle-attached bacteria (**Figure 4.1**). *Cyanobacteria* were exclusively characterized as *Synechococcales*, which represented 20.4 ± 13.3% (D) and 29.6 ± 14.6% (N) of the total identified proteins in both 0.2 and 0.8 µm fractions (**Figure 4.1**). The total *Synechococcales* proteins were significantly more abundant at night (p value = 0.03).

3.2. Diel functioning of the microbial communities

The proteins characterized in the eight metaproteomes were grouped into five functional categories: (i) Protein folding and stress response, (ii) energy metabolism and compound biosynthesis, (iii) replication, transcription, and translation, (iv) transport and (v) cell mobility, structure, and division. Overall, the protein functions detected in the free-living bacterial

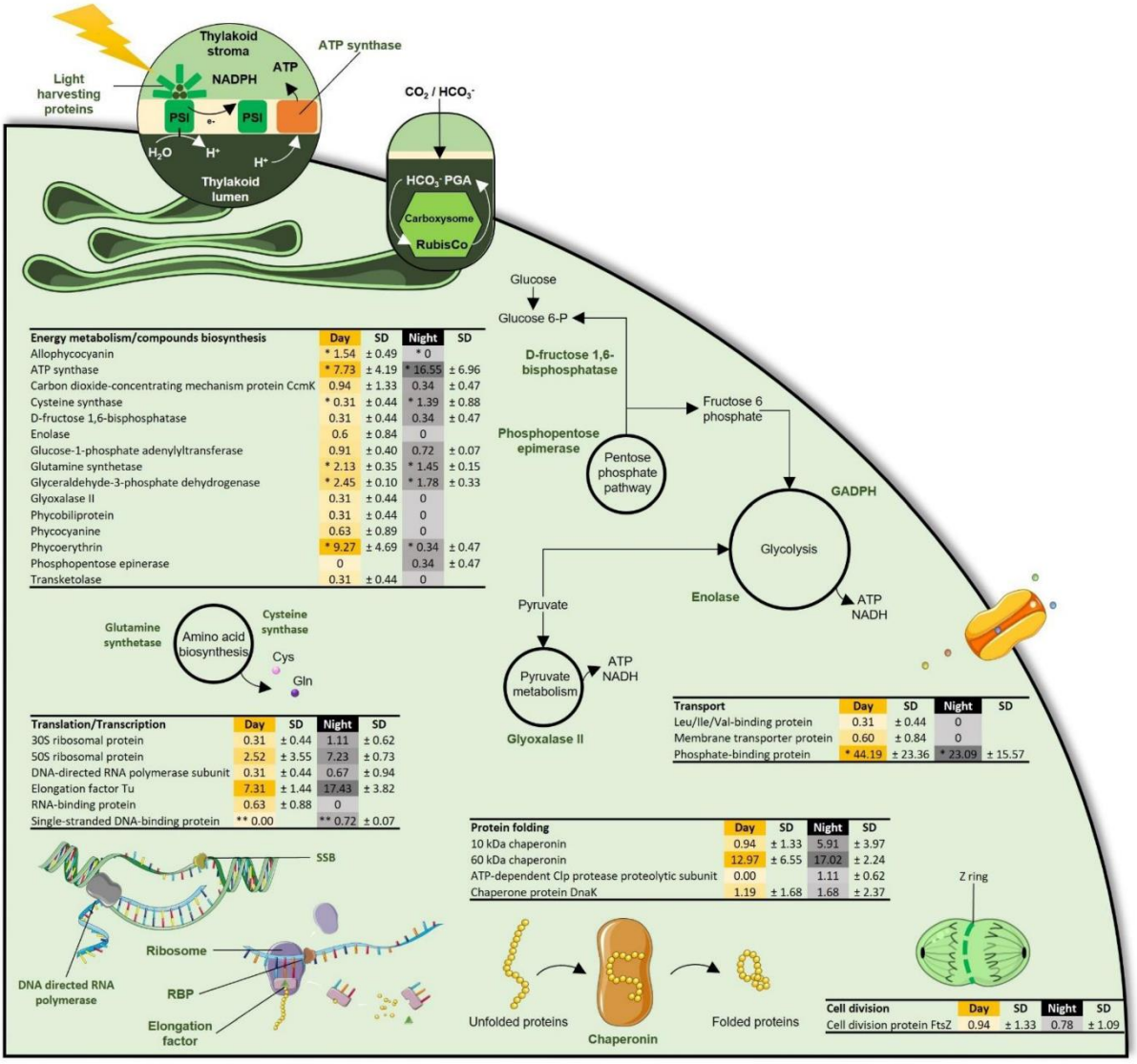
community were found to be stable despite the diel variation, with only two proteins—the glyceraldehyde-3-phosphate dehydrogenase and the actin-like protein—being significantly more represented at day and at night, respectively (**Table 4.2**). In contrast, the proteins expressed by the particle-attached bacteria and *Cyanobacteria* showed more important diel changes, mainly in energy metabolism and compound biosynthesis processes (**Table 4.2**).

3.3. Protein folding and response to stress

Proteins involved in protein folding were detected in the eight metaproteomes, with the 60 kDa chaperonin being the major protein function, followed by the 10 kDa chaperonin and the chaperone protein DnaK (**Table 4.2**). While the protein folding process was equally expressed in the *Cyanobacteria Synechococcales* (**Figure 4.2**), diel variations in chaperonin expression were observed in some free-living and particle-attached bacterial orders: during the day, the 60 kDA chaperonin was more abundant in particle-attached *Rhodobacterales* (p value = 0.10) and in free-living *Rhizobiales* (p value = 0.04), and the chaperone protein DnaK was more represented in particle-attached *Rhizobiales* (p value = 0.02) and free-living *Pelagibacterales* (p value = 0.02) (**Figures 4.3 and 4.4**). Interestingly, the 10 kDa chaperonin was found to be differentially regulated over time among free-living and particle-attached *Sphingomonadales* (p values = 0.09 and 0.09, respectively), as the trends in protein abundance were higher during the day and night, respectively (**Figures 4.3 and 4.4**).

Table 4.2: Diel variation of the functionality of microbial communities. Values consisted of the average percentage of total unique peptide spectra detected per protein function in free-living bacteria, particle-attached bacteria, and all *Cyanobacteria* during both day (yellow, n = 2) and night (black, n = 2). The least abundant functions (<1% of peptide spectra) were classified in "Other" category. Significative differences between day and night samples are shown with a * (p value ≤ 0.1) or ** (p value ≤ 0.05) and were calculated with a paired t-test.

	Free-Living Bacteria				Particle-Attached Bacteria				*Cyanobacteria*			
Protein folding/ Stress response	**Day**	**SD**	**Night**	**SD**	**Day**	**SD**	**Night**	**SD**	**Day**	**SD**	**Night**	**SD**
10 kDa chaperonin	5.36	±0.22	5.52	±0.74	3.48	±1.66	4.64	±1.32	0.94	±1.33	5.91	±3.97
60 kDa chaperonin	32.46	±1.19	30.29	±5.42	** 20.54	±6.29	** 25.52	±6.96	13.02	±6.48	17.02	±2.24
ATP-dependent Clp protease proteolytic subunit											1.11	±0.62
Chaperone protein DnaK	3.29	±0.04	3.72	±0.62	5.77	±3.22	4.01	±1.01	1.20	±1.70	1.68	±2.37
Cold shock protein	0.66	±0.05	0.97	±0.31								
Rubrerythrin	1.35	±0.17	1.98	±0.84								
Energy metabolism/Compounds biosynthesis												
ATP synthase	3.43	±0.23	2.97	±0.01	** 18.23	±6.28	** 11.68	±5.59	* 7.75	±4.15	* 16.55	±6.96
Aconitate hydratase B	0.12	±0.16			1.74	±2.46						
Cysteine synthase									* 0.31	±0.44	* 1.39	±0.87
Fructose-1,6-bisphosphatase							1.72	±2.42				
Glutamine synthetase	2.56	±0.21	2.17	±0.44	** 1.44	±1.21	** 1.27	±1.22	*2.14	±0.36	* 1.44	±0.14
Glyceraldehyde-3-phosphate dehydrogenase	* 0.20	±0.04	* 0.09	±0.00	3.77	±3.70	4.35	±4.32	*2.46	±0.07	* 1.78	±0.32
Isocitrate dehydrogenase [NADP]					0.87	±1.23	2.49	±2.30				
Molybdopterin molybdenumtransferase					1.73	±0.80	0.41	±0.58				
Phycoerythrin									* 9.30	±4.63	* 0.33	±0.47
Allophycocyanin									* 1.54	±0.48	* 0	
Carbon dioxide-concentrating mechanism protein CcmK	0.04	±0.05	0.05	±0.07					0.94	±1.33	0.34	±0.47
Formate dehydrogenase	0.72	±0.40	0.74	±0.91								
Glucose-1-phosphate adenylyltransferase									0.31	±0.44	0.72	±0.07
Replication/Transcription/Translation												
30S ribosomal protein	2.91	±0.39	2.66	±0.81					0.31	±0.44	1.11	±0.62
50S ribosomal protein	13.47	±1.37	12.86	±4.44	2.32	±1.65	10.40	±7.14	2.52	±3.55	7.23	±0.73
DNA-binding protein HU	7.16	±0.09	7.69	±0.63								
DNA-directed RNA polymerase	0.54	±0.10	0.76	±0.23	7.77	±8.51	9.52	±4.73	0.31	±0.44	0.67	±0.94
Elongation factor	5.62	±0.47	6.77	±2.98	6.94	±0.05	9.40	±3.58	7.36	±1.51	17.43	±3.82
Histone-like protein	0.16	±0.10	0.15	±0.07	* 11.25	±5.22	* 3.80	±1.89				
Glycine-tRNA ligase					1.44	±1.21	1.25	±0.55				
Ribosomal protein S12 methylthiotransferase RimO							1.07	±1.51				
Transport												
Amino-acid ABC transporter-binding protein	5.61	±0.74	6.31	±0.36								
Fructose import binding protein FrcB	1.13	±0.50	1.26	±0.60								
Phosphate-binding protein	0.28	±0.15	0.29	±0.28					44.56	±23.88	23.09	±15.58
Cell motility, structure, and division												
Cell division protein FtsZ									0.94	±1.33	0.78	±1.09
Actin-like protein	* 0.70	±0.23	* 0.90	±0.15								
Tubulin					3.77	±2.07	2.57	±3.04				
Peptidoglycan-associated lipoprotein	0.37	±0.02	0.75	±0.63								
Flagellin	4.26	±0.41	5.37	±0.40	4.91	±1.18	3.58	±0.40				
Other (<1%)	7.60		5.72		4.03		2.30		4.03		1.40	



Energy metabolism/compounds biosynthesis	Day	SD	Night	SD
Allophycocyanin	* 1.54	± 0.49	* 0	
ATP synthase	* 7.73	± 4.19	* 16.55	± 6.96
Carbon dioxide-concentrating mechanism protein CcmK	0.94	± 1.33	0.34	± 0.47
Cysteine synthase	* 0.31	± 0.44	* 1.39	± 0.88
D-fructose 1,6-bisphosphatase	0.31	± 0.44	0.34	± 0.47
Enolase	0.6	± 0.84	0	
Glucose-1-phosphate adenylyltransferase	0.91	± 0.40	0.72	± 0.07
Glutamine synthetase	* 2.13	± 0.35	* 1.45	± 0.15
Glyceraldehyde-3-phosphate dehydrogenase	* 2.45	± 0.10	* 1.78	± 0.33
Glyoxalase II	0.31	± 0.44	0	
Phycobiliprotein	0.31	± 0.44	0	
Phycocyanine	0.63	± 0.89	0	
Phycoerythrin	* 9.27	± 4.69	* 0.34	± 0.47
Phosphopentose epinerase	0		0.34	± 0.47
Transketolase	0.31	± 0.44	0	

Translation/Transcription	Day	SD	Night	SD
30S ribosomal protein	0.31	± 0.44	1.11	± 0.62
50S ribosomal protein	2.52	± 3.55	7.23	± 0.73
DNA-directed RNA polymerase subunit	0.31	± 0.44	0.67	± 0.94
Elongation factor Tu	7.31	± 1.44	17.43	± 3.82
RNA-binding protein	0.63	± 0.88	0	
Single-stranded DNA-binding protein	** 0.00		** 0.72	± 0.07

Transport	Day	SD	Night	SD
Leu/Ile/Val-binding protein	0.31	± 0.44	0	
Membrane transporter protein	0.60	± 0.84	0	
Phosphate-binding protein	* 44.19	± 23.36	* 23.09	± 15.57

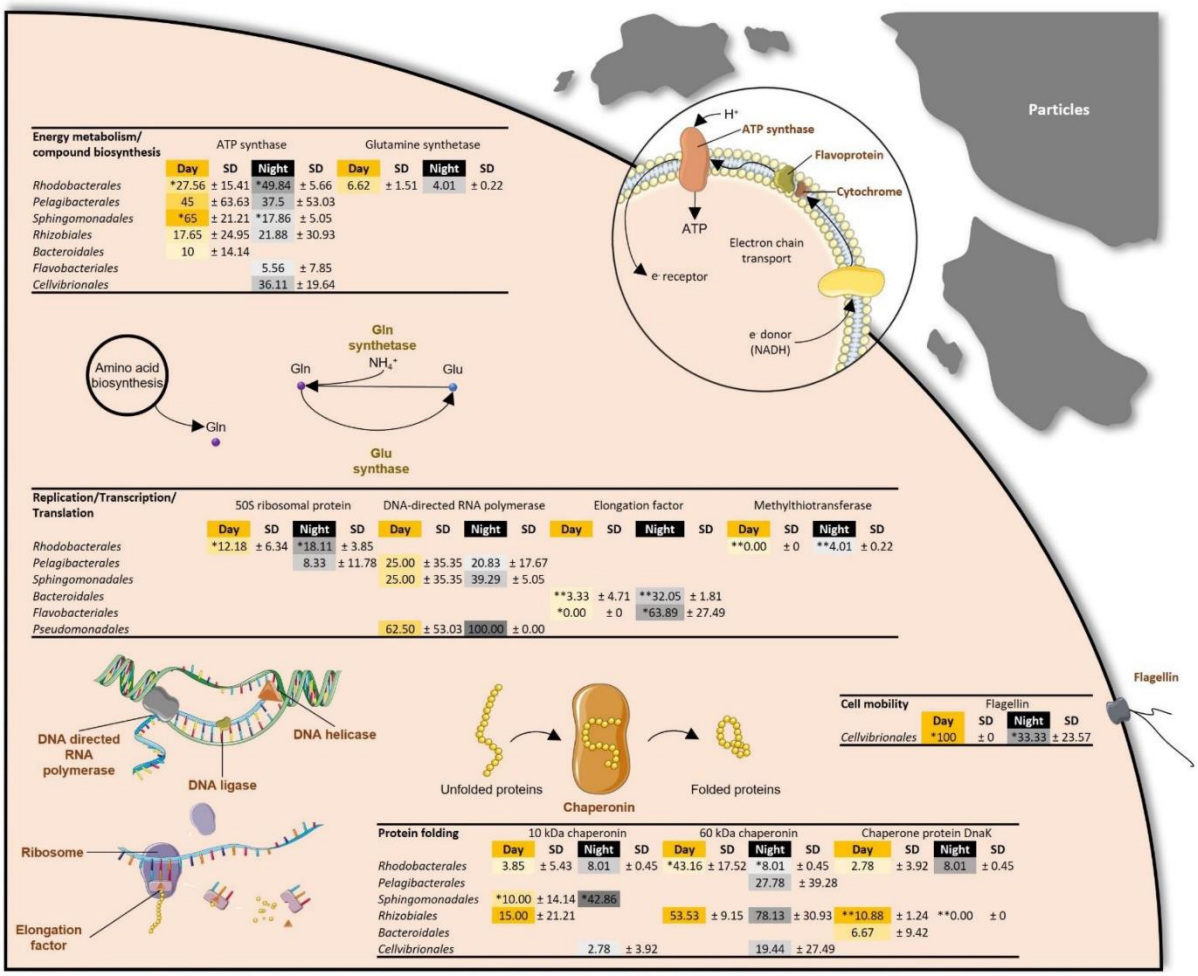
Protein folding	Day	SD	Night	SD
10 kDa chaperonin	0.94	± 1.33	5.91	± 3.97
60 kDa chaperonin	12.97	± 6.55	17.02	± 2.24
ATP-dependent Clp protease proteolytic subunit	0.00		1.11	± 0.62
Chaperone protein DnaK	1.19	± 1.68	1.68	± 2.37

Cell division	Day	SD	Night	SD
Cell division protein FtsZ	0.94	± 1.33	0.78	± 1.09

Figure 4.2: Representation of the diel cycle impact on the cellular processes of *Synechococcales*. Values consisted of the average percentage of total unique peptide spectra detected per protein function detected during day (yellow, n = 2) and night (black, n = 2) in all *Synechococcales* characterized in the 0.2 and 0.8 µm fractions. Significative differences between day and night samples are shown with a * (p value ≤ 0.1) or ** (p value ≤ 0.05) and were calculated with a paired t-test.

Proteins involved in stress response processes were exclusively characterized in a few free-living bacterial orders, including *Rhodobacterales*, *Pelagibacterales*, *Flavobacteriales* and *Cellvibrionales* (**Table 4.2 and Figure 4.4**). Interestingly, the catalase-peroxidase and superoxide dismutase [Fe], which both take part to the oxidative stress response, were exclusively detected during the day in *Rhodobacterales* (**Figure 4.4**). In contrast, the thioredoxin was only observed during the day in *Flavobacteriales* and during the night in

Cellvibrionales (**Figure 4.4**), while the rubrerythrin and the cold-shock protein were expressed by *Pelagibacterales* during both the day and the night (**Figure 4.4**).



Energy metabolism/ compound biosynthesis

	ATP synthase				Glutamine synthetase			
	Day	SD	Night	SD	Day	SD	Night	SD
Rhodobacterales	*27.56	± 15.41	*49.84	± 5.66	6.62	± 1.51	4.01	± 0.22
Pelagibacterales	45	± 63.63	37.5	± 53.03				
Sphingomonadales	*65	± 21.21	*17.86	± 5.05				
Rhizobiales	17.65	± 24.95	21.88	± 30.93				
Bacteroidales	10	± 14.14						
Flavobacteriales			5.56	± 7.85				
Cellvibrionales			36.11	± 19.64				

Replication/Transcription/ Translation

	50S ribosomal protein				DNA-directed RNA polymerase				Elongation factor				Methylthiotransferase			
	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD
Rhodobacterales	*12.18	± 6.34	*18.11	± 3.85									**0.00	± 0	**4.01	± 0.22
Pelagibacterales			8.33	± 11.78	25.00	± 35.35	20.83	± 17.67								
Sphingomonadales					25.00	± 35.35	39.29	± 5.05								
Bacteroidales									**3.33	± 4.71	**32.05	± 1.81				
Flavobacteriales									*0.00	± 0	*63.89	± 27.49				
Pseudomonadales					62.50	± 53.03	100.00	± 0.00								

Cell mobility

	Flagellin			
	Day	SD	Night	SD
Cellvibrionales	*100	± 0	*33.33	± 23.57

Protein folding

	10 kDa chaperonin				60 kDa chaperonin				Chaperone protein DnaK			
	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD
Rhodobacterales	3.85	± 5.43	8.01	± 0.45	*43.16	± 17.52	*8.01	± 0.45	2.78	± 3.92	8.01	± 0.45
Pelagibacterales							27.78	± 39.28				
Sphingomonadales	*10.00	± 14.14	*42.86									
Rhizobiales	15.00	± 21.21			53.53	± 9.15	78.13	± 30.93	**10.88	± 1.24	**0.00	± 0
Bacteroidales									6.67	± 9.42		
Cellvibrionales			2.78	± 3.92			19.44	± 27.49				

Figure 4.3: Representation of the diel cycle impact on the cellular processes of particle-attached bacteria. Values consisted of the average percentage of total unique peptide spectra detected per protein function detected during day (yellow, n = 2) and night (black, n = 2) in all particle-attached bacteria characterized in the 0.8 µm fractions. Significant differences between day and night samples are shown with a * (p value ≤ 0.1) or ** (p value ≤ 0.05) and were calculated with a paired t-test.

3.4. Replication, transcription, and translation

Replication, transcription, and translation processes, mainly represented by the DNA-directed RNA polymerase, the 30S and 50S ribosomal proteins and the elongation factor, were characterized in both day and night metaproteomes (**Table 4.2**). These biological processes were equally represented during the day and at night in *Synechococcales* (**Figure 4.2**).

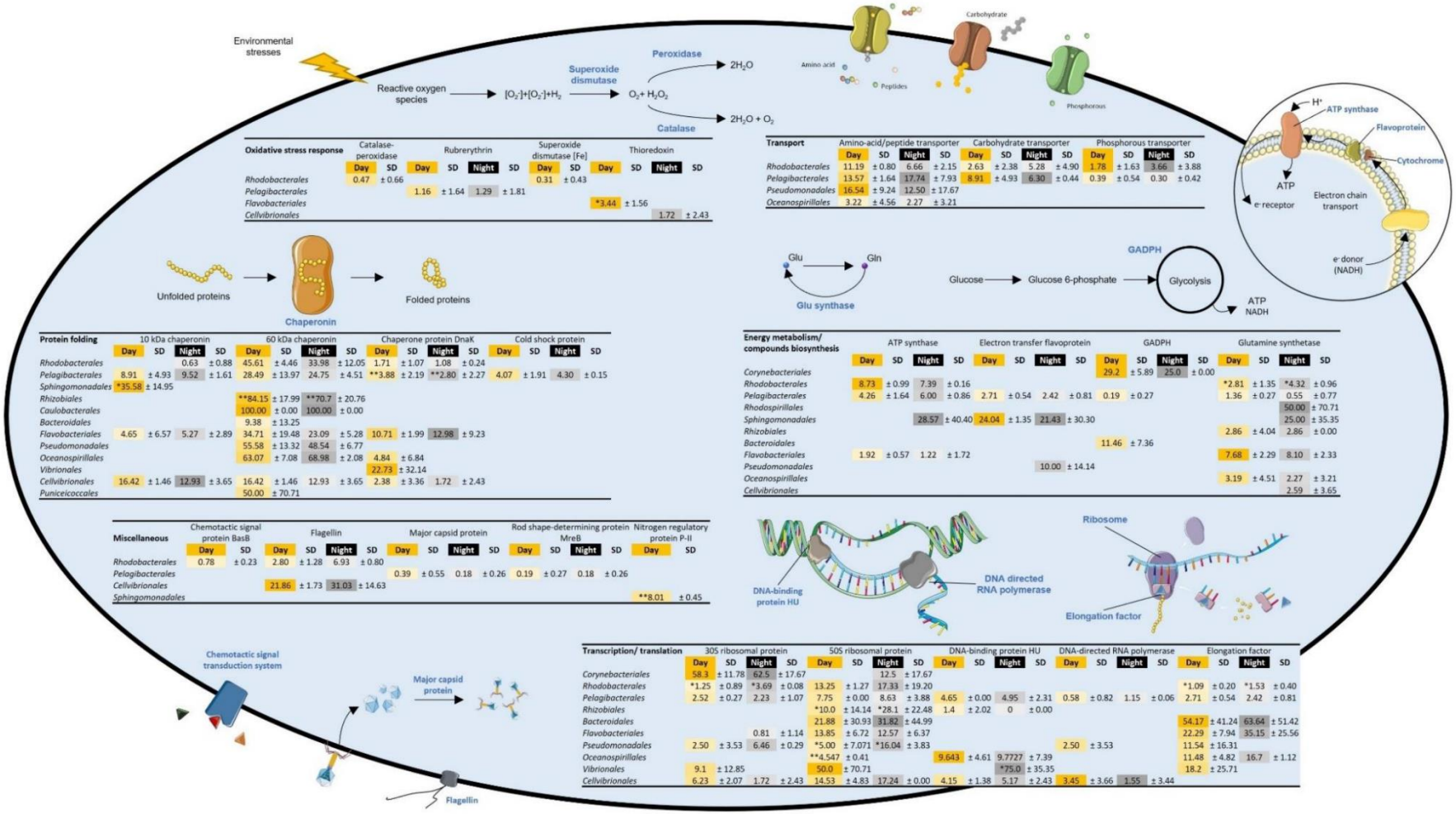
Within the attached bacterial fraction, translational proteins (i.e., 50S ribosomal protein and elongation factor) were over-represented at night in *Rhodobacterales*, *Bacteroidales* or *Flavobacteriales*, while the abundance of the DNA-directed RNA polymerase was similar during both day and night (**Figure 4.3**). The 30S and 50S ribosomal proteins, the elongation factor, the DNA-binding protein HU and the DNA-directed RNA polymerase were also characterized in numerous free-living bacterial orders (**Figure 4.4**). These proteins were not impacted by the diel cycle except in *Rhodobacterales*, *Rhizobiales* and *Pseudomonadales*, where the translation process seemed more important during the night (**Figure 4.4**).

1.1. Energy metabolism and compounds biosynthesis

Energy metabolism and compound biosynthesis accounted for the most diverse functional category and was particularly represented in the particle-attached bacteria and in *Cyanobacteria* metaproteomes (**Table 4.2**). The ATP synthase was the dominant protein and showed diel variation in both the particle-attached bacterial community (p value = 0.02) and in *Cyanobacteria* (p value = 0.07), while it was stable in the free-living bacterial fraction (**Table 4.2).** Other proteins related to the energy metabolism and involved in the pentose phosphate pathway, the glycolysis or the pyruvate metabolism were found to be regulated over time in *Cyanobacteria* (**Table 4.2**). Photosynthesis proteins such as the phycoerythrin and the allophycocyanin were clearly synchronized with daytime in *Synechococcales* (p value = 0.10 and 0.07, respectively) (**Figure 4.2**).

1.2. Transport and cell division, structure, and mobility

Interestingly, the amino acid biosynthesis pathway was represented in all metaproteomes by the glutamine synthase, involved in glutamine metabolism via the incorporation of ammonium ion into glutamate (Forchhammer, 2007) (**Table 4.2**).



Oxidative stress response

	Catalase-peroxidase		Rubrerythrin				Superoxide dismutase [Fe]		Thioredoxin			
	Day	SD	Day	SD	Night	SD	Day	SD	Day	SD	Night	SD
Rhodobacterales	0.47	± 0.66					0.31	± 0.43				
Pelagibacterales			1.16	± 1.64	1.29	± 1.81						
Flavobacteriales									*3.44	± 1.56		
Cellvibrionales											1.72	± 2.43

Transport

	Amino-acid/peptide transporter				Carbohydrate transporter				Phosphorous transporter			
	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD
Rhodobacterales	11.19	± 0.80	6.66	± 2.15	2.63	± 2.38	5.28	± 4.90	1.78	± 1.63	3.66	± 3.88
Pelagibacterales	13.57	± 1.64	17.74	± 7.93	8.91	± 4.93	6.30	± 0.44	0.39	± 0.54	0.30	± 0.42
Pseudomonadales	16.54	± 9.24	12.50	± 17.67								
Oceanospirillales	3.22	± 4.56	2.27	± 3.21								

Protein folding

	10 kDa chaperonin				60 kDa chaperonin				Chaperone protein DnaK				Cold shock protein			
	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD
Rhodobacterales			0.63	± 0.88	45.61	± 4.46	33.98	± 12.05	1.71	± 1.07	1.08	± 0.24				
Pelagibacterales	8.91	± 4.93	9.52	± 1.61	28.49	± 13.97	24.75	± 4.51	**3.88	± 2.19	**2.80	± 2.27	4.07	± 1.91	4.30	± 0.15
Sphingomonadales	*35.58	± 14.95														
Rhizobiales					**84.15	± 17.99	**70.7	± 20.76								
Caulobacterales					100.00	± 0.00	100.00	± 0.00								
Bacteroidales					9.38	± 13.25										
Flavobacteriales	4.65	± 6.57	5.27	± 2.89	34.71	± 19.48	23.09	± 5.28	10.71	± 1.99	12.98	± 9.23				
Pseudomonadales					55.58	± 13.32	48.54	± 6.77								
Oceanospirillales					63.07	± 7.08	68.98	± 2.08	4.84	± 6.84						
Vibrionales									22.73	± 32.14						
Cellvibrionales	16.42	± 1.46	12.93	± 3.65	16.42	± 1.46	12.93	± 3.65	2.38	± 3.36	1.72	± 2.43				
Puniceicoccales					50.00	± 70.71										

Energy metabolism/ compounds biosynthesis

	ATP synthase				Electron transfer flavoprotein				GADPH				Glutamine synthetase			
	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD
Corynebacteriales									29.2	± 5.89	25.0	± 0.00				
Rhodobacterales	8.73	± 0.99	7.39	± 0.16									*2.81	± 1.35	*4.32	± 0.96
Pelagibacterales	4.26	± 1.64	6.00	± 0.86	2.71	± 0.54	2.42	± 0.81	0.19	± 0.27			1.36	± 0.27	0.55	± 0.77
Rhodospirillales															50.00	± 70.71
Sphingomonadales			28.57	± 40.40	24.04	± 1.35	21.43	± 30.30							25.00	± 35.35
Rhizobiales													2.86	± 4.04	2.86	± 0.00
Bacteroidales									11.46	± 7.36						
Flavobacteriales	1.92	± 0.57	1.22	± 1.72									7.68	± 2.29	8.10	± 2.33
Pseudomonadales							10.00	± 14.14								
Oceanospirillales													3.19	± 4.51	2.27	± 3.21
Cellvibrionales															2.59	± 3.65

Miscellaneous

	Chemotactic signal protein BasB				Flagellin				Major capsid protein				Rod shape-determining protein MreB				Nitrogen regulatory protein P-II	
	Day	SD	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD
Rhodobacterales	0.78	± 0.23	2.80	± 1.28	6.93	± 0.80										
Pelagibacterales							0.39	± 0.55	0.18	± 0.26	0.19	± 0.27	0.18	± 0.26		
Cellvibrionales			21.86	± 1.73	31.03	± 14.63										
Sphingomonadales															**8.01	± 0.45

Transcription/ translation

	30S ribosomal protein				50S ribosomal protein				DNA-binding protein HU				DNA-directed RNA polymerase				Elongation factor			
	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD	Day	SD	Night	SD
Corynebacteriales	58.3	± 11.78	62.5	± 17.67			12.5	± 17.67												
Rhodobacterales	*1.25	± 0.89	*3.69	± 0.08	13.25	± 1.27	17.33	± 19.20									*1.09	± 0.20	*1.53	± 0.40
Pelagibacterales	2.52	± 0.27	2.23	± 1.07	7.75	± 0.00	8.63	± 3.88	4.65	± 0.00	4.95	± 2.31	0.58	± 0.82	1.15	± 0.06	2.71	± 0.54	2.42	± 0.81
Rhizobiales					*10.0	± 14.14	*28.1	± 22.48	1.4	± 2.02	0	± 0.00								
Bacteroidales					21.88	± 30.93	31.82	± 44.99									54.17	± 41.24	63.64	± 51.42
Flavobacteriales			0.81	± 1.14	13.85	± 6.72	12.57	± 6.37									22.29	± 7.94	35.15	± 25.56
Pseudomonadales	2.50	± 3.53	6.46	± 0.29	*5.00	± 7.071	*16.04	± 3.83					2.50	± 3.53			11.54	± 16.31		
Oceanospirillales					**4.547	± 0.41			9.643	± 4.61	9.7727	± 7.39					11.48	± 4.82	16.7	± 1.12
Vibrionales	9.1	± 12.85			50.0	± 70.71					*75.0	± 35.35					18.2	± 25.71		
Cellvibrionales	6.23	± 2.07	1.72	± 2.43	14.53	± 4.83	17.24	± 0.00	4.15	± 1.38	5.17	± 2.43	3.45	± 3.66	1.55	± 3.44				

Figure 4.4: Representation of the diel cycle impact on the cellular processes of free-living bacteria. Values consisted of the average percentage of total unique peptide spectra detected per protein function detected during day (yellow, n = 2) and night (black, n = 2) in all free-living bacteria characterized in the 0.2 µm fractions. Protein functions displaying only day or night values (e.g., catalase-peroxidase) means that they were exclusively detected during the day or the night period. Significant differences between day and night samples are shown with a * (p value ≤ 0.1) or ** (p value ≤ 0.05) and were calculated with a paired t-test.

This protein displayed diel variations in *Synechococcales* (p value = 0.07) and in free-living *Rhodobacterales* (p value = 0.06), where it was dominant at day and night, respectively (**Figures 4.2 and 4.4**). In *Synechococcales*, the amino acid biosynthesis pathway was also characterized by the cysteine synthase, which showed contrasting diel variability with the glutamine synthase as it peaked at night (p value = 0.09) (**Figure 4.2**).

Numerous amino acid/peptides, carbohydrate and phosphorous transporters were observed in free-living bacterial metaproteomes (**Figure 4.4**). The phosphorous transporters were the major proteins identified in *Cyanobacteria* (**Figure 4.2**). In contrast, no transport related protein was detected in the particle-attached fraction. While the phosphorous transporters in *Synechococcales* were over expressed during the day (p value = 0.08) (**Figure 4.2**), no diel variation was observed in transporter abundance of free-living bacteria (**Figure 4.4**).

Among the last protein functions observed, the cell division protein (FtsZ) was present in *Synechococcales* metaproteomes, and its expression was not affected by the diel cycle (**Figure 4.2**). Flagellin protein, expressed by *Cellvibrionales*, showed contrasting diel regulation depending on the bacterial lifestyle as it was stable over time in the free-living fraction and peaked during the day in the particle-attached fraction (p value = 0.08) (**Figures 4.3 and 4.4**). This protein was also detected in free-living *Rhodobacterales* where no diel variation was observed (**Figure 4.4**). However, proteins associated with the chemotaxis system in free-living *Rhodobacterales* were found to express at day only (p value = 0.07) (**Figure 4.4**). Finally, the major capsid protein and the rod shape-determining protein MreB were characterized in free-living *Pelagibacterales* during both the day and the night (**Figure 4.4**).

2. Discussion

This study compared day and night protein abundance between free-living and particle-attached bacteria from an oligotrophic marine surface environment. Metaproteomic analyses were performed on duplicate filters for both conditions (day/night) and pore-sized fractions (0.2/0.8 µm). Although we appreciate that more replicates could be performed, the low standard deviation of our samples allowed us to provide the first overview of protein diel variations at the sea surface. The total identified proteins of both bacterial fractions were consistent with previous metaproteomic studies conducted in marine oligotrophic surface waters (Morris et al., 2002; Sowell et al., 2009; Williams et al., 2012; Williams et al., 2013) (**Table 4.1**). The taxonomic distribution of the OSD14 metagenome showed that the most abundant members of the community were *Proteobacteria*, followed by *Bacteroidetes* and *Cyanobacteria* (**Table 4.1**). These taxa were previously reported as numerically abundant in coastal marine oligotrophic environments, such as the Mediterranean Sea (Feingersch et al., 2010), the Antarctic (Williams et al., 2012; Williams et al., 2013) and Atlantic (Georges et al., 2014; Morris et al., 2010) surface waters. The metaproteome taxonomic structure was found to be similar to that of OSD14 metagenome (**Table 4.1**), indicating a correlation between abundant vs. metabolically active community members. Moreover, free-living and particle-attached metaproteomes showed high similarities in taxonomic distribution. Overlaps in the structure of both bacterial fractions were previously reported within microbial assemblages of the Mediterranean Sea, where the colonization of particles was suggested to be largely mediated by free-living bacteria present in the surrounding water (Ghiglione et al., 2007; Hollibaugh et al., 2000; Moeseneder et al., 2001).

Interestingly, the representation of *Cyanobacteria* was higher in the metaproteomes than in the OSD14 metagenome (**Table 4.1**). All identified *Cyanobacteria* proteins were classified

as *Synechococcales* (**Figure 4.1**), which is a main contributor to the primary production in oligotrophic water during summer (Mella-Flores et al., 2011). In contrast, *Bacteroidetes*, mainly characterized as *Flavobacteriia*, were less represented in the metaproteomes (**Table 4.1**). *Flavobacteriia* are bloom-associated bacteria known for degrading phytoplankton-derived compounds (Buchan et al., 2014). The NW Mediterranean Sea is characterized by spring and autumnal phytoplankton blooms separated by an oligotrophic summer (Mayot et al., 2017), which could explain why this group was less abundant at the protein level at the sampling time.

Comparison of day and night metaproteomes revealed differences between free-living and particle-attached bacteria (**Figure 4.1**). Proteins of free-living bacteria were the most represented in all samples and peaked during the day, while proteins expressed by particle-attached bacteria showed higher abundance during the night (**Table 4.1**). Similar observations were reported for the bacterial activity measured by 3H-leucine incorporation in the NW Mediterranean Sea in summer (Ghiglione et al., 2007). The activity of attached bacteria depends on the nature and concentration of aggregates and suspended particles, which represent hot-spots for microbial processes (Simon et al., 2002). In the upper layer of the NW Mediterranean Sea during summer, the release of organic material from photosynthetic microorganisms and zooplankton was suggested to be a major factor driving the diel variation in particle-attached bacterial activity (Ghiglione et al., 2007). In this study, metaproteomic analyses showed that particle-attached *Flavobacteriales* proteins were more abundant at night (**Figure 4.1**). Therefore, the representation of *Flavobacteriales* could increase at night in response of zooplankton feeding on phytoplankton and releasing organic matter (Ghiglione et al., 2007).

Time-keeping mechanisms in *Synechococcales* are well described and show that circadian clock regulates patterns of genetic expression throughout the day using external variable clues (e.g., light, temperature and/or redox cycles) to scale to the environment (Cohen & Golden, 2015). Distinct diel profiles of protein abundance were observed between

metaproteomes in *Synechococcales* (**Figure 4.1**). Proteins involved in photosynthesis and phosphate transport showed a clear trend in abundance that was higher in the day than at night (**Table 4.2 and Figure 4.2**). These results confirmed previous comparative day/night (meta)-transcriptomic studies that showed a higher abundance of transcripts for photosynthesis during the day compared with the night (Ottesen et al., 2014; Poretsky et al., 2009). In contrast, proteins involved in housekeeping functions such as protein folding, translation, transcription, and cell division displayed similar abundances in both day and night samples (**Table 4.2 and Figure 4.2**). Proteins involved in catabolic pathways including glycolysis, pyruvate metabolism and respiration were also observed in similar abundance in the day and at night, except for the ATP synthase, which was twice more represented during the night (**Figure 4.2**). This suggested that *Synechococcales* maintained housekeeping activity independently of diel variation but increased ATP production during the night when photosynthesis is shut down.

Chaperonin proteins, which are characterized as ubiquitous in many marine ecosystems (Sowell et al., 2011; Sowell et al., 2009; Williams et al., 2012), were highly represented in both day and night community metaproteomes (**Table 4.2**). During summer, bacteria in the euphotic layer are exposed to high UV radiation, altering both proteins and DNA structure. Mechanisms such as protein folding, reactive oxygen species reduction and protein biosynthesis are essential for coping with protein damage and maintaining proper cellular functions (Matallana-Surget et al., 2013). Chaperonin abundance was not impacted by diel cycle in most bacterial orders, with a few exceptions including *Rhodobacterales*, *Sphingomonadales*, *Rhizobiales* and *Pelagibacterales* (**Figures 4.2 and 4.3**). Proteins involved in protein biosynthesis (i.e., ribosomal protein and elongation factor) were more abundant at night in free-living and particle-attached *Rhodobacterales* (**Figures 4.2 and 4.3**). Interestingly, proteins involved in oxidative stress response such as the catalase-peroxidase and the superoxide dismutase [Fe] were only observed during the day in free-living *Rhodobacterales* (**Figure 4.4**). Thioredoxin

was exclusively detected in the day in *Flavobacteriales* and at night in *Cellvibrionales* (**Figure 4.4**). In contrast, rubrerythrin and cold shock protein were present in *Pelagibacterales* during both the day and night (**Figure 4.4**). These observations suggested that protein regulation in response to environmental stress is taxa-specific and depends on lifestyle (free-living vs. particle-attached). Protein regulation in a protein repair system could be time-gated in bacterial orders such as *Rhodobacterales* or constitutive in other such as *Pelagibacterales*.

Proteins involved in compounds transport were detected in free-living bacteria and *Cyanobacteria* during both the day and the night (**Table 4.2**). Glutamine synthetase, involved in nitrogen metabolism, was detected in all bacterial fractions (**Table 4.2, Figures 4.2 and 4.4**). The abundance of transporters in the free-living fraction and the overall characterization of glutamine synthetase suggested an adaptation to an oligotrophic environment, where a strong competition for limiting nutrients such as nitrogen and phosphorous was reported (Hoch et al., 2006; Kuipers et al., 2000; Williams & Cavicchioli, 2014). Interestingly, no transporter was identified in the attached-bacterial fraction (**Table 4.2 and Figure 4.3**). This suggested lesser environmental pressure for the expression of nutrient transporters in attached bacteria since nutrients are more readily available in the particle microenvironment. In contrast, free-living bacteria and *Cyanobacteria* could depend on constitutive expression of transporters for efficient nutrient scavenging (Williams & Cavicchioli, 2014).

Pelagibacterales dominated the free-living bacterial community, in both the metagenome and the metaproteomes (**Figure 4.1**). *Pelagibacterales* include proteorhodopsin-containing photoheterotrophs such as Pelagibacter (SAR11), which is known to be abundant and highly active in the ocean (Morris et al., 2002). Metatranscriptomic studies showed evidence of diel periodicity in many of their gene transcripts (Ottesen et al., 2014). Here, proteins involved in protein folding, stress response and replication, transcription and translation were the main functions characterized in *Pelagibacterales* metaproteomes (**Figures**

4.3 and 4.4). Despite the higher trend in protein abundance of free-living *Pelagibacterales* at night (**Figure 4.1**), no significant change was observed in the aforementioned biological processes (**Figure 4.1**). This suggests that *Pelagibacterales* constitutively express diverse housekeeping genes required for the maintenance of basal cellular functions that are essential to protect the cell against molecular damage and environmental changes. Protein expression regulation could take place at transcript level, thus limiting energy losses from diel protein turnover (Waldbauer et al., 2012).

3. Conclusions

This study provided the first overview on the picoplanktonic response to diel variation at the protein level and demonstrated taxa-specific diel protein regulation from surface marine microbial communities. Taxonomic overlaps were observed between free-living and particle-attached bacteria, where protein abundance peaked at day and at night, respectively. The photoautotrophs *Synechococcales* showed distinct diel protein profiles with light-dependent functions synchronized with daytime. Similarly, diel variations in (photo)-heterotrophic bacteria were observed, thus revealing distinct adaptation strategies with essential regulations in environmental stress response. This study provided preliminary results reinforcing the hypothesis that the functioning of free-living and particle-attached communities could be time-gated. Additional work, including observational studies with more sampling replicates and laboratory-based investigations, is needed to further understand the response of these communities to diel changes and to decipher the cellular mechanisms involved in the diel adaptation of (photo)-heterotroph microorganisms.

Chapter 5: Molecular clocks in Prokaryotes

Towards the Discovery of Novel Molecular Clocks in Prokaryotes

Géron, **A**., Werner, J., Wattiez, R., & Matallana-Surget, S.
Critical Review in Microbiology (under review)

Abstract

Diel cycle is of enormous biological importance as it imposes daily oscillation in environmental conditions, which temporally structures most ecosystems. Organisms developed biological time-keeping mechanisms – circadian clocks – that provide a significant fitness advantage over competitors by optimizing the synchronization of their biological activities. While circadian clocks are ubiquitous in Eukaryotes, they are so far only characterized in *Cyanobacteria* within Prokaryotes. However, growing evidence suggests that circadian clocks are widespread in the bacterial and archaeal domains. As Prokaryotes are at the heart of crucial environmental processes and are essential to human health, unraveling their time-keeping systems provides numerous applications in medical research, environmental sciences, and biotechnology. In this review, we elaborate on how novel circadian clocks in Prokaryotes offer research and development perspectives. We compare and contrast the different circadian systems in *Cyanobacteria* and discuss about their evolution and taxonomic distribution. We necessarily provide an updated phylogenetic analysis of bacterial and archaeal species that harbor homologs of the main cyanobacterial clock components. Finally, we elaborate on new potential clock-controlled microorganisms that represent opportunities of ecological and industrial relevance in prokaryotic groups such as anoxygenic photosynthetic bacteria, methanogenic archaea, methanotrophs or sulfate-reducing bacteria.

1. Introduction

The Earth's rotation around its axis causes the 24-hour day and night cycle known as the diel cycle. This recurrent phenomenon induces daily fluctuations of solar radiation and/or temperature, which regulate the functioning of most ecosystems. One of the most significant adaptations to the diel cycle is the emergence of circadian clocks in both Eukaryotes and Prokaryotes (Saini et al., 2019). Circadian clocks are endogenous molecular systems that entrain to environmental zeitgeber (i.e., external cue such as light) and allow temporal coordination of intracellular processes. These clocks follow a 24-hour period oscillation that remains in a constant condition (free-running) and over a range of physiological temperatures (temperature-compensated) (Golden & Canales, 2003). Although circadian clocks are ubiquitous in Eukaryotes, prokaryotic circadian clocks were so far only characterized in *Cyanobacteria* (for a review see Johnson & Rust, 2021). The cyanobacterial clock is based on the KaiABC proteins complex that acts as an oscillator that synchronizes with environmental signals and subsequently controls rhythmic gene expression (Kondo, 2007; Markson et al., 2013).

Unravelling the time-keeping mechanisms of non-cyanobacterial species is crucial regarding the central role Prokaryotes play in all ecosystems. This fundamental knowledge could lead to the incorporation of their temporal structures in broad technological applications, which will undoubtedly offer great research and development opportunities. Over the past decades, a growing body of evidence suggested that time-controlled mechanisms are widespread in Prokaryotes (Schmelling et al., 2021). Pioneer bioinformatical analyses revealed that homologs of the Kai proteins were present across a broad range of bacterial and archaeal phyla (Dvornyk et al., 2003; Loza-Correa et al., 2010; Schmelling et al., 2017). Microbial community-wide diel rhythms are observed in various biomes and are attributed to non-

cyanobacterial species (Ottesen et al., 2014; Hörnlein et al., 2020; Géron et al., 2021). Moreover, rhythmic activities are found in prokaryotic species harboring Kai protein homologs such as purple bacteria (Min et al., 2005; Van Praag et al., 2000; Ma et al., 2016), non-photosynthetic bacteria (Sartor et al., 2019; Paulose et al., 2019; Eelderink-Chen et al., 2021) and extremophile archaea (Whitehead et al., 2009; Maniscalco et al. 2014).

In this review, we discuss the numerous applications of biological rhythms of Prokaryotes in research and development from human health to the industry. We present the known circadian systems in Prokaryotes, and we describe their properties and evolutionary paths. With the development of next-generation sequencing the number of complete, draft and metagenome-assembled genomes exponentially increased since the last phylogenetic analyses of Kai protein homologs (Dvornyk et al., 2003; Loza-Correa et al., 2010; Schmelling et al., 2017). Therefore, we provide an updated analysis of the presence of Kai proteins among bacterial and archaeal species, and we elaborate on new potentially clock-controlled candidates in taxonomic groups such as anoxygenic photosynthetic bacteria or methanogenic archaea.

2. Research focus on Prokaryotic time-keeping systems: rationales and impacts

Prokaryotes are essential to humans for numerous reasons (**Figure 5.1**). As a matter of fact, microorganisms outnumber human cells by tenfold and are crucial in food degradation, nutrient absorption, or defense against pathogens (Venkova et al., 2018). Prokaryotes also play a significant role in symbiosis with plants by providing nutrients as well as growth hormones (Backer et al., 2018). They also degrade organic matter and contribute to biogeochemical cycles (Madsen, 2011). Moreover, prokaryotes are central in numerous biotechnological applications (Caplice & Fitzgerald, 1999; Demain & Sanchez, 2009; Akinsemolu, 2018). In that context, research focus on the chronobiology (i.e., the study of the temporal biology of organisms) of

prokaryotes in natural and artificial environments would provide important translational opportunities in the medical, ecological, and industrial fields, which could promote and maintain our health and well-being.

The human clock scales with environmental changes and affects the behaviour including timing and type of food intake (**Figure 5.1**) (Matenchuk et al., 2020). Consequently, the gut microbiome experiences daily oscillations leading to taxonomic, functional, and spatial variations inducing local and systemic effects on the host (Deaver et al., 2018; Wu et al., 2018; Godinho-Silva et al., 2019; Nobs et al., 2019; Saran et al., 2020). Although it is unknown if the gut microbiota possesses an endogenous circadian clock, there is growing evidence supporting the interconnection between the host's and the microbiota's circadian rhythms, their impact on host physiology, and their association with diseases such as diabetes (Thaiss et al., 2014; Kaczmarek et al., 2017; Heinemann et al., 2021; Choi et al., 2021).

Furthermore, circadian rhythms of host immunity are interlinked to bacterial and viral infections (Curtis et al., 2014; Pearson et al., 2021). With the prevalence of antibiotic-resistance bacteria threatening global health, food security, and development, chronotherapy (i.e., time-specific therapy) capitalizing on the circadian rhythms of both the host and the infectious agents could lead to a more efficient treatment strategy (Pearson et al., 2021). Similarly, plant's physiology is regulated by the circadian clock, which likely generates a rhythmic environment for the root-associated microorganisms (i.e., the rhizobacteria). The rhizobacteria feed on root exudates and, in exchange, provide great benefits such as plant growth promotion, nutrient accessibility (e.g., nitrogen fixation), or protection from predators and phytopathogens (**Figure 5.1**) (Gould et al., 2018; Hubbard et al., 2018; Lu et al., 2021). In that context, the root-colonizing bacteria could use circadian programs to coordinate their metabolism with rhythms of the host to maximize the advantage of the bidirectional interaction (Asif et al., 2019). Plant growth-promoting rhizobacteria are of high economic importance and are commercialized to

improve biomass production and quality of numerous crops (Saharan & Nehra, 2011). Therefore, deciphering the time-controlled interaction between rhizobacteria and plants could be a promising source of valuable discoveries that can have immediate applications in the sustainable agriculture field (Maddur Puttaswamy, 2019).

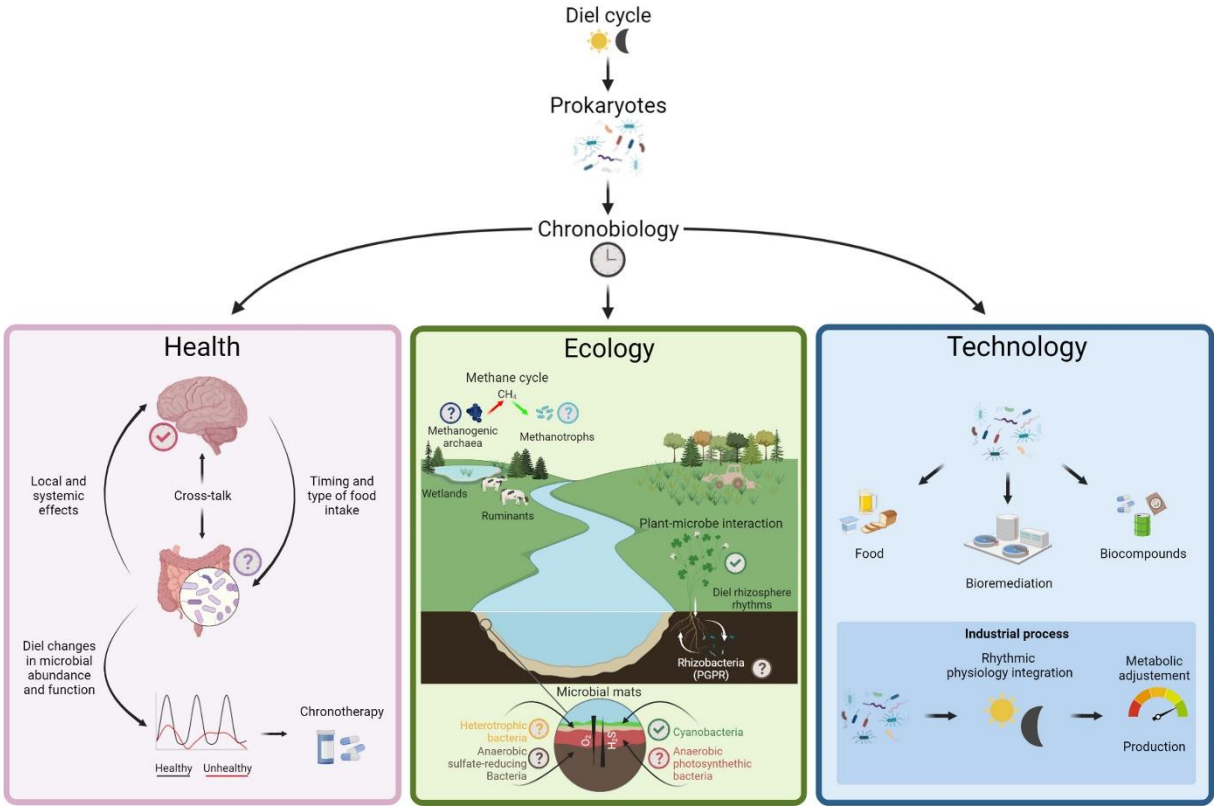


Figure 5.1: Schematic representation of chronobiology-based applications of Prokaryotes. The metabolism of humans and plants is controlled by a circadian clock that impacts their commensal microbial communities. In return, the microbiome fluctuates in terms of taxonomy, functioning and localization, which influences their host's physiology. A better understanding on potential time-keeping mechanisms in the microorganisms associated with plants and animals will allow more accurate models and open new perspectives on optimizing these relationships. In addition, the industrial and technological fields would potentially benefit from the integration of the natural rhythmic physiology of Prokaryotes into utilization and manufacturing processes.

Chronobiology research focusing on anoxygenic photosynthetic organisms (purple and green (non)-sulfur bacteria), sulfate-reducing bacteria, methanogenic Archaea, or methanotrophs also represent interesting applications for both fields of ecology and microbiology-based industries (**Figure 5.1**). Anoxygenic photosynthetic bacteria are often

encountered in microbial mats where they interact with *Cyanobacteria* that possess a well-defined circadian clock (Prieto-Barajas et al., 2018). As they can use light as an energy source, anoxygenic photosynthetic bacteria would greatly benefit from a time-keeping mechanism that would allow them to synchronize with a both biotic and abiotic external signals oscillating daily (Hörnlein et al., 2020). Because of microbial mat communities displaying a highly versatile metabolism that allows them to use a broad range of substrate, there is considerable interest in industrial uses of mats, especially for water treatment and for cleaning up pollution (Abed et al., 2020; George et al., 2020). Deciphering the response to diel changes in methanogenic archaeal, methanotrophic bacterial and sulfate reducing bacterial communities is crucial as they are deeply involved in methane and sulfur cycles. Methane is the second most important greenhouse gas on Earth and methanogenic archaea produce about 70% through the methanogenesis process (Conrad, 2009). Therefore, time-specific models of these microorganisms in their natural (e.g., microbial mats) or artificial environments (e.g., wastewater treatment plants) could provide insights to better understand their ecological functions and allow optimization of resource management (Wang et al., 2017; Enzmann et al. 2018). Finally, Prokaryotes are widely used in biocompounds production, bioremediation, or treatment of organic and industrial wastes (**Figure 5.1**) (Abdel-Aziz et al., 2017; Alizadeh-Sani et al., 2018; Kallscheuer et al., 2019). More insights into their rhythmic physiologies would allow the integration of their temporal programs into industrial processes, which could induce metabolic adjustments and lead to improved qualitative and quantitative production (Sartor et al., 2019).

3. The circadian systems in Prokaryotes: the clock versus the hourglass

Prior to mid-1980, Prokaryotes were thought to have neither the resource nor the need to possess circadian systems. As many bacteria can divide several times over a 24h-cycle, it

was believed that cellular functions would not be coupled to a biological clock. This dogma, known as the "circadian-infradian" rule, was refuted in 1986 when Huang and colleagues discovered circadian rhythmicity of nitrogen fixation and amino acid uptake in the cyanobacterium *Synechococcus* sp. RF-1 (Huang et al., 1990; Chen et al., 1991). Subsequently, *Synechococcus elongatus* PCC 7942 emerged as the model organism for studying the mechanism of the clock using luciferase reporter assay (Kondo et al., 1993). The clock in *Synechococcus elongatus* PCC 7942 is based on a core oscillator that consists of the KaiA, KaiB and KaiC proteins and determines diel patterns of gene expression, timing of cell division, or metabolic switches (**Figure 5.2**) (Dong et al., 2010; Mori et al., 1996; Pattanayek et al., 2014; Puszynska & O'Shea 2017).

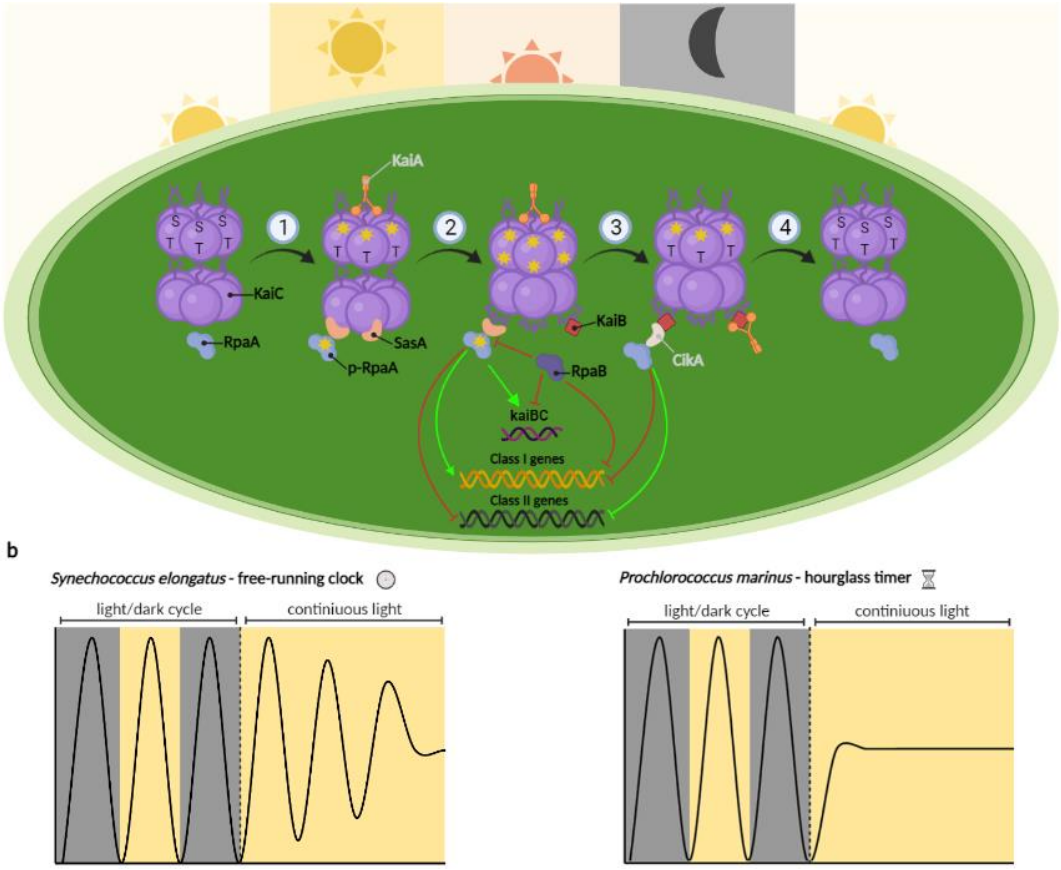


Figure 5.2: Circadian clocks in *Cyanobacteria*. (a): Schematic representation of the two circadian clock systems in Prokaryotes: *Synechococcus elongatus* PCC 7942 and *Prochlorococcus marinus* MED4 models. The phosphorylation/dephosphorylation cycle of KaiC is represented with the different key component of the output pathway that controls the gene expression. The proteins KaiA and CikA, labelled in grey, are missing from the *Prochlorococcus* clock model (Axman et al., 2014). **(b):** Difference between the stability of oscillation provided by a free-running clock in *S. elongatus* and an hourglass-like timer in *P. marinus* in the absence of light/dark cycle.

Throughout the course of day and night, KaiC, the central clock component, synchronizes with environmental signals and undergoes a phosphorylation and dephosphorylation cycle (**Figure 5.2**) (Nishiwaki et al., 2000; Nishiwaki et al., 2004; Nishiwaki et al., 2012). At dawn, unphosphorylated KaiC has a loose and unstacked structure, exposing A-Loops where KaiA can bind as the morning progresses (Kim et al., 2008). Throughout the day, KaiA stimulates KaiC autokinase activity and thus, threonine and serine become phosphorylated. By dusk, both threonine and serine residues of KaiC are fully phosphorylated and KaiC becomes stiff and stacked, which hides the binding site for KaiA and exposes a B-Loop that constitutes a binding site for KaiB. As the night progresses, KaiA can no longer bind to KaiC and get sequestered by KaiB, subsequently activating the KaiC autophosphatase (Kitayama et al., 2003; Chang et al., 2012; Tseng et al., 2014). The dephosphorylated threonine and serine residues induce KaiC to return to its unphosphorylated state with exposed A-loops (Nishiwaki et al., 2012). The Kai oscillator interacts with circadian output components that subsequently controls physiological processes, such as photosynthesis, glycogen metabolism, and cell division. In *S. elongatus*, these output components include the histidine kinase SasA, its cognate response regulator RpaA, and the phosphatase CikA (Iwasaki et al., 2000; TaKai et al., 2006). During the day, SasA binds to the phosphorylated KaiC protein, induces its autophosphorylation and transfers its phosphate group to RpaA, which induces phosphorylated RpaA to accumulate (Iwasaki et al., 2000; TaKai et al., 2006). At dusk, KaiB competes with SasA for binding with KaiC and, as the night progresses, KaiB sequesters KaiA and attracts CikA (Tseng et al., 2014). CikA binds to the KaiBC, thus forming a complex that acts as a phosphatase that dephosphorylates RpaA (Chang et al., 2015). By the end of the night, the RpaA phosphorylation level is very low. Therefore, the interaction of SasA and CikA with the Kai oscillator creates a phosphorylation cycle of RpaA that peaks at dusk (Markson et al., 2013). In its phosphorylated form, RpaA acts as a transcription factor that activates the transcription of

KaiBC genes and Class I genes and represses Class II genes (Markson et al., 2013). On the other hand, non-phosphorylated RpaA activates Class II genes and represses Class I genes. RpaB inhibits RpaA phosphorylation and represses KaiBC and Class I genes (Kappell & Waasbergen, 2007; Espinosa et al., 2015).

Interestingly, the KaiABC clock mechanism of *S. elongatus* PCC 7942 is not generalized in *Cyanobacteria*. Indeed, the globally distributed marine *Prochlorococcus* harbors a simplified clock mechanism that lacks KaiA and where KaiC compensates with an enhanced autophosphorylation activity (**Figure 5.2**) (Axmann et al., 2009). In addition, *Prochlorococcus marinus* MED4 is missing the output component CikA. The clock depends on the KaiBC complex formation, which allows SasA to bind with KaiB and induces RpaA phosphorylation (Axmann et al., 2009). Phosphorylated RpaA binds to the KaiBC promoter and regulates its transcription, which controls the clock expression. Although the reduced KaiBC-system in *Prochlorococcus* provides genome-wide expression rhythms (Zinser et al., 2009), it differs from the circadian clock of *S. elongatus* in that it cannot maintain oscillation in the absence of external signals (Holtzendorff et al., 2008). Therefore, the *Prochlorococcus* clock rather works as an hourglass, requiring daily signals to reset the cycle (Mullineaux & Stanewsky, 2009). *Prochlorococcus* could benefit from the reduced KaiBC-system in its habitat – the near-equatorial oceans – where a free-running clock may not be essential because of the high regularity of diel conditions (Axmann et al., 2009). In contrast, *Synechococcus elongatus* is found at higher longitude, where the seasonality and environment can often disturb the regularity of the signals which could make a clock system with stable rhythm indispensable. This hypothesis is reinforced by recent numerical simulations using a stochastic modeling approach that showed that an hourglass-like system outperforms a free-running clock in small organisms evolving in stable environment, which is the case of *Prochlorococcus* that is smaller by an order of magnitude than *Synechococcus* (Chew et al., 2018). The evolution of clock

systems in *Cyanobacteria* could expand our perspective away from the specific free-running clock system of *Synechococcus elongatus* and lead us to focus more on environmentally driven time-keeping systems in Prokaryotes (Chew et al., 2018).

4. Distribution of the Kai proteins

Computational analyses helped to unravel the origins and the evolutionary paths of the Kai proteins within Prokaryotes (Dvornyk et al., 2003; Loza-Correa et al., 2010; Schmelling et al., 2017). Using the Kai protein sequences from *Synechococcus elongatus* PCC 7942 as a template for sequence alignment-based discovery, bioinformatical analyses revealed that the Kai proteins are not restricted to *Cyanobacteria* but instead are rather widespread in Prokaryotes. From an evolutionary perspective, KaiC is the oldest Kai protein, and its homologs are found in both archaea and bacteria (**Figure 5.3a**). Dvornyk and co-workers suggested that the predecessor of KaiC (pKaiC) was present in the Last Universal Common Ancestor (LUCA) about 3,800 million years ago (Mya) (Dvornyk et al., 2003). During the next 300 Mya, pKaiC duplicated and subsequently fusion to form a double-domain structure that is essential for the functioning of the circadian oscillation. It was hypothesized that the single-domain version of *kaiC* gene was horizontally transferred from Bacteria to Archaea where *kaiC* subsequently duplicated and fusion (Dvornyk et al., 2003). Multiple lateral transfers of the *kaiC* genes within and between major bacterial clades are thought to be one of the major factors shaping their evolution. The origin of KaiB is estimated between ~ 3,500 and 2,320 Mya, which coincides to the beginning of "the age of *Cyanobacteria*" (**Figure 5.3a**). When *Cyanobacteria* started to produce oxygen and thus replaced the reducing geochemical environment on Earth, biological clocks could have stated significant adaptative advantages as they provide anticipative control on a wide variety of vital metabolic cycles (Dvornyk et al., 2003). KaiA, however, the youngest

evolutionarily component of the free-running cyanobacterial pacemaker, occurs in *Cyanobacteria* about 1,000 Mya (**Figure 5.3a**).

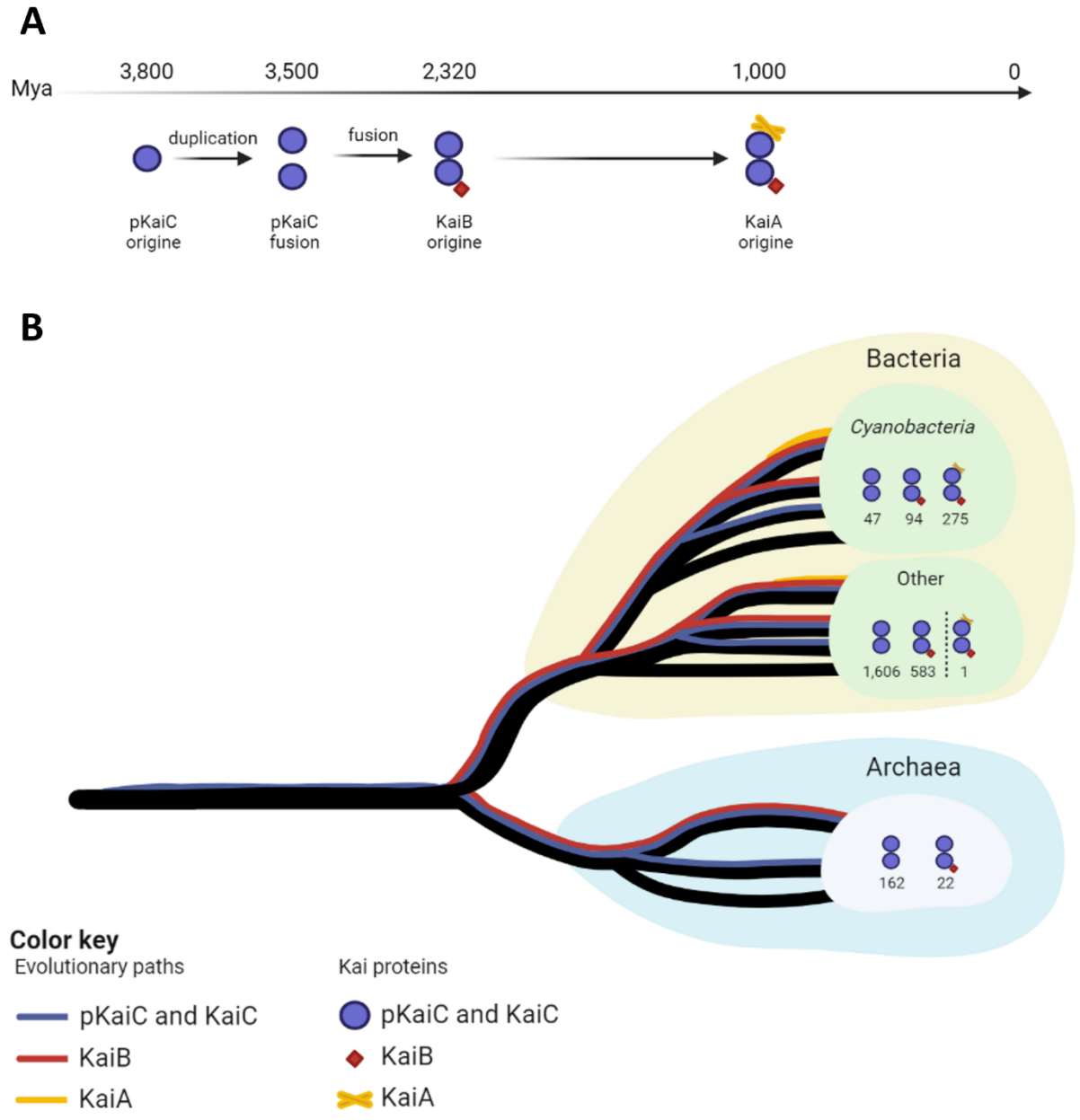


Figure 5.3: Schematic view of the distribution of Kai protein homologs in Prokaryotes. **a**: Kai evolutionary events were adapted from Dvornyk et al., 2003. **b**: Tree branches and Kai evolutionary paths are not on scale and do not reflect the evolution of kai genes in Prokaryotes but illustrate the distribution of current Kai protein homologs carriers in Archaea and Bacteria. The number of unique species carrying KaiC homologs alone or with KaiB and/or KaiA are represented for archaea, and bacteria. For clarity purpose, bacterial phyla were not display with the exception of *Cyanobacteria*.

In this review, we used a bioinformatic approach to search for all the prokaryotic KaiA, KaiB and KaiC protein sequences and update the list of archaeal and bacterial species harboring Kai homologs (**Appendix F**). In combination with the increasing number of sequence data, reaching more than 200,000 bacterial and archaeal complete or draft genomes in 2020 (Zhang *et al.*, 2020), we identified Kai protein homologs in 186 archaeal and 2,824 bacterial species/strains among the protein sequences available on UniProt in October 2020 (**Table 5.1**). Briefly, we retrieved all known KaiA, KaiB and KaiC protein sequences from UniProt (December 2020), which resulted in 114, 695 and 1,531 protein sequences, respectively. Then, we performed sequence clustering with CD-HIT v. 4.8.1 (Fu et al., 2012) for KaiA, KaiB and KaiC sequences (cutoff = 70%) obtaining 31, 105, and 420 protein sequence clusters, respectively. We determined the most representative sequences for all clusters using Diamond v. 2.0.4 (Buchfink et al., 2015) (coverage > 80%, e-value < 1e-5) and subsequently queried them against the entire UniProt database and we obtained 375, 166,266 and 90,852 sequences. As relatively low stringency was used, false positive matches are not excluded. Therefore, after the removal of duplicate sequences, an InterProScan Pfam search v. 5.47-82.0 was performed on all non-redundant potential KaiA, KaiB and KaiC protein homolog sequences. Only the protein sequences containing the kai functional protein family patterns (KaiA: PF07688, KaiB: PF07689, KaiC: PF06745) and registered as a circadian clock component (protein name) were kept for downstream analysis. This resulted in 374 KaiA, 1,127 KaiB and 4,282 KaiC protein homolog sequences, distributed across 374, 1,127, and 2,825 unique species, respectively. In the light of current genetic information available nowadays, further analyses of Kai carrier phylogeny should be performed to verify the hypotheses of the evolution of the Kai clock components and determine the role of horizontal gene transfer in their current taxonomic distribution (Dvornyk et al., 2003).

Although Kai protein homologs are present in a broad diversity of prokaryotic phyla, they are not distributed equally (**Table 5.1**). KaiC, the most essential and the oldest of the three Kai proteins is found in its functional double-domain version in 184 and 2,621 archaea and bacteria species, respectively. KaiB is observed in 24 archaeal and 1,103 bacterial species. KaiA is almost exclusively present in *Cyanobacteria* (373 species) and in one representative of *Planctomycetes* and *Actinobacteria*. Interestingly, KaiBC is represented in 22 and 677 species of archaea and bacteria, respectively (**Figure 5.3**). In Archaea, KaiBC carriers are almost exclusively *Euryarchaeota*, while in bacteria, they are distributed across 21 phyla which included both photosynthetic and non-photosynthetic microorganisms. Given the fact that a functional KaiBC timer exist in *Prochlorococcus*, it could be speculated that similar time-keeping mechanisms may be widespread in Prokaryotes.

While KaiABC was thought to be exclusively present in *Cyanobacteria*, the fact that KaiA homologs were detected in *Planctomycetes* and Actinobacteria worth further analyses. Sequence similarity analysis using Clustal 2.1 revealed that both sequences share almost 70% of similarity but only about 40% with the *S. elongatus* version of KaiA. These findings suggest that KaiABC-based clock oscillator might exist beyond *Cyanobacteria*. Complementary observations were recently reported by Köbler and colleagues who identified the gene Sll0485 gene in *Synechocystis* sp. PCC 6803 as *kaiA*-like gene that is thought to interact with KaiB3 and KaiC3 in a peripheric kai system involved in the fine-tuning of the core clock oscillator KaiA1B1C1 (Köbler et al., 2021). Orthologs of Sll0485 were identified in cyanobacterial species and in non-cyanobacterial species such as *Bradyrhizobium, Chloroherpeton, Chloroflexus*, *Rhodospirillum*, and *Roseiflexus* (Köbler et al., 2021). Further bioinformatical analysis on *kaiA* and *kaiA*-like genes would undoubtedly provide additional information on the distribution of potential KaiABC-like clock oscillators in Prokaryotes.

Table 5.1: Taxonomic distribution of Kai protein homologs in Archaea and Bacteria at phylum level. The numbers represent the unique species where at least one copy of one homolog was observed. For KaiC, only the protein homologs with at least two KaiC functional domains were considered.

	KaiA	KaiB	KaiC
Archaea			
Euryarchaeota		23	160
Crenarchaeota			2
Thaumarchaeota			1
Unclassified/Candidatus		1	21
Bacteria			
Abditibacteriota			1
Acidobacteria		8	21
Actinobacteria	1	42	67
Aquificae			2
Armatimonadetes		1	3
Bacteroidetes		209	214
Calditrichaeota		3	2
Chlamydiae		1	
Chlorobi		2	2
Chloroflexi		17	23
Cyanobacteria	373	427	450
Deinococcus-Thermus			1
Elusimicrobia		4	2
Firmicutes		4	37
Gemmatimonadetes		3	5
Ignavibacteriae		1	11
Lentisphaerae		2	2
Nitrospinae		2	1
Nitrospirae		8	9
Planctomycetes	1	31	66
Proteobacteria		297	1,666
Spirochaetes		4	3
Synergistetes		1	
Thermodesulfobacteria		1	
Verrucomicrobia		17	24
Unclassified/Candidatus		18	29

5. Towards the discovery of new prokaryotic clock systems

Over the past decades, studies revealed diel rhythmic activities beyond the well-studied *Cyanobacteria* among other prokaryotes such as extremophilic archaea, anoxygenic photosynthetic and heterotrophic bacteria. Here, we summarize the key findings and, based on our updated bioinformatical analyses of the screening of the Kai proteins, we open new

perspectives on prokaryotic time-keeping mechanisms. These candidates include anoxygenic photosynthetic bacteria, methanotrophic bacteria, methanogen archaea, sulfate-reducing bacteria, and diverse plant- and animal-associated bacteria (**Figure 5.4**).

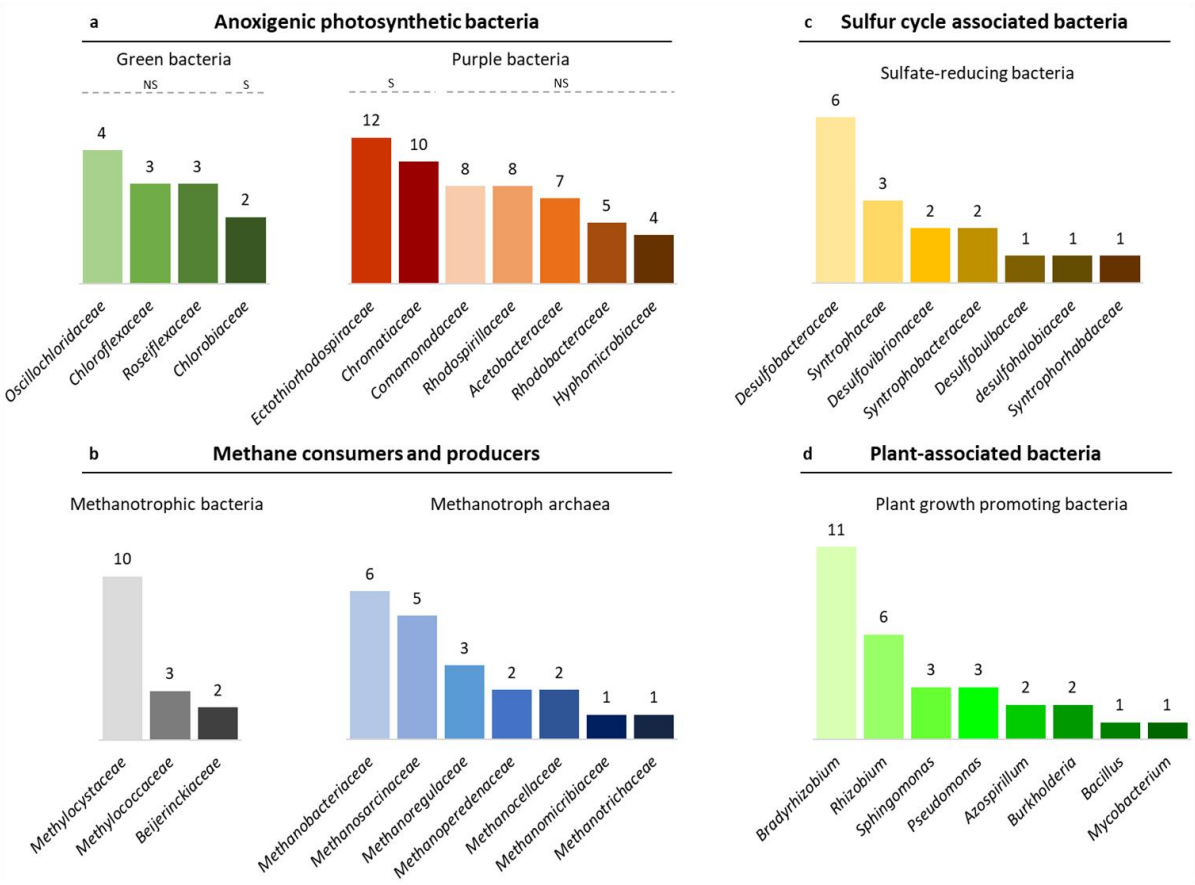


Figure 5.4: Taxonomic distribution of KaiBC carriers at family level. (a): anoxygenic photosynthetic green and purple bacteria (S: sulfur bacteria, NS: non-sulfur bacteria), (b): methanotrophic bacteria and methanotroph archaea, (c): sulfate reducing bacteria, (d): plant-growth promoting bacteria (PGPRs). The numbers represent the unique species where KaiB and KaiC homologs were observed.

We identified a total of 69 species of anoxygenic photosynthetic bacteria that possess KaiBC homolog proteins (**Figure 5.4**). Community-wide diel rhythms in anoxygenic photosynthetic bacterial communities were reported in numerous environmental studies and showed diel patterns in metabolic activity, diversity, or spatio-temporal distribution (van Gemerden et al., 1985; Garcia-Pichel et al., 1994; Fourçans et al., 2006; Fecskeová et al., 2019, Piwosz et al., 2020). Though little experimental evidence at species level is available, diel

rhythmic activity was observed in three anoxygenic photosynthetic purple bacteria: *Rhodobacter sphaeroides* (Min et al., 2005), *Rhodospirillum rubrum* (Van Praag et al., 2000), and *Rhodopseudomonas palustris* (Ma et al., 2016). In a 24h-cyclic environment, *Rhodopseudomonas* exhibits KaiC-dependent fitness enhancement in rhythmic environments but not under constant conditions (Ma et al., 2016). Although the rhythms evidenced in these purple bacteria do not strictly meet all the criteria of a bona fide circadian clock, it was suggested that *Rhodopseudomonas palustris* could harbor an ancestral endogenous circadian program, based on a KaiBC oscillator, that would include some of the canonical properties of the circadian clocks (Ma et al., 2016). We suggest further investigation based on our bioinformatic analysis in the purple-sulfur bacteria *Chromatiaceae*, the purple non-sulfur bacteria *Comamonadaceae*, and the green non-sulfur phototrophic *Chloroflexi* families, as they include several species harboring the KaiBC protein homologs and were previously associated with diel dynamics in their natural environment (Klatt et al., 2013; van Gemerden et al., 1985; Shahraki et al., 2021).

The circadian cycle was previously investigated in extremophilic archaea and revealed rhythmic activity and a potential clock mechanism based on the Kai proteins. For instance, the halophilic genus *Halobacterium* exhibit light-dark entrained daily transcription but not sustained oscillations under constant conditions (Whitehead et al., 2009). It was hypothesized that simpler timing system solely driven by KaiC could exist in the halophilic *Haloferax volcanii* (Maniscalco et al. 2014). Moreover, phosphorylation assays with KaiC homologs of the hyperthermophilic *Thermococcus litoralis* and *Pyrococcus horikoshii* have shown a conserved kinase activity (Schmelling et al. 2017). Here, we significantly broadened the list of potential clock-controlled archaea as we evidenced KaiBC homolog proteins in 22 species of methanogen archaea and two capable of denitrifying anaerobic methane oxidation (DAMO) (**Figure 5.4**). We also observed eighteen sulfate-reducing bacteria species/strains carrying

KaiBC homolog proteins (**Figure 5.4**). Methanogenic archaea and sulfate reducing bacteria are worthy of attention as they were previously associated with diel rhythmic activity in their natural environment, where they play an important role in methane and sulfur cycles (Jørgensen, 1994; Steppe & Paerl, 2002; Fourçans et al., 2007; Louyakis et al., 2017). An environmental metatranscriptomic study revealed that methanogenesis transcripts from methylotrophic methanogen archaea exhibited diel oscillations in thrombolite (Louyakis et al., 2017). In addition, diurnal oscillations of methane production were reported in diverse wetlands such as fens, mire, or peat bog where they are often associated with sulfate reducing bacteria (Mikkelä et al., 1995; Henneberger et al., 2017).

Deciphering how prokaryotes and eukaryotes are temporally embedded is crucial in the transition toward more holistic and sustainable medical and agricultural models. In this review, we evidenced 29 plant growth-promoting rhizobacteria that harbor KaiBC protein homologs (**Figure 5.4**). These bacteria were mainly represented by the genera *Bradyrhizobium* and *Rhizobium*. Both are essential to leguminous plants as they form nodules on their root hair and perform nitrogen fixation, which further promote plant development (Sessitsch et al., 2002). The circadian clock was shown to alter up to 30% of the plant transcriptome including diel carbon fluxes (Michael et al., 2008). Consequently, the plant diel rhythmicity influences the rhizosphere community structure and function (Staley et al., 2017; Hubbard et al., 2018). Unraveling how rhizobacteria anticipate diel variations promoted by plant physiology would be of agroecological interest as the plant fitness and biomass strongly rely on optimal plant/rhizobacteria bidirectional interactions (Pérez-Jaramillo et al., 2016; Hubbard et al., 2018). Although there is so far no evidence of rhythmic activity of rhizobacteria, the soil bacteria *Bacillus subtilis* was shown to display temperature-compensated free-running ~ 24h oscillations upon release to constant dark and temperature conditions (Eelderink-Chen et al., 2021). Interestingly, this bacterial species does not harbor any Kai protein homologs. A similar

case was observed in the gastrointestinal bacterium *Klebsiella aerogenes* that displayed an endogenously generated, temperature-compensated circadian rhythm in swarming motility (Paulose et al., 2019). Thus, it remains essential to elucidate if the other prokaryotes associated with animal and plants displaying circadian rhythms would rely on the Kai-based system. Alternatively, it is worth noting that other molecules are strong candidates for circadian pacemaker: peroxiredoxin proteins are found in almost all organisms and have circadian rhythms of their redox state persisting even in the absence of clock rhythmicity in photosynthetic organisms such as *Cyanobacteria* (Edgar et al., 2012).

6. Conclusion

Life experiences diel cycles since its earliest stages and time-keeping mechanisms emerged as one of the most outstanding adaptations to these perpetual environmental fluctuations. Although only few bacterial species are known to possess an effective molecular circadian clock, evidence of widespread prokaryotic timing system accumulates. These new clocks might share the canonical properties of circadian clock or rather work differently depending on the evolutionary path, the internal properties of the microorganisms, or the specificities of their environment. Prokaryotes are often embedded in complex communities and mutually dependent on well-defined clock-controlled organisms. Therefore, future chronobiology studies should carefully design their experiments when using new clock candidates as growth conditions likely shape the time-keeping mechanisms.

Chapter 6: Diel proteome and circadian genes regulation in *Rhodospirillum rubrum*

Diel cycle proteomic in the purple bacterium *Rhodospirillum rubrum*

Géron, **A**., Decroo C., Matallana-Surget, S & Wattiez, R *(In preparation)*

Abstract

Circadian clocks exist in all domains of Life, including the photosynthetic *Cyanobacteria*. Recent studies have shown evidence of rhythmic activities coupled with the presence of putative circadian clock genes in other prokaryotes, such as purple bacteria. So far, rhythmic activities potentially linked to a circadian clock were reported in three species of purple non-sulfur bacteria, including *Rhodospirillum rubrum*, a remarkable microorganism for its use in fuel and plastic bioproduction. However, little is known on the impact of light and dark cycle on its protein and putative circadian clock gene expression. Here we reported for the first time the proteome regulation of *R. rubrum* grown under continuous light versus light and dark cycle conditions by using a shotgun proteomic analysis. In addition, we measured the impact of light regimes on the expression of the four putative circadian clock genes (*Kai*) at the transcriptional and translational using RT-qPCR and target proteomic, respectively. Data showed that light conditions impacted the whole proteome regulation, especially in the early growth stages. Specific functions were found to be regulated depending on the light or dark period is biological processes such as the energy conversion pathways and general stress response. In addition, the four *Kai* genes were found to be differentially regulated depending on light conditions at both the mRNA and the protein levels. Deciphering the diel cycle impact on purple bacteria will greatly contribute to better understand their ecology, to optimize their use in biotechnology, and to shed light on the origin and evolution of prokaryotic clock mechanisms.

1. Introduction

Circadian clocks are endogenous biological mechanisms that synchronize a wide array of biological processes oscillating with the diel cycle (Saini et al., 2019). While they are ubiquitous in Eukaryotes, they were only characterized in the oxygenic photosynthetic bacteria (i.e., *Cyanobacteria*) in Prokaryotes (Huang et al., 1990; Chen et al., 1991). The circadian clock of *Synechococcus elongatus* – the model organism in the study of the cyanobacterial circadian clock – is based on the Kai proteins (i.e., KaiA, KaiB, and KaiC) that form a core oscillator synchronizing with environmental signals and controlling gene expression (Kondo, 2007; Markson et al., 2013). For more detailed reviews on the KaiABC oscillator functioning and its regulatory network, see Cohen & Golden et al., 2015; Swan et al., 2018; Snijder & Axmann, 2019).

The composition and the functioning of Kai systems in *Cyanobacteria* strongly vary but are less documented (Schmelling et al., 2017). For instance, the globally distributed marine *Prochlorococcus marinus* has a simplified clock mechanism that lacks KaiA (Axmann et al., 2009). In contrast, *Synechocystis* sp. PCC 6803 encodes multiple Kai proteins, which are suggested to be involved in the finetuning of the core oscillator (Aoki & Onai, 2009; Kanesaki et al., 2012; Wiegard et al., 2020; Köbler et al., 2021). Importantly, the occurrence of Kai proteins is not restricted to *Cyanobacteria*. In fact, bioinformatical analyses revealed that homologs of the Kai proteins are widespread in Prokaryotes and present in purple-bacteria (Dvornyk et al., 2003; Loza-Correa et al., 2010; Schmelling et al., 2017). Rhythmic activities of in purple non-sulfur bacteria harboring Kai homologs was reported in three species: *Rhodopseudomonas palustris* (Ma et al., 2016), *Rhodobacter sphaeroides* (Min et al., 2005), and *Rhodospirillum rubrum* (Van Praag et al., 2000). Purple non-sulfur bacteria display high

metabolic versatility and are used in broad biotechnological applications including bioproduction, biofertilization, or wastewater treatment) (De Meur et al., 2020).

In this study we investigated the diel cycle response of the purple non-sulfur and nitrogen fixating *Rhodospirillum rubrum* ATCC 11170 (hereafter *Rhodospirillum rubrum*). *R. rubrum* has a highly versatile metabolism, and it is capable of photoautotrophic and heterotrophic growth (McEwan, 1994). This makes this bacterium particularly interesting for biotechnological applications such as bioplastic and hydrogen fuel production (Bayon-Vincente et al., 2020; Rodríguez et al., 2021). *R. rubrum* possesses two homologs of the circadian clock *kaiB* and *kaiC* genes and one potential homolog of one of the gene *kaiA* of *Synechocystis* sp. PCC 6803 (Köbler et al., 2021). Interestingly, rhythmic activity of the uptake hydrogenase (Hup), an enzyme involved in the consumption of H_2, was reported using an enzyme assay (Van Praag et al., 2000). The authors suggested that the rhythmic hydrogenase activity could be involved in a mechanism coordinating the energy metabolism in *R. rubrum*. However, information is still missing about the general response and the expression of Kai genes of *R. rubrum* under diel cycle.

Here, we performed the most exhaustive molecular study of the impact of light regimes on the proteome and Kai genes expression in *R. rubrum*. Cultures of *R. rubrum* were grown a phytotron under continuous light exposure (LL) and 12h/12h light and dark cycle (LD). The whole proteome fluctuations were measured using shotgun proteomics by Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH-MS) analysis. In addition, we quantified the expression regulation of the two *kaiB* and *kaiC* gene homologs at both the transcriptional and translational levels, using Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR) and Multiple Reaction Monitoring (MRM-MS).

2. Materials and methods

2.1. Bacterial culture conditions and sampling

In this study, *R. rubrum* was grown in two independent experiences for downstream RT-qPCR and proteomics analyses, respectively (**Figure 6.1**). In both experiences, *R. rubrum* was plated on supplemented malate-ammonium medium (SMN) (Fitzmaurice et al., 1989) agar petri dishes at 30°C in complete darkness. Then, single colony forming unit were inoculated in tubes filled with 15ml of SMN liquid medium and pre-cultured at 25°C, under aerobic conditions and shaken at 150 rpm. For each experience, pre-cultures were grown under two different light conditions: (i) continuously light exposure (LL) and (ii) 12/12h light and dark cycle (LD). After 72h, LL and LD pre-cultures with similar optical density (2.1 ± 0.1 at 680nm) were selected and 35µl were inoculated in 150ml of SMN liquid medium (n=3). The culture flasks were exposed to similar light, temperature, and shaking conditions as the pre-culture. Every 8 hours, 6ml of bacterial culture were sampled in each flask, from the end of the lag phase (~ 40h) until the beginning of the stationary phase (~96h). Samples were collected at 04:00 pm (i.e., 48h after the start of the experiment), 00:00 am (i.e., 56h), 08:00 am (i.e., 64h), 04:00 pm (i.e., 72h), 00:00 am (i.e., 80h), 08:00 am (i.e., 88h), and 04:00 pm (i.e., 96h). In total, 7 samples were collected per replicate and per light condition. In the LD condition, light was switched off from 08:00 pm to 08:00 am. The optical density was measured immediately after sampling by spectrophotometry (680nm). Cells were harvested by centrifugation (16,000 × g, 4°C). Cell pellets were washed twice with BupH ^{TM} PBS solution (Thermo Fisher Scientific, Massachusetts, United-States) and stored at -20°C for downstream RT-qPCR and proteomics analyses.

2.2. Protein and RNA isolation

For the RT-qPCR experiment, total RNA was extracted using the RNeasy Protect Bacteria Mini Kit (50) (Qiagen, Hilden, Germany) following fabricant instruction. RNA concentration was measured using a Nanodrop spectrophotometer (Appendix G). The Reverse Transcriptase Core Kit (300) (Eurogentec, Seraing, Belgium) was used following fabricant instruction to generate complementary DNA (cDNA) from the RNA templates. Briefly, RNA template (15ng/µl) was mixed with random nonamers (2.5 µM), dNTP (500µM of each dNTP), $MgCL_2$ (5 mM), EuroScript RT (1.25 U/µl), RNAse inhibitor (0.4 U/µl), RNase free water and reaction buffer. The reverse transcription was performed in a Real-Time thermocycler with an initial step of 10min at 25°C, followed by a revers transcription step of 30min at 48°C and finally, with inactivation of the RT enzyme during 5min at 95°C. cDNA was stored at 4°C until qPCR analyses.

For the proteomic analyses, proteins were extracted using extraction buffer (guanidinium hydrochloride 6M; K_2HPO_4 50mM) and ultrasonication (4°C, 3 × 10 s, amplitude: 20%, IKA U50 sonicator (Staufen, Germany)). The protein content of the supernatant was evaluated using the Bradford assay (Bradford, 1976) with bovine gamma globulin as standard. Then, a total of 50 µg of the proteins was reduced with 1.5-dithioerythritol (DTE), alkylated with iodoacetamide (IAA), and precipitated overnight using acetone at -20°C. Next, the protein pellets were solubilized using 50 mM ammonium bicarbonate containing 1µg of trypsin and incubated overnight at 37°C. Digestion was stopped by adding 0.5% formic acid in water (vol/vol [0.1% final concentration]). After the digestion, the peptides were quantified using the kit Pierce ™ Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, Massachusetts, United-States). Peptides were stored at -20°C for downstream proteomics analyses.

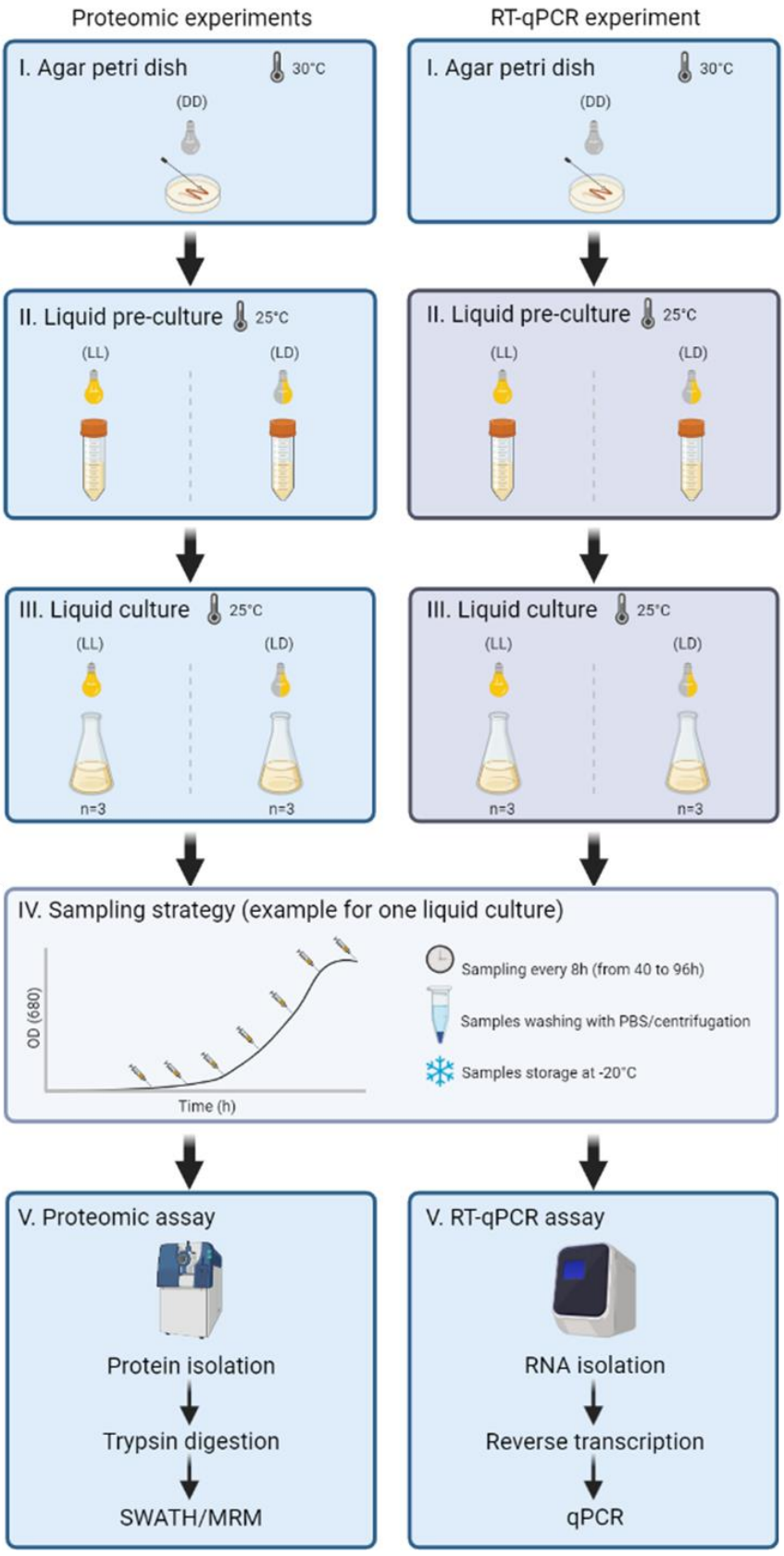


Figure 6.1: Workflow to study the diel cycle impact in *R. rubrum*.

2.3. RT-qPCR analyses

cDNA was further used as template for qPCR assay using the Takyon™ Rox SYBR Core Kit blue dTTP (1250) (Eurogentec, Seraing, Belgium). The primers used for qPCR assay were targeting the Rru_A2542 (hereafter *kaiC1*), Rru_A2543 (hereafter *kaiB1*), Rru_A3294 (hereafter *kaiB2*), and Rru_A3295 (hereafter *kaiC2*) reverse transcripts of *R. rubrum* (**Table 6.1**). The physicochemical properties of each pair of primers were checked using the ThermoFisher Scientific Multiple Primer Analyzer tool (Breslauer et al., 1986). Their specificity was checked using SnapGene Viewer 4.2.11 (Dotmatics, Bishop's Stortford, UK) and using classic PCR of pure culture of *R. rubrum*. A StepOnePlus Real-Time PCR System (Thermo Fisher Scientist) was used for the quantification. The relative concentration of *Kai* was calculated as the ratio of their expression and that of the 16S rRNA using universal primers 518R and 341F (**Table 6.1**) according to Pfaffl et al., 2002 (1). Results were presented as a relative expression ratio of the targeted gene (*kaiB1*, *kaiC1*, *kaiB2*, or *kaiC2*) genes versus a reference gene (*16SrRNA*) for each sample time, in comparison with the expression at T0 (i.e., 48h).

$$\text{Ratio} = (E_{target})^{\Delta CP\ target\ (T0 - Tx)} / (E_{ref})^{\Delta CP\ ref\ (T0 - Tx)} \quad (1)$$

where E is the PCR efficiency calculated as defined in Ramakers et *al*., 2003 and CP is the crossing point of the amplification curve with the threshold.

Table 6.1: Primer sequences targeting *kai* gene homologs in *R. rubrum.*

	Primers	
	Forward (5'-3')	*Reverse (3'-5')*
16S rRNA	CCTACGGGAGGCAGCAG	ATTACCGCGGCTGCTGG
kaiB1	GCCCACGGAAACTAACGCTC	GTTCCGCGCAAATCCGTTC
kaiB2	GATGTGATCGACAGTCCCGC	AGATCAAGGATGCGGCACAC
kaiC1	TTCAGCGTTCTTCCCGTCTC	GACCAGGATGCTTGATCCCC
kaiC2	TCTTTTCCGCCCAGTTCCTG	TCGACGAAGCTCCATTTCCC

qPCR reactions were performed for each DNA sample (n=3) and targeted gene combination using the SYBR ™ Green master mix (Applied Biosystems ™, Waltham, Massachusetts, USA). Each qPCR consisted in a serial of dilution (2.5, 1.2, 0.6 and 0.3 ng/µL) of cDNA template and qPCR mix following fabricant instruction. After an initial denaturation step at 94°C during 10 min, 40 cycles of 15 sec at 95°C, 30 sec at 60°C and 30 sec at 72°C were performed followed by a final denaturation step (60 to 90°C, +0.3°C/min).

2.4. Proteomic analysis

Protein identification and quantification were performed according to a label-free strategy on a UHPLC HRMS platform (Eksigent 2D Ultra-AB Sciex TripleTOF 6600+). For each sample, the peptides were separated in a 15-cm C_{18} column (YMCC18) using a linear acetonitrile gradient (5 to 35% [vol/vol] in 75 min) in water containing 0.1% (vol/vol) formic acid at a flow rate of 5µl min^{-1}. For protein identification, data were acquired in a data-dependent acquisition mode (DDA) for 4µg peptides on column. Mass spectra (MS) were acquired over the range of 400-1,250 m/z in high-resolution mode (resolution > 35,000), with a 250 msec accumulation time. MS/MS spectra were acquired over the range of 100-1,500 m/z. The precursor selection parameters were as follows: intensity threshold, 100 cps; maximum precursors per cycle, 90; accumulation time, 25 msec; and exclusion time after two spectra, 15 sec. These parameters lead to a duty cycle of 4 sec per cycle to ensure that high quality extracted ion chromatograms (XICs) were obtained for peptide quantification. ProteinPilot Software (v5.0.1 – ABSciex, United States) was used to perform database searches against the UniProt database, restricted to *Rhodospirillum rubrum* ATCC 11170 entries. The search parameters included differential amino acid mass shifts for carbamidomethyl cysteine, all biological modifications, missed trypsin cleavage sites.

For whole proteins relative quantitation analyses, the instrument was set in SWATH mode. Briefly, 100 incremental steps were defined as windows of variable m/z values over a 400–1250 m/z mass range. The MS/MS working time for each window was 50 msec, leading to a duty cycle of 5 sec per cycle. The ion chromatogram of the top six fragments of each peptide was extracted, and their area under the curve was integrated. PeakView® software (version 2.1.0.11041, ABSciex, United States) was used for the SWATH processing of all proteins identified considering an FDR below 1% (as determined by ProteinPilot). The retention time (RT) was recalibrated automatically with PepCalMix standard (ABSciex, United States) with retention times in the range of 20–85 min. Intensity of peptides were individually normalized based on a summed area of all peptides for each sample. Only proteins quantified with a minimum of 2 peptides were considered. A non-supervised Pareto-PCA analysis was performed according to the area of protein curve to discriminate comparative groups (MarkerViewTM MarkerView™ 1.2.1 ABSciex, United States). For proteome comparisons, only fold change higher than 1.5 (up-regulated in LL compared to LD) or lower than 0.66 (down-regulated in LL compared to LD) and having a p-value lower than 0.05 were further considered. Each differentially abundant protein's function was checked manually consulting research articles, Uniprot and NCBI data base.

The MRM analyses were performed using a QTRAP 6500+ instrument (SCIEX) fitted with an electrospray ionization source (150 °C, 4500V). Transition selection and MRM method optimization using the Skyline software (20.2.0.343 MacCoss Lab) on protein extracted from *R. rubrum* grown under light and dark cycle. The best transitions (y or b ions) were chosen for each peptide and at least two peptides were analyzed for each protein. Peptides were separated on a C18-reversed phase column (YMC TriArt C18, 0.3mm, 150 mm) and eluted using a gradient of 5–35% acetonitrile with 0.1% formic acid over 20 min at a flow rate of 5 µl/min. MRM data were acquired in scheduled mode with two minutes retention time window and a

maximum cycle time of 1.5 sec. Skyline software (20.2.0.343 MacCoss Lab) was used for visual inspection of MRM data and area under the curve integration. Peak picking for each peptide was manually refined using the transition intensity ratio and retention time as leading parameters. The intensity of all transitions was summed up for each peptide. Protein abundance was obtained as the average of the Ln-transformed area under the curve of each target peptides detected in the three samples.

3. Results and discussion

3.1. Impact of light conditions on *R. rubrum* growth

The impact of the light and dark cycle on growth was negligible in the lag phase and the early exponential phase (**Figure 6.2**). From 72h, the optical density was significantly higher under continuous light exposition and reached a final average maximum of 3.16 ± 0.07 (LL) and 2.69 ± 0.09 (LD) (p-value = 0.002) (**Figure 6.2**). *R. rubrum* is a facultative anaerobe capable of aerobic heterotrophic growth and anaerobic photosynthesis growth (McEwan, 1994). In this experimental setup, heterotrophic growth was clearly favored as *R. rubrum* was in oxic condition. However, the color of the cultures in both LL and LD conditions turned from whitish to reddish overtime, which suggests that photosynthesis was progressively activated (Takaichi, 2009).

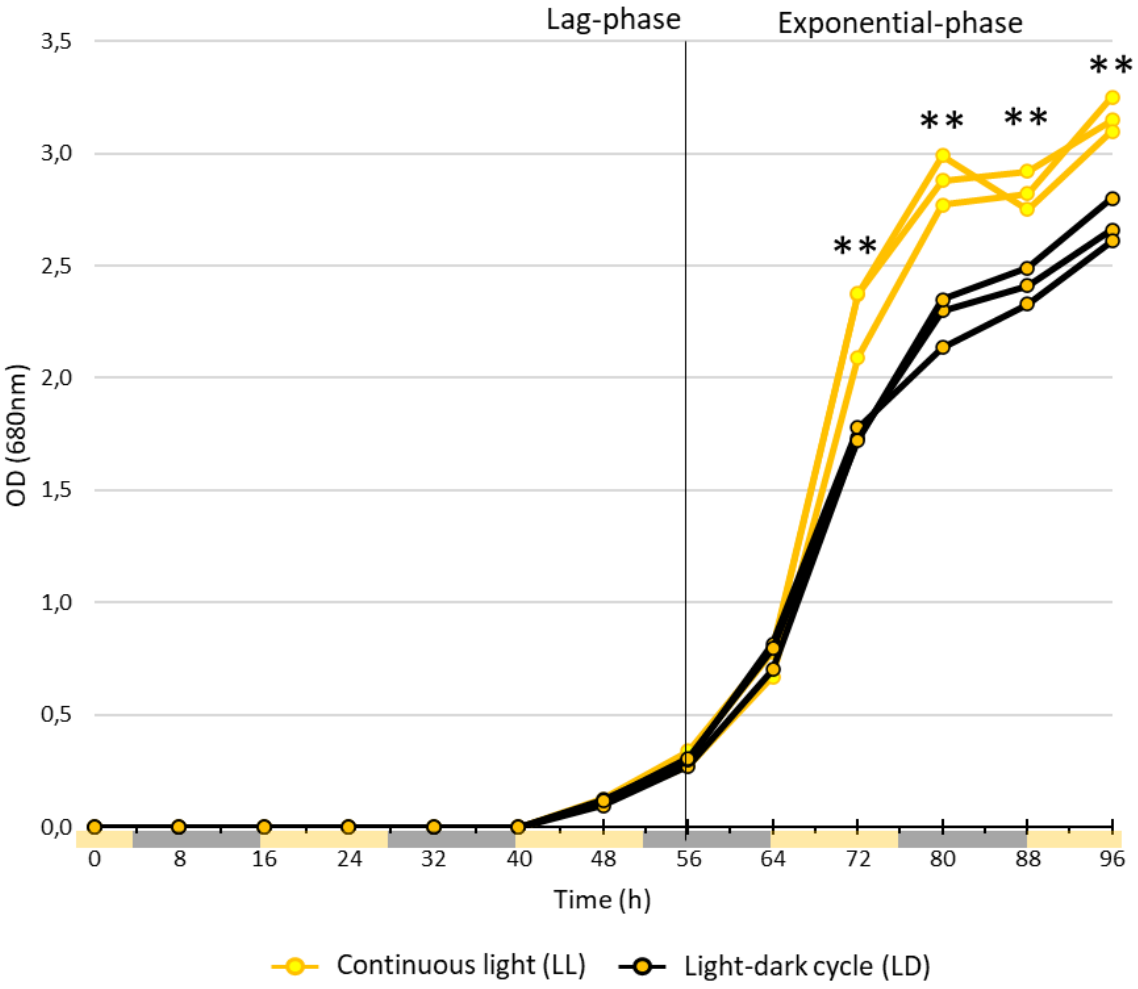


Figure 6.2: Growth curves of *R. rubrum* depending on light conditions. Growth curves were computed based on optical density (680nm). Light and dark phases in the LD condition are represented on the time axis in yellow and grey, respectively. Significative differences between LL and LD samples are shown with a * (p value ≤ 0.1) or ** (p value ≤ 0.05) and were calculated with a t-test.

3.2. Proteomic analysis of *R. rubrum* under LD cycle

For the first time, we presented the quantitative whole proteome analysis of the *R. rubrum* growth conducted over a 56-h time span under LL and LD conditions. A total of 1901 unique proteins were identified, covering 50 % of the predicted *R. rubrum* proteome (3,835 proteins). Non-supervised Pareto-PCA analysis revealed that the LL and LD proteomes were mainly grouped based on the growth stages: i) the lag phase proteomes (LL: 48h and 56h, LD: 48h); ii) the early exponential phase proteomes (LL: 64h, LD: 56h and 64h); iii) the mid exponential phase proteomes (LL and LD: 72h); and iv) the late exponential phase proteomes

(LL and LD: 80h, 88h, and 96h) (**Figure 6.3**). An exception was observed in the earliest stages (groups 1 and 2), where the LL and LD 56h proteomes grouped differently, suggesting a differential proteome regulation between LL and LD cultures in the early exponential phase (**Figure 6.3**).

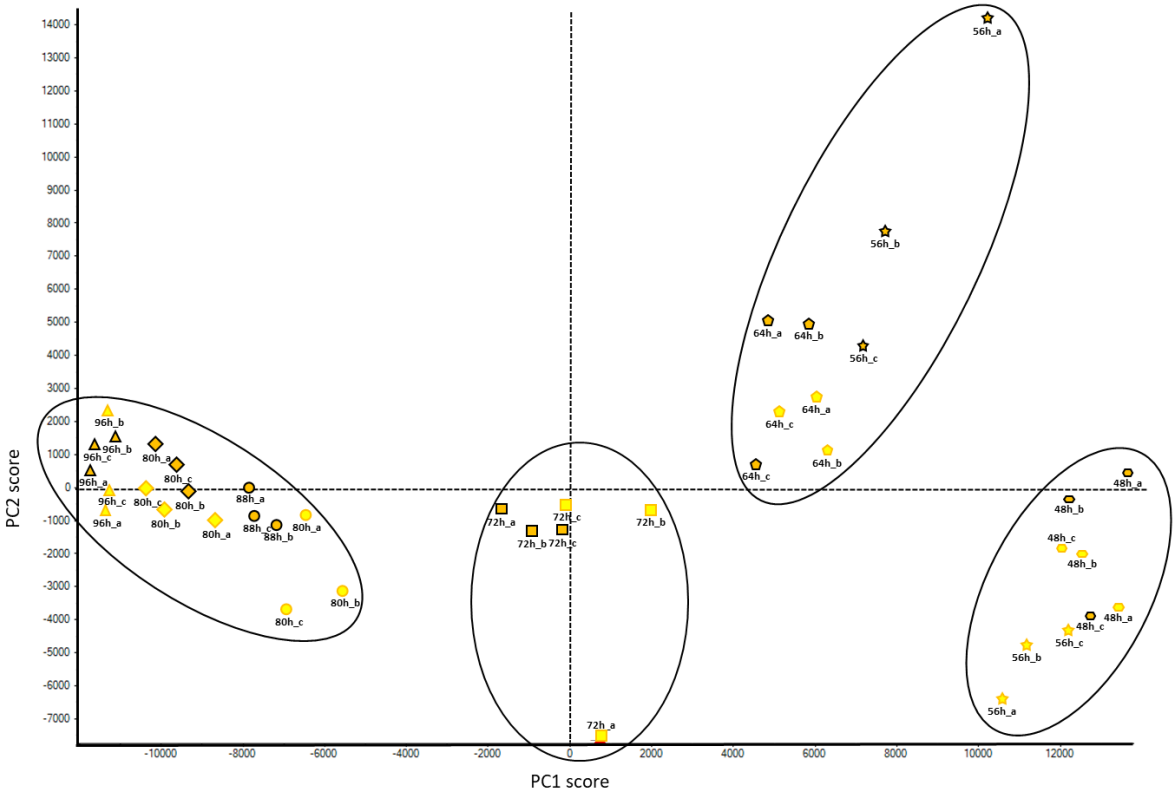


Figure 6.3: Non-supervised Pareto-PCA analysis of *R. rubrum* proteomes under different light conditions. Scatter plot of the first two components PC1 (48.7%) and PC2 (8.0%). Orange filled and black outlined shapes correspond to LD proteomes. Yellow filled and orange outlined shapes correspond to LL proteomes.

3.2.1. Cyclic protein regulation

Next, we investigated the up- and downregulated proteins in LL compared to LD conditions for each sampling point. In total, 548 proteins were found to be significantly up- and/or downregulated at 1 out of 7 time points or more (**Figure 6.4**). The time point when the highest number of regulated proteins was observed was at 56h, with 97 and 149 up- and downregulated proteins, respectively (**Figure 6.4**). This was consistent with the PCA analysis

that showed high distance between LL 56h and LD 56h proteomes (**Figure 6.3**). LL downregulated proteins were also abundant over the late exponential phase (80h-96h), meaning that under LD condition, those functions were upregulated (**Figure 6.4**). To facilitate the interpretation of the protein regulation, we sorted them into protein regulation profiles (**Appendix H**). Most of the proteins (i.e., 411) were up- or downregulated punctually (at one single time point) and only 20 proteins were found to show a cyclic regulation profile (strictly up or down regulated over two consecutive light or dark time points). This might be due to the fact that protein abundance was measured over the bacterial growth when the proteome phenotype undergoes strong fluctuations independently on the light condition (**Figure 6.3**).

Interestingly, we observed that numerous transcriptional regulatory proteins were differentially regulated depending on the light conditions, especially during the time points corresponding to dark phases in LD condition (**Appendix I)**. These proteins were involved in various biological processes such as sugar catabolism, oxidant tolerance and virulence or even metal stress. Although information flows from DNA genes to mRNA transcripts to proteins (Crick, 1970), much remains to be solved regarding how modifications at each of the gene product level can affect downstream activity (Picard et al., 2009). This might have implications for the interpretation of the light and dark cycle impact of *R. rubrum*. Indeed, light-dark induced oscillations at the mRNA level are well established in clock-controlled *Cyanobacteria* such as *Synechococcus* or *Prochlorococcus* but significant discrepancies between the diel oscillations at mRNA and protein levels exist (Ito et al., 2009; Waldbauer et al., 2012; Guerreiro et al., 2014). For instance, in *S. elongatus*, only 14% of its proteome display LD oscillation (Guerreiro et al., 2014), while 30% to 60% of the mRNA levels exhibited cyclic profiles (Ito et al., 2009; Vijayan et al., 2009).

3.2.2. Impact of light conditions on biological processes

To estimate whether the differentially regulated proteins were biologically connected, we searched them in the STRING database (Mering et al., 2003) (**Figure 6.4**). Despite the satisfying proteome coverage obtained for *R. rubrum*, we observed that 114 of the regulated proteins were annotated as uncharacterized functions. These proteins were even found in the most regulated proteins in all proteomes and could be implicated in all kinds of biological processes discussed further.

This analysis revealed that some of the up and down-regulated proteins formed networks with significantly more interactions than expected in a random set of proteins and displayed significative functional enrichments (**Figure 6.4**). These enrichments were observed during the corresponding dark phases in LD condition and during the late growth phase. At the beginning of the first dark phase (i.e., 56h), ribosomal proteins, and proteins involved in oxidative phosphorylation, protein export, and bacterial secretion system were enriched among LL up-regulated proteins. At the end of the first dark phase (i.e., 64h), proteins involved in transcription repair and organic solvent tolerance were enriched among LL up-regulated proteins. In the late growth phase (i.e., 96h), ribosomal proteins and proteins involved in electron transfer activity were enriched among LL down-regulated proteins, while proteins involved in miss-match repair system were enriched among LL up-regulated proteins. These results suggest that under LL condition, *R. rubrum* faces higher stress inducing repair systems to be upregulated. On the other hand, under LD condition, *R. rubrum* might have lowered its metabolic activity (protein synthesis/export and oxidative phosphorylation) during the first dark period then increased it over the late growth phase.

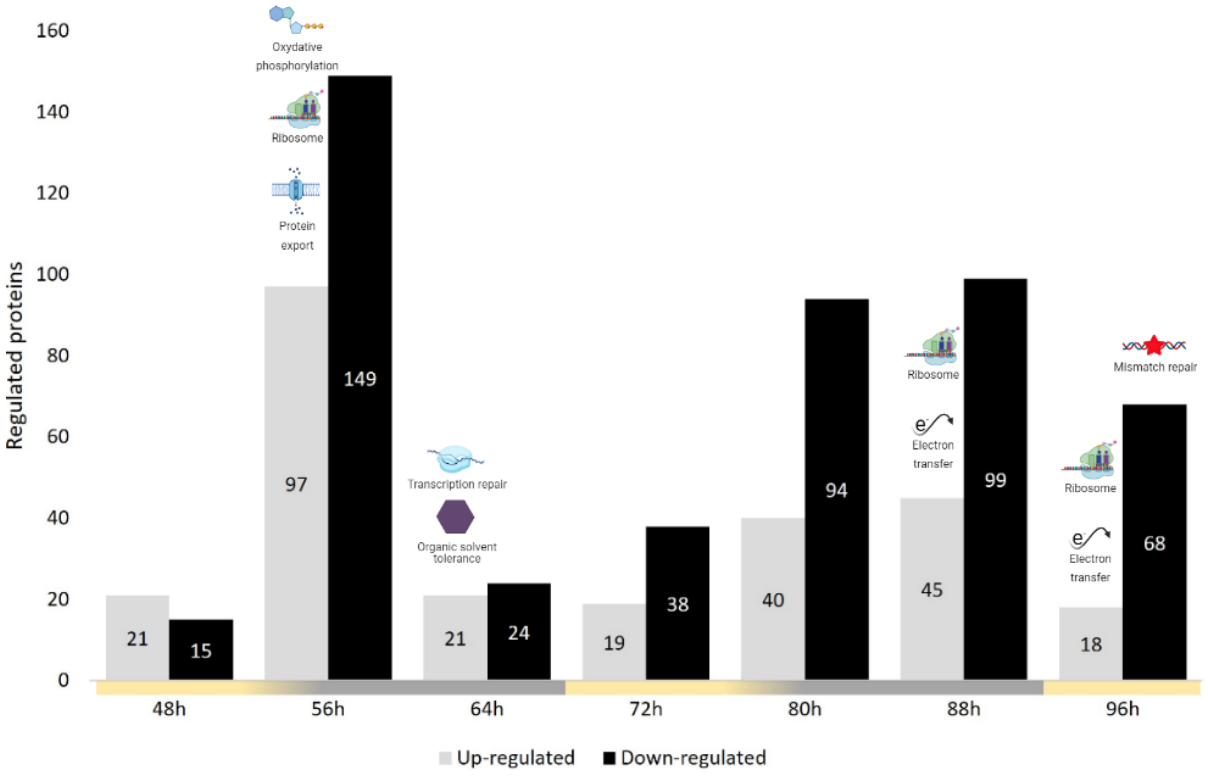


Figure 6.4: Up- and downregulated proteins in LL condition compared to LD condition in *R. rubrum*. Network STRING analysis results have been illustrated for each up- and downregulated protein subsets. The absence of illustration means that no significative functional enrichment was identified. Light and dark phases in the LD condition are represented on the time axis in yellow and grey, respectively.

Differential regulation of proteins involved in photosynthesis and oxygenic respiration were observed over time under different light treatments. In total, 34 proteins were found to be up/down regulated in LL compared to LD conditions (**Appendix J**). These proteins were involved in both the respiratory and the photochemical electron transport systems, the photosynthesis reaction center, the biosynthesis of bacteriochlorophyll, or the synthesis of ATP. While no clear distinct metabolic pattern was observed between LL and LD conditions, the proteins involved in the synthesis of light harvesting proteins were upregulated under LD condition in the early growth (i.e., 48h to 72h). Under LL condition, the proteins involved in anoxygenic respiration were overall upregulated in early growth and then downregulated in late growth under. Since the respiratory and the photochemical electron transport systems of *R.*

rubrum are associated (Thore et al., 1969), further investigation in strict photoautotrophic or heterotrophic growth conditions should better reflect the potential impact of diel cycle on the energetic metabolism regulation of *R. rubrum*.

Stress related proteins were also differentially regulated between LL and LD grown *R. rubrum* cultures (**Appendix K**). The heat shock protein Hsp20, involved in the response to an array of stresses, including hyperthermia, oxygen radicals, or heavy metals (Maio, 1999), was overall upregulated in LD cultures. In contrast, the copper/zinc superoxide dismutase was found to be strongly up-regulated at 56h in LL condition (FC: 4.33, *p-value* = 0.04), which correlates with the strong upregulation of NADH ubiquinone oxidoreductase at the same time in LL (FC: 4.85, p-value = 0.000095), which is a major source of reactive oxygen species (ROS) (Esterházy et al., 2008). Moreover, we observed LL upregulated transcription repair system (**Figure 6.4**), and a strong up-regulation of the SOS-response transcriptional repressor LexA in the LL cultures (FC (72h): 3.89, *p-value* = 0.006, FC (80h): 35.21, p-value = 0.00028). LexA is a DNA repair system repressor which is cleaved from DNA by RecA when DNA is damaged (Slilaty, 1987; Neher, 2003). These results suggest that *R. rubrum* cultures maintained under LL condition are exposed to higher oxidative stress and light induced damages when compared to LD cycles (Bayon-Vincente, 2020).

The nitrogen regulatory protein P-II was observed in one time point (i.e., 56h) and was strongly up-regulated in LD condition during the first dark phase (**Appendix L)**. This observation refers to the only study presenting rhythmic activity of the uptake hydrogenase, which is involved in consumption of H_2 produced by nitrogenase hydrogen (Van Praag et al., 2000). Another small subset of proteins associated with virus defense mechanism (e.g., CRISPR system) or host integration factor were differentially regulated between LL and LD conditions (**Appendix L**). In the environment, viral infection was shown to follow a diel pattern and infection occurred during the night (Winter et al., 2004). We also observed four proteins

involved in motility (i.e., flagellin) that were upregulated during the dark periods as well as chemotaxis related proteins (**Appendix L**). Bacteria utilize motility and chemotaxis (i.e., regulation of motility towards chemical attractants and away from chemical repellents) for optimal growth (Berleman & Bauer, 2005). Interestingly, diel motility patterns have been observed at the environmental level (Grossart et al., 2001; Mitchell & Kogure, 2006). Additionally, in the gastrointestinal bacterium *Klebsiella aerogenes*, an endogenously generated, temperature-compensated circadian rhythm in swarming motility has been recently identified (Paulose et al., 2019).

3.3. *kai* genes expression

The four homologous circadian clock proteins, KaiB1, KaiB2, KaiC1 and KaiC2 were all detected in the SWATH analyses of *R. rubrum*. However, only KaiC1 and KaiC2 were found to be slightly differentially regulated at a single time point (i.e., 88h) (**Appendix L**). Therefore, we quantified the expression of the four kai gene homologs using RT-qPCR and targeted proteomic (**Figures 6.5 and 6.6**). Since measurements were done in samples that come from two different experiments (**Figure 6.1**), direct comparisons were limited.

At the transcriptional level, the expression of *kai* mRNAs in *R. rubrum* cultures maintained under LL and LD conditions was strongly impacted by light condition and display important fluctuation in relative abundance (**Figure 6.5**). In LL condition, *KaiB2* and *KaiC2* mRNAs were overexpressed and reached maximal abundances three and ten times higher than in LD, respectively. In contrast, *kaiB1* and *kaiC1* were twice more expressed in LD than in LL.

At the translational level, the expression of Kai proteins under LD condition showed less variability than under LL condition (**Figure 6.6**). Differences were mostly observed in the late growth phase when LL KaiC1, KaiB2, and KaiC2 proteins displayed higher intensity compared with LD. In contrast, the intensity of LL KaiB1 was higher than LD KaiB1 until 64h then remained similar.

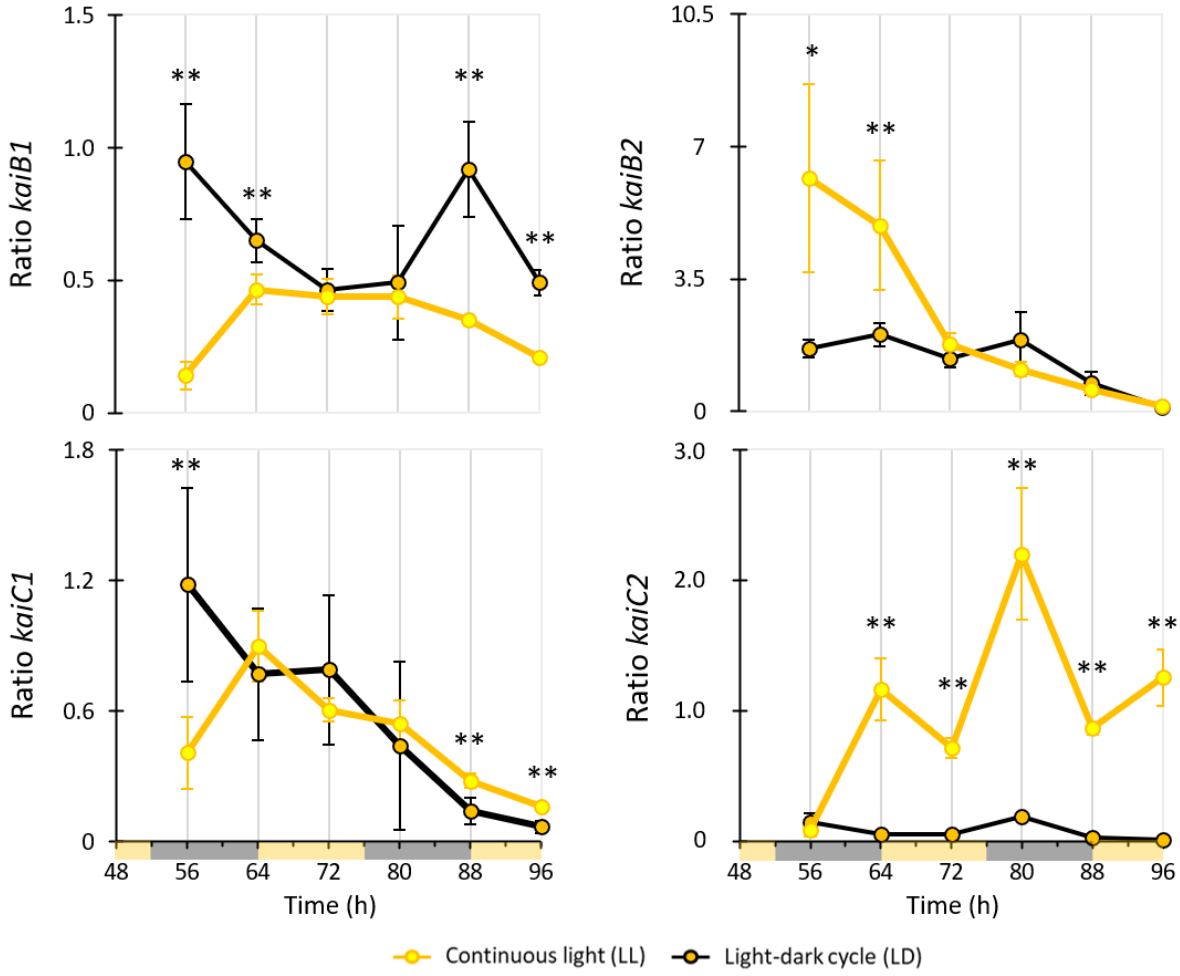


Figure 6.5: Evolution of the abundance of Kai proteins in *R. rubrum* depending on light conditions. Results were presented as a relative expression ratio of the targeted gene (*kaiB1*, *kaiC1*, *kaiB2*, or *kaiC2*) genes versus a reference gene (*16SrRNA*) for each sample time, in comparison with the expression at T0 (i.e., 48h). Light and dark phases in the LD condition are represented on the time axis in yellow and grey, respectively. Significative differences between LL and LD samples are shown with a * (p value ≤ 0.1) or ** (p value ≤ 0.05) and were calculated with a t-test.

While the current data did not allow us to conclude on the rhythmic transcriptional or translational regulation of *kai* genes under light and dark periods, they showed that *R. rubrum* expresses these genes and that light conditions impact their expression at both mRNA and protein level. Although in *S. elongatus*, *kaiB* and *kaiC* genes show diel rhythmicity at both transcriptional and translational levels (Kitayama et al., 2003, Ito et al., 2009), the main regulation of the Kai oscillator occurs at the posttranslational level by phosphorylation and dephosphorylation cycle (Cohen & Golden, 2015). The role of Kai proteins in *R. rubrum* is still unknown and their function might perhaps be related with a timing system that yet remain to be identified. If we have a closer look at the position of *kai* genes in the genome of *R. rubrum*, we see that *kaiB2* and *kaiC2* are preceded by the gene Rru_A3293, recently identified as an ortholog of the *kaiA3* gene in *Synechocystis* by Köbler and colleagues (Köbler et al., 2021). Further measurement of *kai* gene homologs transcriptional and translational expression should include Rru_A3293 for a better visualization of the clock components regulation. Additionally, phosphorylated peptide enrichment mass spectrometry analysis would be more pertinent to measure the regulation of the potential Kai oscillator.

Finally, it is worth mentioning that both *KaiBC1* and *KaiBC2* gene clusters are followed by a histidine kinase protein containing a PAS domain, Rru_A2544 and Rru_A3296, respectively. In the purple non-sulfur bacteria *Rhodobacter sphaeroides* a similar protein was found near the N-terminal region of *kaiBC* gene operon. *R. sphaeroides* of gene expression controlled by oxygen (20.5h under aerobic conditions and 10.6-12.7h under anaerobic conditions). In *R. sphaeroides*, the histidine kinase protein was suggested to be involved in sensing oxygen and transducing redox signals to the central clock (Gonzalez et al., 1998; Min et al., 2005). Therefore, it is possible that the potential time-keeping mechanism of *R. rubrum* is regulated by other type of environmental factor than light, such as oxygen like in *R. sphaeroides* (Min et al., 2005). In their natural environment, purple bacteria are embedded

within complex communities including clock-controlled organisms such as *Cyanobacteria*, which rhythmically release of organic compounds over the course of the day (Hörnlein et al., 2020). Primary production could be another major driver of the rhythmicity in purple bacteria (Piwosz et al., 2020). Further investigations should be conducted using light and other physicochemical parameters variation in order to clarify the role of the Kai proteins in *Rhodospirillum rubrum*.

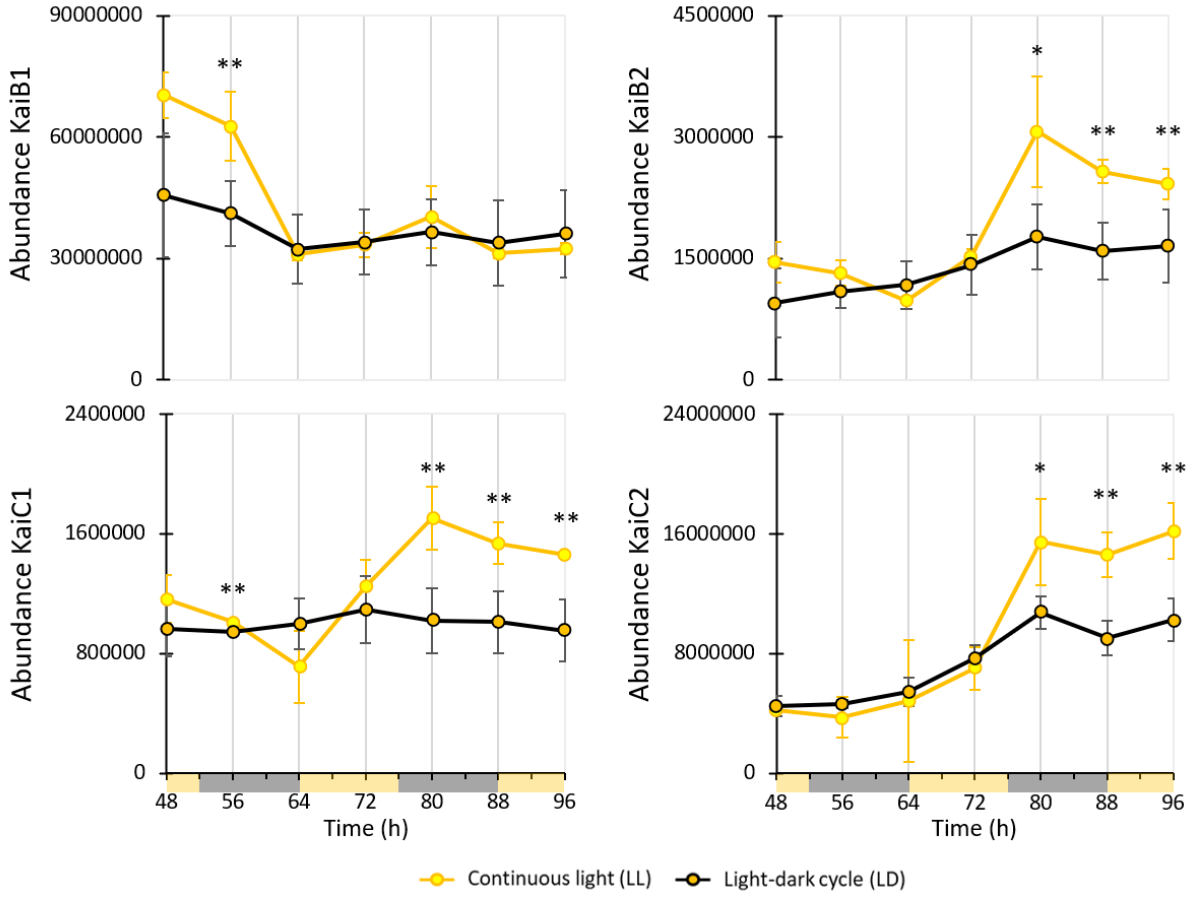


Figure 6.6: Evolution of the relative abundance of *kai* mRNAs in *R. rubrum* depending on light conditions. Light and dark phases in the LD condition are represented on the time axis in yellow and grey, respectively. Significative differences between LL and LD samples are shown with a * (p value ≤ 0.1) or ** (p value ≤ 0.05) and were calculated with a t-test.

4. Conclusion

Deciphering the diel cycle impact on purple bacteria will greatly contribute to understand their ecology, to optimize their use in biotechnology, and to shed light on the origin and evolution of prokaryotic clock mechanisms. Here, we reported for the first time a comprehensive molecular study of the impact of diel cycle in the photoheterotroph *R. rubrum*, previously reported as a potential candidate for circadian rhythm, harboring *kai* genes. We demonstrated that light conditions impacted protein expression, especially in the early growth stage. Numerous proteins involved in transcriptional regulation, energy conversion, stress response, and to a lesser extent, in motility, viral defense and chemotaxis were differently regulated in continuous light exposure compared to light and dark cycle. Light was also found to affect the expression of *kai* genes at both the transcriptional and translational levels, which may be involved in a time-keeping mechanism that yet remain to be elucidated. Further studies should measure the diel transcriptional cycle of *R. rubrum* during continuous culture in chemostat.

Chapter 7: General discussion

1. Diel cycle and circadian clocks in Prokaryotes

With the development of omics approaches, and more specifically the advances in environmental metatranscriptomic and metaproteomics, we are now able to expand our understanding of microbial community dynamics under fluctuating environmental parameters such as diel cycle (Matallana-Surget et al., 2018). In this thesis, we used the metaproteomic approaches to explore the diel functional rhythmicity of oligotrophic marine picoplanktonic communities (**Chapter 4**). In addition, we focused on the diel cycle response of the photoheterotrophic purple bacteria *Rhodospirillum rubrum* (**Chapters 5 and 6**). These studies provided complementarity information allowing to extend our understanding of the drivers of diel dynamics in complex microbial assemblages.

1.1. Is light the main driver of diel rhythmic activity in heterotrophs?

In their natural environment, marine surface bacteria are embedded within complex communities that includes primary producers and consumers (Worden et al., 2015). *Cyanobacteria* are major contributor to primary production and their physiological responses to diel cycle was extensively studied in laboratory. *Cyanobacteria* are known to exhibit diel cycles directly in response to solar radiations and possess circadian clock that control the expression of genes involved in photosynthesis, growth and cell division (Cohen & Golden, 2015). In fact, their clock synchronizes their physiology with the diel cycle and allows them to be prepared when light is available (Bell-Pedersen et al., 2005).

In this work, our metaproteomic analysis revealed a strong diel response of the cyanobacterial *Synechococcales* in the expression of proteins involved in energy conversion

processes such as photosynthesis (**Chapter 4**). This confirmed previous observation of diel oscillation of transcripts involved in light-driven processes (Poretsky et al., 2009; Gilbert et al., 2010; Ottesen et al., 2014). Although the rhythmicity observed at the translational level was less important than that at the transcriptional level, it is coherent with laboratory-based observations. For instance, while *Synechococcus* exhibit 30% to 60% of diel cyclic mRNA (Ito et al., 2009; Vijayan et al., 2009), only 14% of its proteome is rhythmic under light and dark cycle (Guerreiro et al., 2014).

The diel dynamics of photoautotrophs can have structural effects on the heterotrophic bacteria (Morimoto et al., 2020). In our diel metaproteomic study, we observed diel rhythmic abundance of different heterotrophic groups such as *Pelagibacterales* or *Flavobacteriales* in total protein abundance (**Chapter 4**). This could be a direct effect of the diel rhythmicity of *Synechococcales*. Indeed, the rhythmic release of organic matter by photoautotrophs is suggested to influence the expression of transcripts in heterotrophic organisms that can utilize them (Ottesen et al., 2014; Tsai et al., 2012; Aylward et al., 2015; Frischkorn, 2018; McCarren et al., 2010; Straub et al., 2011). In addition, the release of organic matter due to viral lysis and zooplanktonic predation also display a diel pattern, which can indirectly structure the heterotrophic community (Aylward et al., 2017; Kimura et al., 2012; Liu, et al., 2019; Welkie et al., 2019; Yoshida et al., 2018; Deng et al., 2020). Although we observed only few proteins involved in viral infection/defense in our diel metaproteomic study (**Chapter 4**), we detected more of them in our targeted study of *R. rubrum* (**Chapter 6**). Diel rhythmicity of viral infection of *Cyanobacteria* such as *Prochlorococcus* and *Synechococcus* is much more documented than for heterotrophic bacteria (Morimoto et al., 2020). As mentioned before, viruses play a crucial role in diel DOM dynamics (Aylward et al., 2017; Morimoto et al., 2019), which impacts heterotrophs composition and gene expression (Zhao et al., 2019; Fang et al., 2019).

In our targeted study of the diel response of *R. rubrum*, we also observed that 34 proteins involved in its energy metabolism (photosynthesis and oxidative phosphorylation) were differentially regulated depending on the light condition, although we were unable to detect any specific patterns (**Chapter 6**). This indicate that light might not be the sole driver of the diel rhythmic activity in purple bacteria. For example, in *Rhodobacter sphaeroides*, oxygen was found to control the period of circadian oscillations of gene expression (circadian under aerobic condition and ultradian under anaerobic conditions) (Min et al., 2005). In addition, the composition of the culture medium might also influence the rhythmic activity of heterotrophs. Here, *R. rubrum* was cultivated in nutrient rich culture medium, which could have dampened potential diel dynamics that could be driven by daily oscillation of nutrient concentration in natural conditions (Morimoto et al., 2020). Alternatively, further studies could cultivate *R. rubrum* in nutrient poor medium supplemented with daily nutrient inputs. Although it is challenging due to experimental limitations in laboratory, another interesting approach would be to perform diel cycle co-culture experiments combining a clock-controlled *Cyanobacteria* and potential clock-controlled photoheterotrophs and/or heterotrophs. These interspecies models have already provided valuable insights in the phytoplankton-bacterial interactions (Beliaev et al., 2014; Christie-Oleza et al., 2017; Becker et al., 2019). In both experimental designs, stable isotope substrates could be used as tracers to determine temporal patterns in metabolic activity within heterotrophs (i.e., axenic cultures supplemented with daily nutrient input) and between photoautotrophs and heterotrophs (i.e., in co-cultures to assess the impact of primary production on heterotrophs).

1.2. New clock-controlled prokaryotes: how and where to look?

Single-species physiological experimentation in *Cyanobacteria* provided valuable microbial physiology data that allowed the characterization of the sole known prokaryotic

circadian clock mechanism based on Kai proteins (Huang et al., 1990; Chen et al., 1991). In this work, we attempted to characterized new potential clock-controlled candidates based on the distribution of Kai homologs in both Archaea and Bacteria (**Chapter 5**). However, the presence of Kai proteins might not be the sole strategy to characterize new clock mechanisms in Prokaryotes. Indeed, circadian rhythms were recently evidenced in two species that do not possess Kai homologs: the gastrointestinal *Klebsiella aerogenes* (Paulose et al., 2019) and the rhizobacteria *Bacillus subtilis* (Eelderink-Chen et al., 2021). These two organisms are associated with clock-controlled organisms that might play a role in the expression of bacterial circadian rhythms (Paulose et al., 2019; Hörnlein et al., 2020; Eelderink-Chen et al., 2021). Although it remains to elucidate how non-Kai system might work, these observations reinforce the importance of combining top-down and bottom-up approaches when it comes to decipher new clock-controlled Prokaryotes (Röling and van Bodegom, 2014, Morimoto et al., 2020).

1.3. Light-induced stress phenotype

Interestingly, stress response proteins (i.e., light, temperature, or oxidative stresses) were differentially regulated in both the diel metaproteomic analysis of the marine picoplanktonic communities and the targeted study of *R. rubrum* (**Chapters 4 and 6**). Although monocultures grown in laboratory might not be truly representative of what happens in a complex microbial assemblage, the consistency in the observation of a diel impact in the expression of stress response proteins is questioning. Our results suggest that the stress regulation pathways could be time-gated in some bacterial orders such as *Rhodobacterales* or *Rhodospirillales* or constitutive in others like *Pelagibacterales*. The fact that proteins involved in the stress response were detected among the small fraction of rhythmic proteins identified in laboratory grown cultures of *Synechococcus elongatus* supports our observations (Guerreiro et al., 2014). Further studies focusing on the diel regulation of stress response might lead to the

detection of internal molecular timing mechanisms. Additionally, it might also contribute to the improvement of bacterial culture maintenance in both the research and industrial fields by reducing the light-induced stress and achieve more qualitative or quantitative bacterial production (Sartor et al., 2019).

1.4. Recommendations for diel cycle studies

Chronobiology studies focusing on the diel cycle are challenging because of the temporality of the experimental design (i.e., overlapping with the circadian cycle of the experimenter). In addition, in the environment, the stimuli triggering cyclic activity are not always known and if so, not easily controllable.

Although the schedule sampling must be adapted according to human and material resource availability, I recommend measuring biological information during at least three consecutive light and dark periods for optimal coverage and reliable statistical analysis of diel rhythmicity. As sampling is usually limited in environmental studies, I recommend to sample at midday and at night for a better representation of the diel response according to light cycle (**Chapter 6**). Diel cycle laboratory-based studies are usually more flexible and long time series can be afforded, which is useful when it comes to analyze the rhythmicity of biological data. For this purpose, interactive workflow such as the software package DiscoRhythm can be used to estimate cyclical characteristics (e.g., period, phase, amplitude, and statistical significance) in large temporal datasets (Carlucci et al., 2020). However, achieving a long time series observation in fast growing culture can be challenging because the different growth phases usually overlap with the light and dark periods. In our analysis of the diel cycle impact in *R. rubrum*, the gene expression was measured in batch culture within flasks and the proteomes were found to be mainly driven by the growth phase rather than the light conditions (**Chapter 6**). Therefore, I would recommend studying diel rhythmicity in continuous culture using

chemostat. Although reaching stability in continuous bacterial culture can be challenging due to the risk of contamination, and require accurate control of growth parameters, further studies on *R. rubrum* could beneficiate from all the knowledge accumulated in the biotechnological field, where this bacterium is commonly grown in bioreactors (Dadak et al., 2016)

Being able to test the sustainability of wide-community rhythms in the natural environment would provide valuable information on the functioning of potential new endogenous time-keeping mechanisms. However, it is nearly impossible to study rhythmic activity of pelagic microbial communities under constant light or dark conditions without risking complex unknown effects on the studied environment. A solution would be to study diel cycle and circadian rhythms properties of picoplanktonic microbial communities in mesocosms (Crossland & La Point, 1992). This would have the double advantage to allow to easily control light availability and to consider the structural effect of higher trophic level (e.g., zooplankton) on the rhythmic activity of bacterial communities (Deng et al., 2020).

2. Marine metaproteomics: challenges and solutions

Although metaproteomic greatly facilitated the linkages between genotype and phenotype in complex microbial assemblages, this molecular approach is still a growing discipline, facing multiple challenges still hampered by methodological challenges (**Chapters 2 and 3**). Metaproteomics workflow in marine ecosystems is still not standardized and critical steps affect downstream proteins extraction, identification, and subsequent biological conclusions (**Chapter 3**).

2.1. Biomass fractionation

Oligotrophic marine picoplanktonic communities are highly divers and are composed of picoeukaryotes and free-living and particle-attached bacteria. In this thesis, we reported the

different impact of diel cycle on both free-living and particles-attached bacterial communities (**Chapter 5**). These results support the fact that the fractionation of picoplankton is essential to characterize their role in the environment. As oligotrophic waters are characterized by a low biomass concentration, large volumes of seawater optimization steps during samples preparation are usually required to recover sufficient protein concentration for mass spectrometry (Matallana-Surget *et al*., 2018; Saito et al., 2019). Sequential filtration using different pore-size filter is efficient to separate and concentrate the different microbial fractions. Yet, there is no consensus on the methodology for free-living and particle-attached bacteria fractionation. Pre-filtration is not always used, and the pore-size filters used to retrieve the particle-attached fraction vary strongly depending on the study (e.g., Liu et al., 2019; Zhang et al., 2016). However, in environmental metaproteomic, only the most abundant proteins are efficiently detected by mass spectrometry (Matallana-Surget *et al*., 2018). Therefore, pre-filtration is highly recommended to increase the detection of bacterial proteins otherwise hidden by the presence of microeukaryotes material.

2.2. Creation of protein search databases

Bioinformatical processing of metaproteomic data is challenging due to the complexity of microbial communities (Heyer et al., 2017). To address two of the main issues encountered in the characterization of experimental mass spectra, we developed *mPies* (**Chapter 2**). Building a protein search database can be tricky for new metaproteomic users as it requires specific bioinformatic skills depending on the input data (e.g., protein inference from metagenomic data). Although, the use of a metagenomic template is often recommended (May et al., 2016; Timmins-Schiffman et al., 2017), this data is not always available and requires additional expensive analyses (i.e., shotgun metagenomic sequencing). In this context, *mPies* stands out for bringing a unique “all-in” automatized and customizable workflow allowing the

creation of protein search database from assembled and non-assembled metagenomic data or from a list of taxa, which can be obtained by 16S rRNA analysis. Moreover, *mPies* has a "functional database mode" that allows to subset specific protein functions (e.g., circadian clock proteins, plastic degradation pathways, or antibiotic resistance) from public repositories such as UniProt. These functional databases could be an alternative way to analyses metaproteomic data and might improve the characterization of specific biological processes in a complex community. Last but not least, *mPies* also combines different database and automatically remove all sequence redundancies. This approach was shown to be efficient at integrating the information provided using different individual databases (**Chapters 3 and 5**).

2.3. Diversifying the metaproteomic approach

Given the lack of standardized workflow and the numerous critical steps encountered in metaproteomics, we encourage, when it is possible, to diversify the methodology (**Chapter 3**). Indeed, we reported the impact of the protein fractionation (gel-free and gel-based metaproteomics) and the database (TAX-DB, AM-DB, NAM-DB, and Comb-DB). These differences were found in proteins involved in key processes such as nutrients transport, microbial light metabolism, or oxidative stress response, which can lead to misinterpretation of the functioning of the studied microbial communities (**Chapter 3**). Although it might be challenging to combine multiple approaches when several conditions are studied, mock experiments are strongly recommended to identify the best approach to achieve satisfying protein identification yield and pertinent biological observations (Wang et al., 2017; Matallana-Surget et al., 2018). I strongly recommend the collection of additional environmental parameters that can provide complementary information on the functioning of microbial communities. For example, performing a taxonomic diversity analysis (e. g., 16S RNA sequencing) allow to build a pertinent protein search analysis and provide better protein

identification yield than using complete public repositories such as UniProt (**Chapters 1 and 2**). Contrasting the metaproteomic taxonomic structure with 16S RNA-based diversity also allows to better distinguish the community members that are present from those that are metabolically active (**Chapter 5**). Furthermore, the collection of physicochemical parameters such as the nutrient concentration or the light intensity can help to interpret the functions expressed by the microbial communities (Williams & Cavicchioli, 2014) (**Chapters 3 and 5**). Finally, integrating temporality into further microbial community studies is highly recommended as little is known about the functioning of microorganisms at night. Studying their diel dynamics will undoubtedly lead to a better understanding of their ecological role and of their response to environmental stress (**Chapter 5**).

3. Going further

In order to further complete this work, diel wide-community analysis using metatranscriptomic and metaproteomics *in-situ* analysis during a longer time period (i.e., three days minimum at midnight and midday) would help to confirm the marine picoplanktonic protein oscillations observed in this work. In addition, co-culture experiments and mesocosm-based studies would undoubtedly provide a simplified pertinent context to further study the diel rhythmicity and the circadian properties of picoplanktonic bacterial members and communities.

For further research in *R. rubrum*, we would advise to measure its diel activity in either total anoxic or oxic conditions in chemostat to reduce the metabolic interferences and better assess the impact of oxygen on the expression of *kai* genes. In addition, measure of diel rhythmic activity in nutrient poor medium supplemented with daily input of nutrient would allow to determine the role of metabolites on *R. rubrum* diel activity. Transcriptomic analysis would probably provide a clearer view of the potential rhythmic activity of *R. rubrum* as it is in *Synechococcus*. After the detection of rhythmic activities and their corresponding drivers, it

would be essential to determine the role of the *kai* genes. In this purpose, cultures of *kai* gene knockout mutant strains, which are available for *R. rubrum*, could be compared to wild strains under pertinent cyclic conditions. This will allow to identify the role of each *kai* genes in the diel rhythms of *R. rubrum*. Finally, further investigation under constant conditions after specific diel rhythmic entrainment will provide valuable insights helping to understand the time-keeping mechanisms of *R. rubrum* Kai-based clock oscillator.

Metaproteomic was shown to be a valuable tool to investigate the expression of proteins associated to diel cycle response. However, this technology should keep receiving particular attention to facilitate its utilization by non-expert. Therefore, it is urgent to standardize its workflow. In this context, integrated metaproteomic workflow such as MetaProteomeAnalyzer (Schiebenhoefer et al., 2020) is a promising step toward more accessible metaproteomics data analysis and interpretation. Finally, the development of a user-friendly interface for *mPies* could greatly beneficiate the non-expert users.

4. Concluding remarks

The knowledge acquired during this thesis led us to better estimate the potential new clock-controlled bacteria and archaea of high ecological, medical, and industrial importance. This study also provided insights into the diel cycle impact on the functioning of marine picoplanktonic communities and the promising biotech *Rhodospirillum rubrum*. The optimization of analytic tools such as metaproteomics and the use of top-down and bottom-up approaches to characterize species rhythmic activity will undoubtedly open new perspectives on prokaryotic clock systems and contribute to a better understanding of diel microbial ecology.

References

Abdel-Aziz, S., Elsoud, M. M. A., & Anise, A. A. H. (2017). Microbial Biosynthesis: A Repertory of Vital Natural Products. In *Food Biosynthesis* (pp. 25-54). Academic Press.

Abed, R. M. M., Shanti, M., Muthukrishnan, T., Al-Riyami, Z., Pracejus, B., & Moraetis, D. (2020). The Role of Microbial Mats in the Removal of Hexavalent Chromium and Associated Shifts in Their Bacterial Community Composition. *Frontiers in Microbiology, 11*, 12. https://doi.org/10.3389/fmicb.2020.00012

AbSciex. (2019). *Understanding the Pro Group™ Algorithm*. Retrieved 2019-06-12 from https://sciex.com/Documents/manuals/proteinPilot-ProGroup-Algorithm.pdf

Acinas, S. G., Antón, J., & Rodríguez-Valera, F. (1999). Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Applied and Environmental Microbiology, 65*(2), 514-522. https://doi.org/https://doi.org/10.1128/AEM.65.2.514-522.1999

Akinsemolu, A. A. (2018). The role of microorganisms in achieving the sustainable development goals. *Journal of Cleaner Production, 182*, 139-155. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.02.081

Alizadeh-Sani, M., Hamishehkar, H., Khezerlou, A., Azizi-Lalabadi, M., Azadi, Y., Nattagh-Eshtivani, E., Fasihi, M., Ghavami, A., Aynehchi, A., & Ehsani, A. (2018, Jun). Bioemulsifiers Derived from Microorganisms: Applications in the Drug and Food Industry. *Advanced Pharmaceutical Bulletin, 8*(2), 191-199. https://doi.org/10.15171/apb.2018.023

Aoki, S., & Onai, K. (2009). Circadian Clocks of *Synechocystis* sp. Strain PCC 6803, *Thermosynechococcus elongatus, Prochlorococcus* spp., *Trichodesmium* spp. and Other Species. In J. L. Ditty, S. R. Mackey, & C. H. Johnson (Eds.), *Bacterial Circadian Programs* (pp. 259-282). Springer Berlin Heidelberg.

Asif, M., Pervez, A., & Ahmad, R. (2019). Role of Melatonin and Plant-Growth-Promoting Rhizobacteria in the Growth and Development of Plants. *CLEAN – Soil, Air, Water, 47*(6), 1800459. https://doi.org/https://doi.org/10.1002/clen.201800459

Axmann, I. M., Dühring, U., Seeliger, L., Arnold, A., Vanselow, J. T., Kramer, A., & Wilde, A. (2009). Biochemical evidence for a timing mechanism in *Prochlorococcus*. *Journal of Bacteriology, 191*(17), 5342-5347. https://doi.org/10.1128/jb.00419-09

Axmann, I. M., Hertel, S., Wiegard, A., Dörrich, A. K., & Wilde, A. (2014). Diversity of KaiC-based timing systems in marine *Cyanobacteria*. *Marine Genomics, 14*, 3-16. https://doi.org/https://doi.org/10.1016/j.margen.2013.12.006

Aylward, F. O., Eppley, J. M., Smith, J. M., Chavez, F. P., Scholin, C. A., & DeLong, E. F. (2015). Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proceedings of the National Academy of Sciences, 112*(17), 5443-5448. https://doi.org/doi:10.1073/pnas.1502883112

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., & Thingstad, F. (1983). The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series, 10*(3), 257-263. http://www.jstor.org/stable/24814647

Backer, R., Rokem, J. S., Ilangumaran, G., Lamont, J., Praslickova, D., Ricci, E., Subramanian, S., & Smith, D. L. (2018). Plant Growth-Promoting Rhizobacteria: Context, Mechanisms of Action, and Roadmap to Commercialization of Biostimulants for Sustainable Agriculture. *Frontiers in Plant Science, 9*, 1473. https://doi.org/10.3389/fpls.2018.01473

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America, 115*(25), 6506-6511. https://doi.org/doi:10.1073/pnas.1711842115

Basak, N., & Das, D. (2007). The Prospect of Purple Non-Sulfur (PNS) Photosynthetic Bacteria for Hydrogen Production: The Present State of the Art. *World Journal of Microbiology and Biotechnology, 23*(1), 31-42. https://doi.org/10.1007/s11274-006-9190-9

Bastidas Navarro, M., Balseiro, E., & Modenutti, B. (2009). Effect of UVR on lake water and macrophyte leachates in shallow Andean-Patagonian lakes: Bacterial response to changes in optical features. *Photochemistry and Photobiology, 85*(1), 332-340. https://doi.org/ https://doi.org/10.1111/j.1751-1097.2008.00442.x

Bayon-Vicente, G., Wattiez, R., & Leroy, B. (2020). Global Proteomic Analysis Reveals High Light Intensity Adaptation Strategies and Polyhydroxyalkanoate Production in *Rhodospirillum rubrum* Cultivated With Acetate as Carbon Source. *Frontiers in Microbiology, 11*, 464. https://doi.org/10.3389/fmicb.2020.00464

Beck, C., Hertel, S., Rediger, A., Lehmann, R., Wiegard, A., Kölsch, A., Heilmann, B., Georg, J., Hess, W. R., & Axmann, I. M. (2014). Daily expression pattern of protein-encoding genes and small noncoding RNAs in *Synechocystis* sp. strain PCC 6803. *Applied and Environmental Microbiology, 80*(17), 5195-5206. https://doi.org/10.1128/aem.01086-14

Becker, J. W., Hogle, S. L., Rosendo, K., & Chisholm, S. W. (2019). Co-culture and biogeography of *Prochlorococcus* and SAR11. *The ISME Journal, 13*(6), 1506-1519. https://doi.org/10.1038/s41396-019-0365-4

Beckmann, P., & Spizzichino, A. (1987). The scattering of electromagnetic waves from rough surfaces. *Norwood*.

Béja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., Jovanovich, S. B., Gates, C. M., Feldman, R. A., & Spudich, J. L. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science, 289*(5486), 1902-1906. https://doi.org/10.1126/science.289.5486.1902

Beliaev, A. S., Romine, M. F., Serres, M., Bernstein, H. C., Linggi, B. E., Markillie, L. M., Isern, N. G., Chrisler, W. B., Kucek, L. A., Hill, E. A., Pinchuk, G. E., Bryant, D. A., Steven Wiley, H., Fredrickson, J. K., & Konopka, A. (2014). Inference of interactions in cyanobacterial–heterotrophic co-cultures via transcriptome sequencing. *The ISME Journal, 8*(11), 2243-2255. https://doi.org/10.1038/ismej.2014.69

Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., & Zoran, M. J. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics, 6*(7), 544-556. https://doi.org/10.1038/nrg1633

Bergauer, K., Fernandez-Guerra, A., Garcia, J. A. L., Sprenger, R. R., Stepanauskas, R., Pachiadaki, M. G., Jensen, O. N., & Herndl, G. J. (2018). Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences of the United States of America, 115*(3), E400-E408. https://doi.org/doi:10.1073/pnas.1708779115

Berleman, J. E., & Bauer, C. E. (2005). A che-like signal transduction cascade involved in controlling flagella biosynthesis in *Rhodospirillum centenum*. *Molecular Microbiology, 55*(5), 1390-1402. https://doi.org/10.1111/j.1365-2958.2005.04489.x

Bhadra, U., Thakkar, N., Das, P., & Pal Bhadra, M. (2017). Evolution of circadian rhythms: from bacteria to human. *Sleep Medicine, 35*, 49-61. https://doi.org/https://doi.org/10.1016/j.sleep.2017.04.008

Biller, S. J., Berube, P. M., Lindell, D., & Chisholm, S. W. (2015). *Prochlorococcus*: the structure and function of collective diversity. *Nature Reviews Microbiology, 13*(1), 13-27. https://doi.org/10.1038/nrmicro3378

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. https://doi.org/https://doi.org/10.1093/bioinformatics/btu170

Bolhuis, H., Cretoiu, M. S., & Stal, L. J. (2014). Molecular ecology of microbial mats. *FEMS Microbiology Ecology, 90*(2), 335-350. https://doi.org/10.1111/1574-6941.12408

Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans studies plankton at planetary scale. *Science, 348*(6237), 873-873. https://doi.org/10.1126/science.aac5605

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry, 72*, 248-254. https://doi.org/10.1006/abio.1976.9999

Breitbart, M., Bonnain, C., Malki, K., & Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nature microbiology, 3*(7), 754-766. https://doi.org/10.1038/s41564-018-0166-y

Breslauer, K. J., Frank, R., Blöcker, H., & Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America, 83*(11), 3746-3750. https://doi.org/10.1073/pnas.83.11.3746

Bryant, D. A., & Frigaard, N.-U. (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends in Microbiology, 14*(11), 488-496. https://doi.org/10.1016/j.tim.2006.09.001

Bryson, S., Li, Z., Pett-Ridge, J., Hettich, R. L., Mayali, X., Pan, C., & Mueller, R. S. (2016). Proteomic Stable Isotope Probing Reveals Taxonomically Distinct Patterns in Amino Acid Assimilation by Coastal Marine Bacterioplankton. *mSystems, 1*(2). https://doi.org/10.1128/mSystems.00027-15

Buchan, A., González José, M., & Moran Mary, A. (2005). Overview of the Marine *Roseobacter* Lineage. *Applied and Environmental Microbiology, 71*(10), 5665-5677. https://doi.org/10.1128/AEM.71.10.5665-5677.2005

Buchan, A., LeCleir, G. R., Gulvik, C. A., & González, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology, 12*(10), 686-698. https://doi.org/https://doi.org/10.1038/nrmicro3326

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods, 12*(1), 59-60. https://doi.org/https://doi.org/10.1038/nmeth.3176

Button, D., & Robertson, B. (2000). Effect of nutrient kinetics and cytoarchitecture on bacterioplankter size. *Limnology and Oceanography, 45*(2), 499-505. https://doi.org/https://doi.org/10.4319/lo.2000.45.2.0499

Cantarel, B. L., Erickson, A. R., VerBerkmoes, N. C., Erickson, B. K., Carey, P. A., Pan, C., Shah, M., Mongodin, E. F., Jansson, J. K., & Fraser-Liggett, C. M. (2011). Strategies for metagenomic-guided whole-community proteomics of complex microbial environments. *PLOS ONE, 6*(11), e27173. https://doi.org/https://doi.org/10.1371/journal.pone.0027173

Caplice, E., & Fitzgerald, G. F. (1999). Food fermentations: role of microorganisms in food production and preservation. *International Journal of Food Microbiology, 50*(1-2), 131-149. https://doi.org/10.1016/s0168-1605(99)00082-3

Carlucci, M., Kriščiūnas, A., Li, H., Gibas, P., Koncevičius, K., Petronis, A., & Oh, G. (2019). DiscoRhythm: an easy-to-use web application and R package for discovering rhythmicity. *Bioinformatics, 36*(6), 1952-1954. https://doi.org/10.1093/bioinformatics/btz834

Chang, Y.-G., Cohen, S. E., Phong, C., Myers, W. K., Kim, Y.-I., Tseng, R., Lin, J., Zhang, L., Boyd, J. S., Lee, Y., Kang, S., Lee, D., Li, S., Britt, R. D., Rust, M. J., Golden, S. S., & LiWang, A. (2015). Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in *Cyanobacteria*. *Science 349*(6245), 324-328. https://doi.org/10.1126/science.1260031

Chang, Y.-G., Tseng, R., Kuo, N.-W., & LiWang, A. (2012). Rhythmic ring-ring stacking drives the circadian oscillator clockwise. *Proceedings of the National Academy of Sciences, 109*(42), 16847-16851. https://doi.org/doi:10.1073/pnas.1211508109

Chen, T. H., Chen, T. L., Hung, L. M., & Huang, T. C. (1991). Circadian Rhythm in Amino Acid Uptake by *Synechococcus* RF-1. *Plant Physiology, 97*(1), 55-59. https://doi.org/10.1104/pp.97.1.55

Chew, J., Leypunskiy, E., Lin, J., Murugan, A., & Rust, M. J. (2018). High protein copy number is required to suppress stochasticity in the cyanobacterial circadian clock. *Nature Communications, 9*(1), 3004. https://doi.org/10.1038/s41467-018-05109-4

Chisholm, S. W., Frankel, S. L., Goericke, R., Olson, R. J., Palenik, B., Waterbury, J. B., West-Johnsrud, L., & Zettler, E. R. (1992). *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. *Archives of Microbiology, 157*(3), 297-300. https://doi.org/10.1007/BF00245165

Choi, H., Rao, M. C., & Chang, E. B. (2021). Gut microbiota as a transducer of dietary cues to regulate host circadian rhythms and metabolism. *Nature Reviews Gastroenterology & Hepatology, 18*(10), 679-689. https://doi.org/10.1038/s41575-021-00452-2

Christie-Oleza, J. A., Sousoni, D., Lloyd, M., Armengaud, J., & Scanlan, D. J. (2017). Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. *Nature microbiology, 2*(9), 17100. https://doi.org/10.1038/nmicrobiol.2017.100

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., & Wilczynski, B. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics, 25*(11), 1422-1423. https://doi.org/https://doi.org/10.1093/bioinformatics/btp163

Cohen, S. E., & Golden, S. S. (2015). Circadian rhythms in *Cyanobacteria*. *Microbiology and Molecular Biology Reviews, 79*(4), 373-385. https://doi.org/https://doi.org/10.1128/MMBR.00036-15

Cole, J. J., Findlay, S., & Pace, M. L. (1988). Bacterial production in fresh and saltwater ecosystems: a cross-system overview. *Marine ecology progress series. Oldendorf, 43*(1), 1-10. https://doi.org/10.3354/meps043001

Conrad, R. (2009). The global methane cycle: recent advances in understanding the microbial processes involved. *Environmental Microbiology Reports 1*(5), 285-292. https://doi.org/10.1111/j.1758-2229.2009.00038.x

Crick, F. (1970). Central dogma of molecular biology. *Nature, 227*(5258), 561-563. https://doi.org/10.1038/227561a0

Crossland, N. O., & La Point, T. W. (1992). The design of mesocosm experiments. *Environmental Toxicology and Chemistry: An International Journal*, 11(1), 1-4. https://doi.org/10.1002/etc.5620110101

Crump, B. C., Armbrust, E. V., & Baross, J. A. (1999). Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Applied and Environmental Microbiology, 65*(7), 3192-3204. https://doi.org/https://doi.org/10.1128/AEM.65.7.3192-3204.1999

Curtis, A. M., Bellet, M. M., Sassone-Corsi, P., & O'Neill, L. A. (2014). Circadian clock proteins and immunity. *Immunity, 40*(2), 178-186. https://doi.org/10.1016/j.immuni.2014.02.002

Dadak, A., Aghbashlo, M., Tabatabaei, M., Younesi, H., & Najafpour, G. (2016). Exergy-based sustainability assessment of continuous photobiological hydrogen production using anaerobic bacterium *Rhodospirillum rubrum*. *Journal of Cleaner Production*, 139, 157-166. https://doi.org/10.1016/j.jclepro.2016.08.020

De Maio, A. (1999). Heat shock proteins: facts, thoughts, and dreams. *Shock, 11*(1), 1-12. https://doi.org/10.1097/00024382-199901000-00001

De Meur, Q., Deutschbauer, A., Koch, M., Bayon-Vicente, G., Cabecas Segura, P., Wattiez, R., & Leroy, B. (2020). New perspectives on butyrate assimilation in *Rhodospirillum rubrum* S1H under photoheterotrophic conditions. *BMC Microbiology, 20*(1), 126. https://doi.org/10.1186/s12866-020-01814-7

Deaver, J. A., Eum, S. Y., & Toborek, M. (2018). Circadian Disruption Changes Gut Microbiome Taxa and Functional Gene Composition. *Frontiers in Microbiology, 9*, 737. https://doi.org/10.3389/fmicb.2018.00737

DeLong, E. F. (2005). Microbial community genomics in the ocean. *Nature Reviews Microbiology, 3*(6), 459-469. https://doi.org/10.1038/nrmicro1158

DeLong, E. F., Franks, D. G., & Alldredge, A. L. (1993). Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnology and Oceanography, 38*(5), 924-934. https://doi.org/ https://doi.org/10.4319/lo.1993.38.5.0924

Demain, A. L., & Sanchez, S. (2009). Microbial drug discovery: 80 years of progress. *The Journal of Antibiotics, 62*(1), 5-16. https://doi.org/10.1038/ja.2008.16

Deming, J. W., & Baross, J. A. (2000). Survival, dormancy, and nonculturable cells in extreme deep-sea environments. In *Nonculturable microorganisms in the environment* (pp. 147-197). Springer.

Deusch, S., Camarinha-Silva, A., Conrad, J., Beifuss, U., Rodehutscord, M., & Seifert, J. (2017). A structural and functional elucidation of the rumen microbiome influenced by various diets and microenvironments. *Frontiers in Microbiology, 8*, 1605. https://doi.org/https://doi.org/10.3389/fmicb.2017.01605

Ditty, J. L., Mackey, S. R., & Johnson, C. H. (2009). *Bacterial circadian programs*. Springer Science & Business Media.

Dong, G., Yang, Q., Wang, Q., Kim, Y. I., Wood, T. L., Osteryoung, K. W., van Oudenaarden, A., & Golden, S. S. (2010). Elevated ATPase activity of KaiC applies a circadian checkpoint on cell division in *Synechococcus elongatus*. *Cell, 140*(4), 529-539. https://doi.org/10.1016/j.cell.2009.12.042

Dong, H.-P., Hong, Y.-G., Lu, S., & Xie, L.-Y. (2014). Metaproteomics reveals the major microbial players and their biogeochemical functions in a productive coastal system in the northern South China Sea. *Environmental microbiology reports, 6*(6), 683-695. https://doi.org/https://doi.org/10.1111/1758-2229.12188

Dörrich, A. K., Mitschke, J., Siadat, O., & Wilde, A. (2014). Deletion of the *Synechocystis* sp. PCC 6803 kaiAB1C1 gene cluster causes impaired cell growth under light–dark conditions. *Microbiology, 160*(11), 2538-2550. https://doi.org/https://doi.org/10.1099/mic.0.081695-0

Dufresne, A., Garczarek, L., & Partensky, F. (2005). Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome biology, 6*(2), 1-10. https://doi.org/https://doi.org/10.1186/gb-2005-6-2-r14

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., Duprat, S., Galperin, M. Y., Koonin, E. V., & Le Gall, F. (2003). Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences of the United States of America, 100*(17), 10020-10025. https://doi.org/https://doi.org/10.1073/pnas.1733211100

Dvornyk, V., Vinogradova, O., & Nevo, E. (2003). Origin and evolution of circadian clock genes in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America, 100*(5), 2495-2500. https://doi.org/10.1073/pnas.0130099100

Edgar, R. S., Green, E. W., Zhao, Y., van Ooijen, G., Olmedo, M., Qin, X., Xu, Y., Pan, M., Valekunja, U. K., Feeney, K. A., Maywood, E. S., Hastings, M. H., Baliga, N. S., Merrow, M., Millar, A. J., Johnson, C. H., Kyriacou, C. P., O'Neill, J. S., & Reddy, A. B. (2012). Peroxiredoxins are conserved markers of circadian rhythms. *Nature, 485*(7399), 459-464. https://doi.org/10.1038/nature11088

Edmunds Jr, L. N. (1983). Chronobiology at the cellular and molecular levels: models and mechanisms for circadian timekeeping. *American journal of anatomy, 168*(4), 389-431. https://doi.org/ 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

Emerson, S., & Hedges, J. (2008). *Chemical oceanography and the marine carbon cycle*. Cambridge University Press.

Enzmann, F., Mayer, F., Rother, M., & Holtmann, D. (2018). Methanogens: biochemical background and biotechnological applications. *AMB Express, 8*(1), 1. https://doi.org/10.1186/s13568-017-0531-x

Espinosa, J., Boyd, J. S., Cantos, R., Salinas, P., Golden, S. S., & Contreras, A. (2015). Cross-talk and regulatory interactions between the essential response regulator RpaB and cyanobacterial circadian clock output. *Proceedings of the National Academy of Sciences of the United States of America, 112*(7), 2198-2203. https://doi.org/10.1073/pnas.1424632112

Esterházy, D., King, M. S., Yakovlev, G., & Hirst, J. (2008). Production of reactive oxygen species by complex I (NADH:ubiquinone oxidoreductase) from *Escherichia coli* and comparison to the enzyme from mitochondria. *Biochemistry, 47*(12), 3964-3971. https://doi.org/10.1021/bi702243b

Faust, K., & Raes, J. (2012). Microbial interactions: from networks to models. *Nature Reviews Microbiology, 10*(8), 538-550. https://doi.org/10.1038/nrmicro2832

Fecskeová, L. K., Piwosz, K., Hanusová, M., Nedoma, J., Znachor, P., & Koblížek, M. (2019). Diel changes and diversity of pufM expression in freshwater communities of anoxygenic phototrophic bacteria. *Scientific Reports, 9*(1), 18766. https://doi.org/10.1038/s41598-019-55210-x

Feingersch, R., Suzuki, M. T., Shmoish, M., Sharon, I., Sabehi, G., Partensky, F., & Béjà, O. (2010). Microbial community genomics in eastern Mediterranean Sea surface waters. *The ISME Journal, 4*(1), 78-87. https://doi.org/https://doi.org/10.1038/ismej.2009.92

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), 237-240. https://doi.org/doi:10.1126/science.281.5374.237

Fitzmaurice, W. P., Saari, L. L., Lowery, R. G., Ludden, P. W., & Roberts, G. P. (1989). Genes coding for the reversible ADP-ribosylation system of dinitrogenase reductase from *Rhodospirillum rubrum*. *Molecular and General Genetics, 218*(2), 340-347. https://doi.org/10.1007/bf00331287

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., & Martiny, A. C. (2013). Present and future global distributions of the marine *Cyanobacteria Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America, 110*(24), 9824-9829. https://doi.org/10.1073/pnas.1307701110

Forchhammer, K. (2007). Glutamine signalling in bacteria. *Frontiers in Bioscience-Landmark, 12*(1), 358-370. https://doi.org/https://doi.org/10.2741/2069

Fourçans, A., Solé, A., Diestra, E., Ranchou-Peyruse, A., Esteve, I., Caumette, P., & Duran, R. (2006). Vertical migration of phototrophic bacterial populations in a hypersaline microbial mat from Salins-de-Giraud (Camargue, France). *FEMS Microbiology Ecology, 57*(3), 367-377. https://doi.org/10.1111/j.1574-6941.2006.00124.x

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Reis, S., Sheppard, L. J., Jenkins, A., Grizzetti, B., Galloway, J. N., Vitousek, P., Leach, A., Bouwman, A. F., Butterbach-Bahl, K., Dentener, F., Stevenson, D., Amann, M., & Voss, M. (2013). The global nitrogen cycle in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*(1621), 20130164. https://doi.org/10.1098/rstb.2013.0164

Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C., & Huttenhower, C. (2015). Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology, 13*(6), 360-372. https://doi.org/10.1038/nrmicro3451

Frischkorn, K. R., Haley, S. T., & Dyhrman, S. T. (2018). Coordinated gene expression between *Trichodesmium* and its microbiome over day–night cycles in the North Pacific Subtropical Gyre. *The ISME Journal, 12*(4), 997-1007. https://doi.org/10.1038/s41396-017-0041-5

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics, 28*(23), 3150-3152. https://doi.org/10.1093/bioinformatics/bts565

Fuhrman, J. A., Cram, J. A., & Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology, 13*(3), 133-146. https://doi.org/10.1038/nrmicro3417

Galí, M., Simó, R., Vila-Costa, M., Ruiz-González, C., Gasol, J. M., & Matrai, P. (2013). Diel patterns of oceanic dimethylsulfide (DMS) cycling: Microbial and physical drivers. *Global Biogeochemical Cycles, 27*(3), 620-636. https://doi.org/https://doi.org/10.1002/gbc.20047

Gallois, N., Alpha-Bazin, B., Ortet, P., Barakat, M., Piette, L., Long, J., Berthomieu, C., Armengaud, J., & Chapon, V. (2018). Proteogenomic insights into uranium tolerance of a Chernobyl's Microbacterium bacterial isolate. *Journal of Proteomics, 177*, 148-157. https://doi.org/https://doi.org/10.1016/j.jprot.2017.11.021

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research, 43*(Database issue), D261-269. https://doi.org/10.1093/nar/gku1223

Galperin, M. Y., Rigden, D. J., & Fernández-Suárez, X. M. (2015). The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection. *Nucleic Acids Research, 43*(D1), D1-D5. https://doi.org/10.1093/nar/gku1241

Garcia-Pichel, F., Mechling, M., & Castenholz, R. W. (1994). Diel Migrations of Microorganisms within a Benthic, Hypersaline Mat Community. *Applied and Environmental Microbiology, 60*(5), 1500-1511. https://doi.org/10.1128/aem.60.5.1500-1511.1994

Gasol, J. M., Doval, M. D., Pinhassi, J., Calderón-Paz, J. I., Guixa-Boixareu, N., Vaqué, D., & Pedrós-Alió, C. (1998). Diel variations in bacterial heterotrophic activity and growth in the northwestern Mediterranean Sea. *Marine Ecology Progress Series, 164*, 107-124. https://doi.org/10.3354/meps164107

George, D. M., Vincent, A. S., & Mackey, H. R. (2020). An overview of anoxygenic phototrophic bacteria and their applications in environmental biotechnology for sustainable Resource recovery. *Biotechnology Reports (Amst), 28*, e00563. https://doi.org/10.1016/j.btre.2020.e00563

Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K., & Walsh, D. A. (2014). Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *The ISME Journal, 8*(6), 1301-1313. https://doi.org/https://doi.org/10.1038/ismej.2013.234

Géron, A., Werner, J., Lebaron, P., Wattiez, R., & Matallana-Surget, S. (2021). Diel Protein Regulation of Marine Picoplanktonic Communities Assessed by Metaproteomics. *Microorganisms, 9*(12), 2621. https://doi.org/https://doi.org/10.3390/microorganisms9122621

Géron, A., Werner, J., Wattiez, R., Lebaron, P., & Matallana-Surget, S. (2019). Deciphering the functioning of microbial communities: shedding light on the critical steps in metaproteomics. *Frontiers in Microbiology, 10*, 2395. https://doi.org/https://doi.org/10.3389/fmicb.2019.02395

Ghiglione, J. F., Mevel, G., Pujo-Pay, M., Mousseau, L., Lebaron, P., & Goutx, M. (2007). Diel and seasonal variations in abundance, activity, and community structure of particle-attached and free-living bacteria in NW Mediterranean Sea. *Microbial Ecology, 54*(2), 217-231. https://doi.org/10.1007/s00248-006-9189-7

Gilbert, J. A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., & Joint, I. (2010). The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic'study of seasonal and diel temporal variation. *PLOS ONE, 5*(11), e15545. https://doi.org/https://doi.org/10.1371/journal.pone.0015545

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappé, M. S., Short, J. M., Carrington, J. C., & Mathur, E. J. (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science, 309*(5738), 1242-1245. https://doi.org/doi:10.1126/science.1114057

Gòdia, F., Albiol, J., Montesinos, J. L., Pérez, J., Creus, N., Cabello, F., Mengual, X., Montras, A., & Lasseur, C. (2002). MELISSA: a loop of interconnected bioreactors to develop life support in Space. *Journal of Biotechnology, 99*(3), 319-330. https://doi.org/https://doi.org/10.1016/S0168-1656(02)00222-5

Godinho-Silva, C., Domingues, R. G., Rendas, M., Raposo, B., Ribeiro, H., da Silva, J. A., Vieira, A., Costa, R. M., Barbosa-Morais, N. L., Carvalho, T., & Veiga-Fernandes, H. (2019). Light-entrained and brain-tuned circadian circuits regulate ILC3s and gut homeostasis. *Nature, 574*(7777), 254-258. https://doi.org/10.1038/s41586-019-1579-3

Goericke, R., & Repeta, D. J. (1992). The pigments of *Prochlorococcus marinus*: The presence of divinylchlorophyll a and b in a marine procaryote. *Limnology and Oceanography, 37*(2), 425-433. https://doi.org/https://doi.org/10.4319/lo.1992.37.2.0425

Golden, S. S., & Canales, S. R. (2003). Cyanobacterial circadian clocks--timing is everything. *Nature Reviews: Microbiology, 1*(3), 191-199. https://doi.org/10.1038/nrmicro774

Gonzalez, G., Gilles-Gonzalez, M. A., Rybak-Akimova, E. V., Buchalova, M., & Busch, D. H. (1998). Mechanisms of autoxidation of the oxygen sensor FixL and Aplysia myoglobin: implications for oxygen-binding heme proteins. *Biochemistry, 37*(28), 10188-10194. https://doi.org/10.1021/bi980529x

Gould, P. D., Domijan, M., Greenwood, M., Tokuda, I. T., Rees, H., Kozma-Bognar, L., Hall, A. J., & Locke, J. C. (2018). Coordination of robust single cell rhythms in the Arabidopsis circadian clock via spatial waves of gene expression. *Elife, 7*. https://doi.org/10.7554/eLife.31700

Grossart, H.-P. (2010). Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. *Environmental microbiology reports, 2*(6), 706-714. https://doi.org/https://doi.org/10.1111/j.1758-2229.2010.00179.x

Grossart, H.-P., Riemann, L., & Azam, F. (2001). Bacterial motility in the sea and its ecological Implications. *Aquatic Microbial Ecology 25*, 247-258. https://doi.org/10.3354/ame025247

Grüning, B., Dale, R., Sjödin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., Valieris, R., & Köster, J. (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nature Methods, 15*(7), 475-476. https://doi.org/10.1038/s41592-018-0046-7

Guerreiro, A. C. L., Benevento, M., Lehmann, R., van Breukelen, B., Post, H., Giansanti, P., Maarten Altelaar, A. F., Axmann, I. M., & Heck, A. J. R. (2014). Daily rhythms in the cyanobacterium *Synechococcus elongatus* probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Molecular & Cellular Proteomics, 13*(8), 2042-2055. https://doi.org/10.1074/mcp.m113.035840

Hanada, S. (2016). Anoxygenic Photosynthesis —A Photochemical Reaction That Does Not Contribute to Oxygen Reproduction—. *Microbes and Environments, 31*(1), 1-3. https://doi.org/10.1264/jsme2.ME3101rh

Hawley, A. K., Brewer, H. M., Norbeck, A. D., Paša-Tolić, L., & Hallam, S. J. (2014). Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proceedings of the National Academy of Sciences, 111*(31), 11395-11400. https://doi.org/https://doi.org/10.1073/pnas.1322132111

Hedges, J., & Oades, J. (1997). Comparative organic geochemistries of soils and marine sediments. *Organic Geochemistry, 27*(7-8), 319-361. https://doi.org/https://doi.org/10.1016/S0146-6380(97)00056-9

Heinemann, M., Ratiner, K., & Elinav, E. (2021). Basic Biology of Rhythms and the Microbiome. In *Circadian Rhythms in Bacteria and Microbiomes* (pp. 317-328). Springer, Cham.

Hellweger, F. L., Jabbur, M. L., Johnson, C. H., van Sebille, E., & Sasaki, H. (2020). Circadian clock helps cyanobacteria manage energy in coastal and high latitude ocean. *The ISME Journal, 14*(2), 560-568. https://doi.org/10.1038/s41396-019-0547-0

Henneberger, R., Cheema, S., Folini, M., & Zeyer, J. (2017). Diurnal Patterns of Greenhouse Gas Fluxes in a Swiss Alpine Fen. *Wetlands, 37*(2), 193-204. https://doi.org/10.1007/s13157-016-0844-6

Herbst, F. A., Lünsmann, V., Kjeldal, H., Jehmlich, N., Tholey, A., von Bergen, M., Nielsen, J. L., Hettich, R. L., Seifert, J., & Nielsen, P. H. (2016). Enhancing metaproteomics--The value of models and defined environmental microbial systems. *Proteomics, 16*(5), 783-798. https://doi.org/10.1002/pmic.201500305

Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G., & Benndorf, D. (2017). Challenges and perspectives of metaproteomic data analysis. *Journal of Biotechnology, 261*, 24-36. https://doi.org/https://doi.org/10.1016/j.jbiotec.2017.06.1201

Hoch, M. P., Snyder, R. A., Jeffrey, W. H., Dillon, K. S., & Coffin, R. B. (2006). Expression of glutamine synthetase and glutamate dehydrogenase by marine bacterioplankton: Assay optimizations and efficacy for assessing nitrogen to carbon metabolic balance *in situ*. *Limnology and Oceanography Methods, 4*(9), 308-328. https://doi.org/https://doi.org/10.4319/lom.2006.4.308

Hohmann-Marriott, M. F., & Blankenship, R. E. (2011). Evolution of Photosynthesis. *Annual Review of Plant Biology, 62*(1), 515-548. https://doi.org/10.1146/annurev-arplant-042110-103811

Hollibaugh, J. T., Wong, P. S., & Murrell, M. C. (2000). Similarity of particle-associated and free-living bacterial communities in northern San Francisco Bay, California. *Aquatic Microbial Ecology, 21*(2), 103-114. https://doi.org/10.3354/ame021103

Holtzendorff, J., Partensky, F., Mella, D., Lennon, J. F., Hess, W. R., & Garczarek, L. (2008). Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium *Prochlorococcus marinus* PCC 9511. *Journal of Biological Rhythms, 23*(3), 187-199. https://doi.org/10.1177/0748730408316040

Hörnlein, C., Confurius-Guns, V., Grego, M., Stal, L. J., & Bolhuis, H. (2020). Circadian clock-controlled gene expression in co-cultured, mat-forming *Cyanobacteria*. *Scientific Reports, 10*(1), 14095. https://doi.org/10.1038/s41598-020-69294-3

Huang, R.-C. (2018). The discoveries of molecular mechanisms for the circadian rhythm: The 2017 Nobel Prize in Physiology or Medicine. *Biomedical Journal, 41*(1), 5-8. https://doi.org/https://doi.org/10.1016/j.bj.2018.02.003

Huang, T.-C., & Lin, R.-F. (2009). Circadian Rhythm of *Cyanothece* RF-1(*Synechococcus* RF-1). In J. L. Ditty, S. R. Mackey, & C. H. Johnson (Eds.), *Bacterial Circadian Programs* (pp. 39-61). Springer Berlin Heidelberg.

Huang, T.-C., Tu, J., Chow, T.-J., & Chen, T.-H. (1990). Circadian Rhythm of the Prokaryote *Synechococcus* sp. RF-1. *Plant Physiology, 92*(2), 531-533. https://doi.org/10.1104/pp.92.2.531

Hubbard, C. J., Brock, M. T., van Diepen, L. T., Maignien, L., Ewers, B. E., & Weinig, C. (2018). The plant circadian clock influences rhizosphere community structure and function. *Isme journal, 12*(2), 400-410. https://doi.org/10.1038/ismej.2017.172

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., & Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology, 12*(6), e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics, 11*(1), 1-11. https://doi.org/https://doi.org/10.1186/1471-2105-11-119

Imhoff, J. F. (2008). Systematics of Anoxygenic Phototrophic Bacteria. In R. Hell, C. Dahl, D. Knaff, & T. Leustek (Eds.), *Sulfur Metabolism in Phototrophic Organisms* (pp. 269-287). Springer Netherlands.

Iqbal, J., Nelson, J. A., & McCulley, R. L. (2013). Fungal endophyte presence and genotype affect plant diversity and soil-to-atmosphere trace gas fluxes. *Plant and Soil, 364*(1), 15-27. https://doi.org/10.1007/s11104-012-1326-0

Ito, H., Mutsuda, M., Murayama, Y., Tomita, J., Hosokawa, N., Terauchi, K., Sugita, C., Sugita, M., Kondo, T., & Iwasaki, H. (2009). Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. *Proceedings of the National Academy of Sciences of the United States of America, 106*(33), 14168-14173. https://doi.org/10.1073/pnas.0902587106

Iwasaki, H., Williams, S. B., Kitayama, Y., Ishiura, M., Golden, S. S., & Kondo, T. (2000). A kaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in *Cyanobacteria*. *Cell, 101*(2), 223-233. https://doi.org/10.1016/s0092-8674(00)80832-6

Jagtap, P., Goslinga, J., Kooren, J. A., McGowan, T., Wroblewski, M. S., Seymour, S. L., & Griffin, T. J. (2013). A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics, 13*(8), 1352-1357. https://doi.org/https://doi.org/10.1002/pmic.201200352

John, J. S. (2011). *SeqPrep: Tool for Stripping Adaptors and/or Merging Paired Reads with Overlap into Single Reads.* Retrieved 2021-11-8 from https://githubcom/jstjohn/SeqPrep

Johnson, C. H., & Rust, M. J. (2021). *Circadian Rhythms in Bacteria and Microbiomes* (Vol. 409). Springer.

Johnson, Z. I., & Lin, Y. (2009). *Prochlorococcus*: Approved for export. *Proceedings of the National Academy of Sciences, 106*(26), 10400-10401. https://doi.org/10.1073/pnas.0905187106

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., & Chisholm, S. W. (2006). Niche Partitioning Among *Prochlorococcus* Ecotypes Along Ocean-Scale Environmental Gradients. *Science, 311*(5768), 1737-1740. https://doi.org/10.1126/science.1118052

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics, 30*(9), 1236-1240. https://doi.org/10.1093/bioinformatics/btu031

Jørgensen, B. B. (1994). Sulfate reduction and thiosulfate transformations in a cyanobacterial mat during a diel oxygen cycle. *FEMS Microbiology Ecology, 13*(4), 303-312. https://www.sciencedirect.com/science/article/pii/016864969490068X

Kaczmarek, J. L., Thompson, S. V., & Holscher, H. D. (2017). Complex interactions of circadian rhythms, eating behaviors, and the gastrointestinal microbiota and their potential impact on health. *Nutrition Reviews, 75*(9), 673-682. https://doi.org/10.1093/nutrit/nux036

Kallscheuer, N., Classen, T., Drepper, T., & Marienhagen, J. (2019). Production of plant metabolites with applications in the food industry using engineered microorganisms. *Current Opinion in Biotechnology, 56*, 7-17. https://doi.org/10.1016/j.copbio.2018.07.008

Kan, J., Hanson, T. E., Ginter, J. M., Wang, K., & Chen, F. (2005). Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline systems, 1*(1), 1-7. https://doi.org/10.1186/1746-1448-1-7

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2018). New approach for understanding genome variations in KEGG. *Nucleic Acids Research, 47*(D1), D590-D595. https://doi.org/10.1093/nar/gky962

Kanesaki, Y., Shiwa, Y., Tajima, N., Suzuki, M., Watanabe, S., Sato, N., Ikeuchi, M., & Yoshikawa, H. (2012). Identification of substrain-specific mutations by massively parallel whole-genome resequencing of *Synechocystis* sp. PCC 6803. *DNA Research, 19*(1), 67-79. https://doi.org/10.1093/dnares/dsr042

Kappell, A. D., & van Waasbergen, L. G. (2007). The response regulator RpaB binds the high light regulatory 1 sequence upstream of the high-light-inducible hliB gene from the cyanobacterium *Synechocystis* PCC 6803. *Archives of Microbiology, 187*(4), 337-342. https://doi.org/10.1007/s00203-007-0213-1

Karner, M., DeLong, E. & Karl, D. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature, 409*, 507–510. https://doi.org/10.1038/35054051

Kim, Y.-I., Dong, G., Carruthers, C. W., Golden, S. S., & LiWang, A. (2008). The day/night switch in KaiC, a central oscillator component of the circadian clock of *Cyanobacteria*. *Proceedings of the National Academy of Sciences of the United States of America, 105*(35), 12825-12830. https://doi.org/10.1073/pnas.0800526105

Kirchman, D. L. (2008). New light on an important microbe in the ocean. *Proceedings of the National Academy of Sciences of the United States of America, 105*(25), 8487-8488. https://doi.org/10.1073/pnas.0804196105

Kitayama, Y., Iwasaki, H., Nishiwaki, T., & Kondo, T. (2003). KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. *EMBO Journal, 22*(9), 2127-2134. https://doi.org/10.1093/emboj/cdg212

Klamt, S., Grammel, H., Straube, R., Ghosh, R., & Gilles, E. D. (2008). Modeling the electron transport chain of purple non-sulfur bacteria. *Molecular Systems Biology, 4*(1), 156. https://doi.org/https://doi.org/10.1038/msb4100191

Klatt, C. G., Liu, Z., Ludwig, M., Kühl, M., Jensen, S. I., Bryant, D. A., & Ward, D. M. (2013). Temporal metatranscriptomic patterning in phototrophic Chloroflexi inhabiting a microbial mat in a geothermal spring. *Isme journal, 7*(9), 1775-1789. https://doi.org/10.1038/ismej.2013.52

Kleiner, M. (2019). Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *mSystems, 4*(3), e00115-00119. https://doi.org/10.1128/mSystems.00115-19

Köbler, C., Schmelling, N. M., Pawlowski, A., Spät, P., Scheurer, N. M., Berwanger, L., Maček, B., Axmann, I. M., & Wilde, A. (2021). A chimeric KaiA-like regulator extends the nonstandard KaiB3-KaiC3 clock system in bacteria. *bioRxiv*. https://doi.org/10.1101/2021.07.20.453058

Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews, 39*(6), 854-870. https://doi.org/10.1093/femsre/fuv032

Kondo, T. (2007). A cyanobacterial circadian clock based on the Kai oscillator. *Cold Spring Harbor Symposia on Quantitative Biology, 72*, 47-55. https://doi.org/10.1101/sqb.2007.72.029

Kondo, T., & Ishiura, M. (2000). The circadian clock of *Cyanobacteria*. *Bioessays, 22*(1), 10-15. https://doi.org/10.1002/(SICI)1521-1878(200001)22:1<10::AID-BIES4>3.0.CO;2-A

Kondo, T., Strayer, C. A., Kulkarni, R. D., Taylor, W., Ishiura, M., Golden, S. S., & Johnson, C. H. (1993). Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in *Cyanobacteria*. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(12), 5672-5676. https://doi.org/10.1073/pnas.90.12.5672

Koonin, E. V. (2009). Evolution of genome architecture. *The International Journal of Biochemistry & Cell Biology*, *41*(2), 298-306. https://doi.org/https://doi.org/10.1016/j.biocel.2008.09.015

Köster, J., & Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, *28*(19), 2520-2522. https://doi.org/10.1093/bioinformatics/bts480

Kuipers, B., van Noort, G. J., Vosjan, J., & Herndl, G. J. (2000). Diel periodicity of bacterioplankton in the euphotic zone of the subtropical Atlantic Ocean. *Marine Ecology Progress Series*, *201*, 13-25. https://doi.org/10.3354/meps201013

Kujawinski, E. B. (2011). The impact of microbial metabolism on marine dissolved organic matter. *Annual review of marine science*, *3*, 567-599. https://doi.org/https://doi.org/10.1146/annurev-marine-120308-081003

Lacerda, C. M. R., & Reardon, K. F. (2009). Environmental proteomics: applications of proteome profiling in environmental microbiology and biotechnology. *Briefings in Functional Genomics*, *8*(1), 75-87. https://doi.org/10.1093/bfgp/elp005

Lami, R., Cottrell Matthew, T., Ras, J., Ulloa, O., Obernosterer, I., Claustre, H., Kirchman David, L., & Lebaron, P. (2007). High Abundances of Aerobic Anoxygenic Photosynthetic Bacteria in the South Pacific Ocean. *Applied and Environmental Microbiology*, *73*(13), 4198-4205. https://doi.org/10.1128/AEM.02652-06

Leary, D. H., Hervey, W. J., Deschamps, J. R., Kusterbeck, A. W., & Vora, G. J. (2013). Which metaproteome? The impact of protein extraction bias on metaproteomic analyses. *Molecular and Cellular Probes*, *27*(5), 193-199. https://doi.org/https://doi.org/10.1016/j.mcp.2013.06.003

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674-1676. https://doi.org/10.1093/bioinformatics/btv033

Liu, K., Hou, J., Liu, Y., Hu, A., Wang, M., Wang, F., Chen, Y., & Gu, Z. (2019). Biogeography of the free-living and particle-attached bacteria in Tibetan lakes. *FEMS Microbiology Ecology*, *95*(7). https://doi.org/10.1093/femsec/fiz088

Liu, R., Liu, Y., Chen, Y., Zhan, Y., & Zeng, Q. (2019). Cyanobacterial viruses exhibit diurnal rhythms during infection. *Proceedings of the National Academy of Sciences of the United States of America, 116*(28), 14077-14082. https://doi.org/doi:10.1073/pnas.1819689116

Lodewyckx, C., Vangronsveld, J., Porteous, F., Moore, E. R. B., Taghavi, S., Mezgeay, M., & der Lelie, D. v. (2002). Endophytic Bacteria and Their Potential Applications. *Critical Reviews in Plant Sciences, 21*(6), 583-606. https://doi.org/10.1080/0735-260291044377

Louyakis, A. S., Mobberley, J. M., Vitek, B. E., Visscher, P. T., Hagan, P. D., Reid, R. P., Kozdon, R., Orland, I. J., Valley, J. W., Planavsky, N. J., Casaburi, G., & Foster, J. S. (2017). A Study of the Microbial Spatial Heterogeneity of Bahamian Thrombolites Using Molecular, Biochemical, and Stable Isotope Analyses. *Astrobiology, 17*(5), 413-430. https://doi.org/10.1089/ast.2016.1563

Loza-Correa, M., Gomez-Valero, L., & Buchrieser, C. (2010). Circadian clock proteins in prokaryotes: hidden rhythms? *Frontiers in Microbiology, 1*, 130. https://doi.org/10.3389/fmicb.2010.00130

Lu, T., Zhang, Z., Li, Y., Zhang, Q., Cui, H., Sun, L., Peijnenburg, W., Peñuelas, J., Zhu, L., Zhu, Y. G., Chen, J., & Qian, H. (2021). Does biological rhythm transmit from plants to rhizosphere microbes? *Environmental Microbiology, 23*(11), 6895-6906. https://doi.org/10.1111/1462-2920.15820

Lund, P. A. (2009). Multiple chaperonins in bacteria – why so many? *FEMS Microbiology Reviews, 33*(4), 785-800. https://doi.org/10.1111/j.1574-6976.2009.00178.x

Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., Li, K., Zhang, G., Jin, Z., He, F., Hermjakob, H., & Zhu, Y. (2018). iProX: an integrated proteome resource. *Nucleic Acids Research, 47*(D1), D1211-D1217. https://doi.org/10.1093/nar/gky869

Ma, P., Mori, T., Zhao, C., Thiel, T., & Johnson, C. H. (2016). Evolution of KaiC-Dependent Timekeepers: A Proto-circadian Timing Mechanism Confers Adaptive Fitness in the Purple Bacterium Rhodopseudomonas palustris. *PLoS Genetics, 12*(3), e1005922. https://doi.org/10.1371/journal.pgen.1005922

Maddur Puttaswamy, R. (2019). Circadian Rhythms in Plant-Microbe Interaction: For Better Performance of Bioinoculants in the Agricultural Fields. In B. Giri, R. Prasad, Q.-S. Wu, & A. Varma (Eds.), *Biofertilizers for Sustainable Agriculture and Environment* (pp. 317-332). Springer International Publishing.

Madigan, M. T. (1995). Microbiology of Nitrogen Fixation by Anoxygenic Photosynthetic Bacteria. In R. E. Blankenship, M. T. Madigan, & C. E. Bauer (Eds.), *Anoxygenic Photosynthetic Bacteria* (pp. 915-928). Springer Netherlands.

Madsen, E. L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Current Opinion in Biotechnology, 22*(3), 456-464. https://doi.org/10.1016/j.copbio.2011.01.008

Maguire, M., Coates, A. R., & Henderson, B. (2002). Chaperonin 60 unfolds its secrets of cellular communication. *Cell Stress & Chaperones, 7*(4), 317-329. https://doi.org/10.1379/1466-1268(2002)007<0317:cuisoc>2.0.co;2

Makoto, O., & Tsutomu, I. (1984). *Methods in marine zooplankton ecology*.

Maniscalco, M., Nannen, J., Sodi, V., Silver, G., Lowrey, P. L., & Bidle, K. A. (2014). Light-dependent expression of four cryptic archaeal circadian gene homologs. *Frontiers in Microbiology, 5*, 79. https://doi.org/10.3389/fmicb.2014.00079

Markson, J. S., Piechura, J. R., Puszynska, A. M., & O'Shea, E. K. (2013). Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell, 155*(6), 1396-1408. https://doi.org/10.1016/j.cell.2013.11.005

Mary, I., Garczarek, L., Tarran, G. A., Kolowrat, C., Terry, M. J., Scanlan, D. J., Burkill, P. H., & Zubkov, M. V. (2008). Diel rhythmicity in amino acid uptake by *Prochlorococcus*. *Environmental Microbiology, 10*(8), 2124-2131. https://doi.org/https://doi.org/10.1111/j.1462-2920.2008.01633.x

Matallana-Surget, S., Cavicchioli, R., Fauconnier, C., Wattiez, R., Leroy, B., Joux, F., Raftery, M. J., & Lebaron, P. (2013). Shotgun Redox Proteomics: Identification and Quantitation of Carbonylated Proteins in the UVB-Resistant Marine Bacterium, *Photobacterium angustum* S14. *PLOS ONE, 8*(7), e68112. https://doi.org/10.1371/journal.pone.0068112

Matallana-Surget, S., Jagtap, P. D., Griffin, T. J., Beraud, M., & Wattiez, R. (2018). Comparative metaproteomics to study environmental changes. In *Metagenomics* (pp. 327-363). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-08-102268-9.00017-3

Matallana-Surget, S., Joux, F., Raftery, M. J., & Cavicchioli, R. (2009). The response of the marine bacterium *Sphingopyxis alaskensis* to solar radiation assessed by quantitative proteomics. *Environmental Microbiology, 11*(10), 2660-2675. https://doi.org/https://doi.org/10.1111/j.1462-2920.2009.01992.x

Matallana-Surget, S., & Wattiez, R. (2013). Impact of Solar Radiation on Gene Expression in Bacteria. *Proteomes, 1*(2), 70-86. https://doi.org/https://doi.org/10.3390/proteomes1020070

Matenchuk, B. A., Mandhane, P. J., & Kozyrskyj, A. L. (2020). Sleep, circadian rhythm, and gut microbiota. *Sleep Medicine Reviews, 53*, 101340. https://doi.org/10.1016/j.smrv.2020.101340

Matias Rodrigues, J. F., Schmidt, T. S., Tackmann, J., & von Mering, C. (2017). MAPseq: highly efficient k-mer search with confidence estimates, for rRNA sequence analysis. *Bioinformatics, 33*(23), 3808-3810. https://doi.org/https://doi.org/10.1093/bioinformatics/btx517

May, D. H., Timmins-Schiffman, E., Mikan, M. P., Harvey, H. R., Borenstein, E., Nunn, B. L., & Noble, W. S. (2016). An Alignment-Free "Metapeptide" Strategy for Metaproteomic Characterization of Microbiome Samples Using Shotgun Metagenomic Sequencing. *Journal of proteome research, 15*(8), 2697-2705. https://doi.org/10.1021/acs.jproteome.6b00239

Mayer, L. M., Schick, L. L., Skorko, K., & Boss, E. (2006). Photodissolution of particulate organic matter from sediments. *Limnology and Oceanography, 51*(2), 1064-1071. https://doi.org/https://doi.org/10.4319/lo.2006.51.2.1064

Mayot, N., d'Ortenzio, F., Taillandier, V., Prieur, L., De Fommervault, O. P., Claustre, H., Bosse, A., Testor, P., & Conan, P. (2017). Physical and biogeochemical controls of the phytoplankton blooms in North Western Mediterranean Sea: A multiplatform approach over a complete annual cycle (2012–2013 DEWEX experiment). *Journal of Geophysical Research: Oceans, 122*(12), 9999-10019. https://doi.org/https://doi.org/10.1002/2016JC012052

McCarren, J., Becker, J. W., Repeta, D. J., Shi, Y., Young, C. R., Malmstrom, R. R., Chisholm, S. W., & DeLong, E. F. (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences of the United States of America, 107*(38), 16420-16427. https://doi.org/doi:10.1073/pnas.1010732107

McEwan, A. G. (1994). Photosynthetic electron transport and anaerobic metabolism in purple non-sulfur phototrophic bacteria. *Antonie Van Leeuwenhoek, 66*(1-3), 151-164. https://doi.org/10.1007/bf00871637

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealson, K., Pierce, N. E., ..., & Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America, 110*(9), 3229-3236. https://doi.org/10.1073/pnas.1218525110

Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahé, F., Bariat, L., Courties, C., Marie, D., Ras, J., & Mauriac, R. (2011). Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming? *Biogeosciences, 8*(9), 2785-2804. https://doi.org/https://doi.org/10.5194/bg-8-2785-2011

Michael, T. P., Mockler, T. C., Breton, G., McEntee, C., Byer, A., Trout, J. D., Hazen, S. P., Shen, R., Priest, H. D., Sullivan, C. M., Givan, S. A., Yanovsky, M., Hong, F., Kay, S. A., & Chory, J. (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genetics, 4*(2), e14. https://doi.org/10.1371/journal.pgen.0040014

Mikkelä, C., Sundh, I., Svensson, B. H., & Nilsson, M. (1995). Diurnal variation in methane emission in relation to the water table, soil temperature, climate and vegetation cover in a Swedish acid mire. *Biogeochemistry, 28*(2), 93-114. https://doi.org/10.1007/BF02180679

Min, H., Guo, H., & Xiong, J. (2005). Rhythmic gene expression in a purple photosynthetic bacterium, *Rhodobacter sphaeroides*. *FEBS Letters, 579*(3), 808-812. https://doi.org/10.1016/j.febslet.2005.01.003

Mitchell, J. G., & Kogure, K. (2006). Bacterial motility: links to the environment and a driving force for microbial physics. *FEMS Microbiology Ecology, 55*(1), 3-16. https://doi.org/10.1111/j.1574-6941.2005.00003.x

Mitsui, A., Cao, S., Takahashi, A., & Arai, T. (1987). Growth synchrony and cellular parameters of the unicellular nitrogen-fixing marine cyanobacterium, *Synechococcus* sp. strain Miami BG 043511 under continuous illumination. *Physiologia Plantarum, 69*(1), 1-8. https://doi.org/https://doi.org/10.1111/j.1399-3054.1987.tb01938.x

Moeseneder, M. M., Winter, C., & Herndl, G. J. (2001). Horizontal and vertical complexity of attached and free-living bacteria of the eastern Mediterranean Sea, determined by 16S rDNA and 16S rRNA fingerprints. *Limnology and Oceanography, 46*(1), 95-107. https://doi.org/https://doi.org/10.4319/lo.2001.46.1.0095

Morel, A., Ahn, Y.-H., Partensky, F., Vaulot, D., & Claustre, H. (1993). *Prochlorococcus* and *Synechococcus*: a comparative study of their optical properties in relation to their size and pigmentation. *Journal of Marine Research, 51*(3), 617-649. https://doi.org/https://doi.org/10.1357/0022240933223963

Mori, T., Binder, B., & Johnson, C. H. (1996). Circadian gating of cell division in *Cyanobacteria* growing with average doubling times of less than 24 hours. *Proceedings of the National Academy of Sciences of the United States of America, 93*(19), 10183-10188. https://doi.org/10.1073/pnas.93.19.10183

Mori, T., & Johnson, C. H. (2001). Circadian programming in *Cyanobacteria*. *Seminars in Cell & Developmental Biology, 12*(4), 271-278. https://doi.org/https://doi.org/10.1006/scdb.2001.0254

Morimoto, D., Šulčius, S., Tominaga, K., & Yoshida, T. (2020). Chapter Four - Predetermined clockwork microbial worlds: Current understanding of aquatic microbial diel response from model systems to complex environments. In G. M. Gadd & S. Sariaslani (Eds.), *Advances in Applied Microbiology* (Vol. 113, pp. 163-191). Academic Press.

Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S., & Rocap, G. (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *The ISME Journal, 4*(5), 673-685. https://doi.org/10.1038/ismej.2010.4

Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., & Giovannoni, S. J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature, 420*(6917), 806-810. https://doi.org/https://doi.org/10.1038/nature01240

Mullineaux, C. W., & Stanewsky, R. (2009). The rolex and the hourglass: a simplified circadian clock in *Prochlorococcus*? *Journal of Bacteriology, 191*(17), 5333-5335. https://doi.org/10.1128/jb.00719-09

Muth, T., Behne, A., Heyer, R., Kohrs, F., Benndorf, D., Hoffmann, M., Lehtevä, M., Reichl, U., Martens, L., & Rapp, E. (2015b). The MetaProteomeAnalyzer: A Powerful Open-Source Software Suite for Metaproteomics Data Analysis and Interpretation. *Journal of proteome research, 14*(3), 1557-1565. https://doi.org/10.1021/pr501246w

Muth, T., Kolmeder, C. A., Salojärvi, J., Keskitalo, S., Varjosalo, M., Verdam, F. J., Rensen, S. S., Reichl, U., de Vos, W. M., Rapp, E., & Martens, L. (2015a). Navigating through metaproteomics data: A logbook of database searching. *Proteomics, 15*(20), 3439-3453. https://doi.org/https://doi.org/10.1002/pmic.201400560

Nawrocki, E. P., Kolbe, D. L., & Eddy, S. R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics, 25*(10), 1335-1337. https://doi.org/https://doi.org/10.1093/bioinformatics/btp157

NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 46*(D1), D8-d13. https://doi.org/10.1093/nar/gkx1095

Neher, S. B., Flynn, J. M., Sauer, R. T., & Baker, T. A. (2003). Latent ClpX-recognition signals ensure LexA destruction after DNA damage. *Genes & Development, 17*(9), 1084-1089. https://doi.org/10.1101/gad.1078003

Nesvizhskii, A. I. (2010). A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *Journal of Proteomics, 73*(11), 2092-2123. https://doi.org/https://doi.org/10.1016/j.jprot.2010.08.009

Nishiwaki, T., Iwasaki, H., Ishiura, M., & Kondo, T. (2000). Nucleotide binding and autophosphorylation of the clock protein KaiC as a circadian timing process of *Cyanobacteria*. *Proceedings of the National Academy of Sciences of the United States of America, 97*(1), 495-499. https://doi.org/10.1073/pnas.97.1.495

Nishiwaki, T., & Kondo, T. (2012). Circadian autodephosphorylation of cyanobacterial clock protein KaiC occurs via formation of ATP as intermediate. *Journal of Biological Chemistry, 287*(22), 18030-18035. https://doi.org/10.1074/jbc.M112.350660

Nishiwaki, T., Satomi, Y., Nakajima, M., Lee, C., Kiyohara, R., Kageyama, H., Kitayama, Y., Temamoto, M., Yamaguchi, A., Hijikata, A., Go, M., Iwasaki, H., Takao, T., & Kondo, T. (2004). Role of KaiC phosphorylation in the circadian clock system of *Synechococcus elongatus* PCC 7942. *Proceedings of the National Academy of Sciences of the United States of America, 101*(38), 13927-13932. https://doi.org/10.1073/pnas.0403906101

Nobs, S. P., Tuganbaev, T., & Elinav, E. (2019). Microbiome diurnal rhythmicity and its impact on host physiology and disease risk. *EMBO Rep, 20*(4). https://doi.org/10.15252/embr.201847129

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research, 27*(5), 824-834. https://doi.org/10.1101/gr.213959.116

Obernosterer, I., Sempere, R., & J. Herndl, G. (2001). Ultraviolet radiation induces reversal of the bioavailability of DOM to marine bacterioplankton. *Aquatic Microbial Ecology, 24*(1), 61-68. https://doi.org/10.3354/ame024061

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), E488-497. https://doi.org/10.1073/pnas.1222099110

Ottesen, E. A., Young, C. R., Gifford, S. M., Eppley, J. M., Marin III, R., Schuster, S. C., Scholin, C. A., & DeLong, E. F. (2014). Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science, 345*(6193), 207-212. https://doi.org/https://doi.org/10.1126/science.1252476

Pace, N. R. (1997). A Molecular View of Microbial Diversity and the Biosphere. *Science, 276*(5313), 734-740. https://doi.org/10.1126/science.276.5313.734

Partensky, F., Blanchot, J., & Vaulot, D. (1999). Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bulletin de l'Institut océanographique de Monaco*, 457-476.

Pattanayak, G. K., Phong, C., & Rust, M. J. (2014). Rhythms in energy storage control the ability of the cyanobacterial circadian clock to reset. *Current Biology, 24*(16), 1934-1938. https://doi.org/10.1016/j.cub.2014.07.022

Paulose, J. K., Cassone, C. V., Graniczkowska, K. B., & Cassone, V. M. (2019). Entrainment of the Circadian Clock of the Enteric Bacterium *Klebsiella aerogenes* by Temperature Cycles. *iScience, 19*, 1202-1213. https://doi.org/10.1016/j.isci.2019.09.007

Pearson, J. A., Voisey, A. C., Boest-Bjerg, K., Wong, F. S., & Wen, L. (2021). Circadian Rhythm Modulation of Microbes During Health and Infection. *Frontiers in Microbiology, 12*, 721004. https://doi.org/10.3389/fmicb.2021.721004

Pérez-Jaramillo, J. E., Mendes, R., & Raaijmakers, J. M. (2016). Impact of plant domestication on rhizosphere microbiome assembly and functions. *Plant Molecular Biology, 90*(6), 635-644. https://doi.org/10.1007/s11103-015-0337-7

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research, 29*(9), e45. https://doi.org/10.1093/nar/29.9.e45

Pible, O., & Armengaud, J. (2015). Improving the quality of genome, protein sequence, and taxonomy databases: a prerequisite for microbiome meta-omics 2.0. *Proteomics, 15*(20), 3418-3423. https://doi.org/10.1002/pmic.201500104

Picard, F., Dressaire, C., Girbal, L., & Cocaign-Bousquet, M. (2009). Examination of post-transcriptional regulations in prokaryotes by integrative biology. *Comptes Rendus Biologies, 332*(11), 958-973. https://doi.org/10.1016/j.crvi.2009.09.005

Piwosz, K., Vrdoljak, A., Frenken, T., González-Olalla, J. M., Šantić, D., McKay, R. M., Spilling, K., Guttman, L., Znachor, P., Mujakić, I., Fecskeová, L. K., Zoccarato, L., Hanusová, M., Pessina, A., Reich, T., Grossart, H. P., & Koblížek, M. (2020). Light and Primary Production Shape Bacterial Activity and Community Composition of Aerobic Anoxygenic Phototrophic Bacteria in a Microcosm Experiment. *mSphere, 5*(4). https://doi.org/10.1128/mSphere.00354-20

Poretsky, R. S., Hewson, I., Sun, S., Allen, A. E., Zehr, J. P., & Moran, M. A. (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology, 11*(6), 1358-1375. https://doi.org/https://doi.org/10.1111/j.1462-2920.2008.01863.x

Prieto-Barajas, C. M., Valencia-Cantero, E., & Santoyo, G. (2018). Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. *Electronic Journal of Biotechnology, 31*, 48-56. https://doi.org/https://doi.org/10.1016/j.ejbt.2017.11.001

Puszynska, A. M., & O'Shea, E. K. (2017). Switching of metabolic programs in response to light availability is an essential function of the cyanobacterial circadian output pathway. *Elife, 6*. https://doi.org/10.7554/eLife.23210

Ramakers, C., Ruijter, J. M., Deprez, R. H., & Moorman, A. F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters, 339*(1), 62-66. https://doi.org/10.1016/s0304-3940(02)01423-4

Rappé, M. S., Connon, S. A., Vergin, K. L., & Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature, 418*(6898), 630-633. https://doi.org/10.1038/nature00917

Rechenberger, J., Samaras, P., Jarzab, A., Behr, J., Frejno, M., Djukovic, A., Sanz, J., González-Barberá, E. M., Salavert, M., López-Hontangas, J. L., Xavier, K. B., Debrauwer, L., Rolain, J.-M., Sanz, M., Garcia-Garcera, M., Wilhelm, M., Ubeda, C., & Kuster, B. (2019). Challenges in Clinical Metaproteomics Highlighted by the Analysis of Acute Leukemia Patients with Gut Colonization by Multidrug-Resistant Enterobacteriaceae. *Proteomes, 7*(1), 2. https://www.mdpi.com/2227-7382/7/1/2

Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research, 38*(20), e191-e191. https://doi.org/10.1093/nar/gkq747

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., ..., & Chisholm, S. W. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature, 424*(6952), 1042-1047. https://doi.org/10.1038/nature01947

Rodríguez, A., Hernández-Herreros, N., García, J. L., & Auxiliadora Prieto, M. (2021). Enhancement of biohydrogen production rate in *Rhodospirillum rubrum* by a dynamic CO-feeding strategy using dark fermentation. *Biotechnology for Biofuels 14*(1), 168. https://doi.org/10.1186/s13068-021-02017-6

Röling, W. F., & van Bodegom, P. M. (2014). Toward quantitative understanding on microbial community structure and functioning: a modeling-centered approach using degradation of marine oil spills as example. *Frontiers in Microbiology, 5*, 125. https://doi.org/https://doi.org/10.3389/fmicb.2014.00125

Russo, D. A., Couto, N., Beckerman, A. P., & Pandhal, J. (2016). A Metaproteomic Analysis of the Response of a Freshwater Microbial Community under Nutrient Enrichment. *Frontiers in Microbiology, 7*. https://doi.org/10.3389/fmicb.2016.01172

Saharan, B., & Nehra, V. (2011). Plant growth promoting rhizobacteria: a critical review. *Life Sciences and Medicine Research, 21*(1), 30.

Saini, R., Jaskolski, M., & Davis, S. J. (2019). Circadian oscillator proteins across the kingdoms of life: structural aspects. *BMC Biology, 17*(1), 13. https://doi.org/10.1186/s12915-018-0623-3

Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M. R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd, A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of proteome research, 18*(4), 1461-1476. https://doi.org/10.1021/acs.jproteome.8b00761

Saito, M. A., Dorsk, A., Post, A. F., McIlvin, M. R., Rappé, M. S., DiTullio, G. R., & Moran, D. M. (2015). Needles in the blue sea: Sub-species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome. *Proteomics, 15*(20), 3521-3531. https://doi.org/https://doi.org/10.1002/pmic.201400630

Salmela, M. J., & Weinig, C. (2019). The fitness benefits of genetic variation in circadian clock regulation. *Current Opinion in Plant Biology, 49*, 86-93. https://doi.org/https://doi.org/10.1016/j.pbi.2019.06.003

Sánchez-Baracaldo, P., & Cardona, T. (2020). On the origin of oxygenic photosynthesis and *Cyanobacteria*. *New Phytologist, 225*(4), 1440-1446. https://doi.org/https://doi.org/10.1111/nph.16249

Saran, A. R., Dave, S., & Zarrinpar, A. (2020). Circadian Rhythms in the Pathogenesis and Treatment of Fatty Liver Disease. *Gastroenterology, 158*(7), 1948-1966.e1941. https://doi.org/10.1053/j.gastro.2020.01.050

Sartor, F., Eelderink-Chen, Z., Aronson, B., Bosman, J., Hibbert, L. E., Dodd, A. N., Kovács Á, T., & Merrow, M. (2019). Are There Circadian Clocks in Non-Photosynthetic Bacteria? *Biology (Basel), 8*(2). https://doi.org/10.3390/biology8020041

Schmelling, N. M., Lehmann, R., Chaudhury, P., Beck, C., Albers, S.-V., Axmann, I. M., & Wiegard, A. (2017). Minimal tool set for a prokaryotic circadian clock. *BMC Evolutionary Biology, 17*(1), 169. https://doi.org/10.1186/s12862-017-0999-7

Schmelling, N. M., Scheurer, N., Köbler, C., Wilde, A., & Axmann, I. M. (2021). Diversity of Timing Systems in *Cyanobacteria* and Beyond. In C. H. Johnson & M. J. Rust (Eds.), *Circadian Rhythms in Bacteria and Microbiomes* (pp. 179-202). Springer International Publishing.

Schneider, T., Schmid, E., de Castro, J. V., Jr., Cardinale, M., Eberl, L., Grube, M., Berg, G., & Riedel, K. (2011). Structure and function of the symbiosis partners of the lung lichen (Lobaria pulmonaria L. Hoffm.) analyzed by metaproteomics. *Proteomics, 11*(13), 2752-2756. https://doi.org/10.1002/pmic.201000679

Serrano-Villar, S., Rojo, D., Martínez-Martínez, M., Deusch, S., Vázquez-Castellanos, J. F., Bargiela, R., Sainz, T., Vera, M., Moreno, S., Estrada, V., Gosalbes, M. J., Latorre, A., Seifert, J., Barbas, C., Moya, A., & Ferrer, M. (2016). Gut Bacteria Metabolism Impacts Immune Recovery in HIV-infected Individuals. *EBioMedicine, 8*, 203-216. https://doi.org/https://doi.org/10.1016/j.ebiom.2016.04.033

Sessitsch, A., Howieson, J. G., Perret, X., Antoun, H., & Martínez-Romero, E. (2002). Advances in *Rhizobium* Research. *Critical Reviews in Plant Sciences, 21*(4), 323-378. https://doi.org/10.1080/0735-260291044278

Shahraki, A. H., Chaganti, S. R., & Heath, D. D. (2021). Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. *Microbial Ecology, 81*(1), 1-13. https://doi.org/10.1007/s00248-020-01539-0

Sharma, V. K. (2003). Adaptive Significance of Circadian Clocks. *Chronobiology International, 20*(6), 901-919. https://doi.org/10.1081/CBI-120026099

Simon, M., Grossart, H.-P., Schweitzer, B., & Ploug, H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. *Aquatic Microbial Ecology, 28*(2), 175-211. https://doi.org/10.3354/ame028175

Slilaty, S. N., & Little, J. W. (1987). Lysine-156 and serine-119 are required for LexA repressor cleavage: a possible mechanism. *Proceedings of the National Academy of Sciences of the United States of America, 84*(12), 3987-3991. https://doi.org/10.1073/pnas.84.12.3987

Smith, R. M., & Williams, S. B. (2006). Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*. *Proceedings of the National Academy of Sciences of the United States of America, 103*(22), 8564-8569. https://doi.org/10.1073/pnas.0508696103

Snijder, J., & Axmann, I. M. (2019). The Kai-Protein Clock-Keeping Track of *Cyanobacteria's* Daily Life. *Subcellular Biochemistry, 93*, 359-391. https://doi.org/10.1007/978-3-030-28151-9_12

Sommaruga, R., Hofer Julia, S., Alonso-Sáez, L., & Gasol Josep, M. (2005). Differential Sunlight Sensitivity of Picophytoplankton from Surface Mediterranean Coastal Waters. *Applied and Environmental Microbiology, 71*(4), 2154-2157. https://doi.org/10.1128/AEM.71.4.2154-2157.2005

Sowell, S. M., Abraham, P. E., Shah, M., Verberkmoes, N. C., Smith, D. P., Barofsky, D. F., & Giovannoni, S. J. (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *The ISME Journal, 5*(5), 856-865. https://doi.org/10.1038/ismej.2010.168

Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., Carlson, C. A., Smith, R. D., & Giovanonni, S. J. (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *The ISME Journal, 3*(1), 93-105. https://doi.org/10.1038/ismej.2008.83

Stal, L. J., & Cretoiu, M. S. (2016). *The marine microbiome*. Springer.

Stal, L. J., & Krumbein, W. E. (1987). Temporal separation of nitrogen fixation and photosynthesis in the filamentous, non-heterocystous cyanobacterium *Oscillatoria* sp. *Archives of Microbiology, 149*(1), 76-80. https://doi.org/10.1007/BF00423140

Staley, C., Ferrieri, A. P., Tfaily, M. M., Cui, Y., Chu, R. K., Wang, P., Shaw, J. B., Ansong, C. K., Brewer, H., Norbeck, A. D., Markillie, M., do Amaral, F., Tuleski, T., Pellizzaro, T., Agtuca, B., Ferrieri, R., Tringe, S. G., Paša-Tolić, L., Stacey, G., & Sadowsky, M. J. (2017). Diurnal cycling of rhizosphere bacterial communities is associated with shifts in carbon metabolism. *Microbiome, 5*(1), 65. https://doi.org/10.1186/s40168-017-0287-1

Staley, J. T., & Konopka, A. (1985). Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology, 39*(1), 321-346. https://doi.org/10.1146/annurev.mi.39.100185.001541

Steppe, T. F., & Paerl, H. W. (2002). Potential N2 fixation by sulfate-reducing bacteria in a marine intertidal microbial mat. *Aquatic Microbial Ecology, 28*(1), 1-12. https://doi.org/10.3354/ame028001

Straub, C., Quillardet, P., Vergalli, J., de Marsac, N. T., & Humbert, J.-F. (2011). A Day in the Life of *Microcystis aeruginosa* Strain PCC 7806 as Revealed by a Transcriptomic Analysis. *PLOS ONE, 6*(1), e16208. https://doi.org/10.1371/journal.pone.0016208

Sun, Z., & Blanchard, J. L. (2014). Strong Genome-Wide Selection Early in the Evolution of *Prochlorococcus* Resulted in a Reduced Genome through the Loss of a Large Number of Small Effect Genes. *PLOS ONE, 9*(3), e88837. https://doi.org/10.1371/journal.pone.0088837

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., ..., & Velayoudon, D. (2015). Structure and function of the global ocean microbiome. *Science, 348*(6237), 1261359. https://doi.org/10.1126/science.1261359

Swan, J. A., Golden, S. S., LiWang, A., & Partch, C. L. (2018). Structure, function, and mechanism of the core circadian clock in *Cyanobacteria*. *Journal of Biological Chemistry, 293*(14), 5026-5034. https://doi.org/10.1074/jbc.TM117.001433

Sweeney, B. M., & Borgese, M. B. (1989). A circadian rhythm in cell division in a prokaryote, the cyanobacterium *Synechococcus* wh78031 *Journal of Phycology, 25*(1), 183-186. https://doi.org/https://doi.org/10.1111/j.0022-3646.1989.00183.x

Takai, N., Nakajima, M., Oyama, T., Kito, R., Sugita, C., Sugita, M., Kondo, T., & Iwasaki, H. (2006). A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in *Cyanobacteria*. *Proceedings of the National Academy of Sciences of the United States of America, 103*(32), 12109-12114. https://doi.org/doi:10.1073/pnas.0602955103

Takaichi, S. (2009). Distribution and Biosynthesis of Carotenoids. In C. N. Hunter, F. Daldal, M. C. Thurnauer, & J. T. Beatty (Eds.), *The Purple Phototrophic Bacteria* (pp. 97-117). Springer Netherlands.

Tamaki, H., Sekiguchi, Y., Hanada, S., Nakamura, K., Nomura, N., Matsumura, M., & Kamagata, Y. (2005). Comparative Analysis of Bacterial Diversity in Freshwater Sediment of a Shallow Eutrophic Lake by Molecular and Improved Cultivation-Based Techniques. *Applied and Environmental Microbiology, 71*(4), 2162-2169. https://doi.org/10.1128/AEM.71.4.2162-2169.2005

Tanca, A., Palomba, A., Deligios, M., Cubeddu, T., Fraumene, C., Biosa, G., Pagnozzi, D., Addis, M. F., & Uzzau, S. (2013). Evaluating the Impact of Different Sequence Databases on Metaproteome Analysis: Insights from a Lab-Assembled Microbial Mixture. *PLOS ONE, 8*(12), e82981. https://doi.org/10.1371/journal.pone.0082981

Tanca, A., Palomba, A., Fraumene, C., Pagnozzi, D., Manghina, V., Deligios, M., Muth, T., Rapp, E., Martens, L., Addis, M. F., & Uzzau, S. (2016). The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome, 4*(1), 51. https://doi.org/10.1186/s40168-016-0196-8

Tedetti, M., & Sempéré, R. (2006). Penetration of Ultraviolet Radiation in the Marine Environment. A Review. *Photochemistry and Photobiology, 82*(2), 389-397. https://doi.org/https://doi.org/10.1562/2005-11-09-IR-733

Teeling, H., Fuchs, B. M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C. M., Kassabgy, M., Huang, S., Mann, A. J., Waldmann, J., Weber, M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F. D., ..., & Amann, R. (2012). Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science, 336*(6081), 608-611. https://doi.org/doi:10.1126/science.1218344

Thaiss, C. A., Zeevi, D., Levy, M., Zilberman-Schapira, G., Suez, J., Tengeler, A. C., Abramson, L., Katz, M. N., Korem, T., Zmora, N., Kuperman, Y., Biton, I., Gilad, S., Harmelin, A., Shapiro, H., Halpern, Z., Segal, E., & Elinav, E. (2014). Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. *Cell, 159*(3), 514-529. https://doi.org/10.1016/j.cell.2014.09.048

The UniProt Consortium. (2016). UniProt: the universal protein knowledgebase. *Nucleic Acids Research, 45*(D1), D158-D169. https://doi.org/10.1093/nar/gkw1099

Thekaekara, M. P. (1976). Solar radiation measurement: Techniques and instrumentation. *Solar Energy, 18*(4), 309-325. https://doi.org/https://doi.org/10.1016/0038-092X(76)90058-X

Thiel, T. (2005). Nitrogen Fixation in Heterocyst-Forming *Cyanobacteria*. In W. Klipp, B. Masepohl, J. R. Gallon, & W. E. Newton (Eds.), *Genetics and Regulation of Nitrogen Fixation in Free-Living Bacteria* (pp. 73-110). Springer Netherlands.

Thore, A., Keister, D. L., & Pietro, A. S. (1969). Studies on the respiratory system of aerobically (Dark) and anaerobically (Light) grown *Rhodospirillum rubrum*. *Archives of Microbiology, 67*(4), 378-396. https://doi.org/10.1007/BF00412584

Timmins-Schiffman, E., May, D. H., Mikan, M., Riffle, M., Frazar, C., Harvey, H. R., Noble, W. S., & Nunn, B. L. (2017). Critical decisions in metaproteomics: achieving high confidence protein annotations in a sea of unknowns. *The ISME Journal, 11*(2), 309-314. https://doi.org/10.1038/ismej.2016.132

Ting, C. S., Rocap, G., King, J., & Chisholm, S. W. (2002). Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends in Microbiology, 10*(3), 134-142. https://doi.org/https://doi.org/10.1016/S0966-842X(02)02319-3

Tomasch, J., Gohl, R., Bunk, B., Diez, M. S., & Wagner-Döbler, I. (2011). Transcriptional response of the photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. *Isme journal, 5*(12), 1957-1968. https://doi.org/10.1038/ismej.2011.68

Tsai, A. Y., Gong, G. C., Sanders, R. W., Chiang, K. P., Huang, J. K., & Chan, Y. F. (2012). Viral lysis and nanoflagellate grazing as factors controlling diel variations of *Synechococcus* spp. summer abundance in coastal waters of Taiwan. *Aquatic Microbial Ecology, 66*(2), 159-167. https://doi.org/https://doi.org/10.3354/ame01566

Tseng, R., Chang, Y. G., Bravo, I., Latham, R., Chaudhary, A., Kuo, N. W., & Liwang, A. (2014). Cooperative KaiA-KaiB-KaiC interactions affect KaiB/SasA competition in the circadian clock of *Cyanobacteria*. *Journal of Molecular Biology, 426*(2), 389-402. https://doi.org/10.1016/j.jmb.2013.09.040

Turley, C., & Stutt, E. (2000). Depth-related cell-specific bacterial leucine incorporation rates on particles and its biogeochemical significance in the Northwest Mediterranean. *Limnology and Oceanography, 45*(2), 419-425. https://doi.org/https://doi.org/10.4319/lo.2000.45.2.0419

Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Reece, J. B. (2017). *Campbell biology*. Pearson Education, Incorporated.

van Gemerden, H., Montesinos, E., Mas, J., & Guerrero, R. (1985). Diel cycle of metabolism of phototrophic purple sulfur bacteria in Lake Cisó (Spain)1. *Limnology and Oceanography, 30*(5), 932-943. https://doi.org/https://doi.org/10.4319/lo.1985.30.5.0932

Van Mooy, B. A. S., Rocap, G., Fredricks, H. F., Evans, C. T., & Devol, A. H. (2006). Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences of the United States of America, 103*(23), 8607-8612. https://doi.org/10.1073/pnas.0600540103

Van Praag, E., Degli Agosti, R., & Bachofen, R. (2000). Rhythmic activity of uptake hydrogenase in the prokaryote *Rhodospirillum rubrum*. *Journal of Biological Rhythms, 15*(3), 218-224. https://doi.org/10.1177/074873040001500303

Vaulot, D., & Marie, D. (1999). Diel variability of photosynthetic picoplankton in the equatorial Pacific. *Journal of Geophysical Research: Oceans, 104*(C2), 3297-3310. https://doi.org/https://doi.org/10.1029/98JC01333

Vaulot, D., Marie, D., Olson, R. J., & Chisholm, S. W. (1995). Growth of *Prochlorococcus*, a Photosynthetic Prokaryote, in the Equatorial Pacific Ocean. *Science, 268*(5216), 1480-1482. https://doi.org/10.1126/science.268.5216.1480

Venkova, T., Yeo, C. C., & Espinosa, M. (2018). Editorial: The Good, The Bad, and The Ugly: Multiple Roles of Bacteria in Human Life. *Frontiers in Microbiology, 9*, 1702. https://doi.org/10.3389/fmicb.2018.01702

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., ..., & Smith, H. O. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science, 304*(5667), 66-74. https://doi.org/10.1126/science.1093857

Vijayan, V., Zuzow, R., & O'Shea, E. K. (2009). Oscillations in supercoiling drive circadian gene expression in *Cyanobacteria*. *Proceedings of the National Academy of Sciences of the United States of America, 106*(52), 22564-22568. https://doi.org/10.1073/pnas.0912673106

Volk, T., & Hoffert, M. I. (1985). Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO2 Changes. In *The Carbon Cycle and Atmospheric CO2: Natural Variations Archean to Present* (pp. 99-110). American Geophysical Union; Geophysical Monograph 32.

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Research, 31*(1), 258-261. https://doi.org/10.1093/nar/gkg034

Waldbauer, J. R., Rodrigue, S., Coleman, M. L., & Chisholm, S. W. (2012). Transcriptome and Proteome Dynamics of a Light-Dark Synchronized Bacterial Cell Cycle. *PLOS ONE, 7*(8), e43432. https://doi.org/10.1371/journal.pone.0043432

Wang, D.-Z., Xie, Z.-X., & Zhang, S.-F. (2014). Marine metaproteomics: Current status and future directions. *Journal of Proteomics, 97*, 27-35. https://doi.org/https://doi.org/10.1016/j.jprot.2013.08.024

Wang, Q., Wei, W., Gong, Y., Yu, Q., Li, Q., Sun, J., & Yuan, Z. (2017). Technologies for reducing sludge production in wastewater treatment plants: State of the art. *Science of the Total Environment, 587-588*, 510-521. https://doi.org/10.1016/j.scitotenv.2017.02.203

Welkie, D. G., Rubin, B. E., Diamond, S., Hood, R. D., Savage, D. F., & Golden, S. S. (2019). A Hard Day's Night: *Cyanobacteria* in Diel Cycles. *Trends in Microbiology, 27*(3), 231-242. https://doi.org/https://doi.org/10.1016/j.tim.2018.11.002

Werner, J., Géron, A., Kerssemakers, J., & Matallana-Surget, S. (2019). *mPies*: a novel metaproteomics tool for the creation of relevant protein databases and automatized protein annotation. *Biology Direct, 14*(1), 1-5. https://doi.org/https://doi.org/10.1186/s13062-019-0253-x

Whitehead, K., Pan, M., Masumura, K., Bonneau, R., & Baliga, N. S. (2009). Diurnally entrained anticipatory behavior in archaea. *PLOS ONE, 4*(5), e5485. https://doi.org/10.1371/journal.pone.0005485

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America, 95*(12), 6578-6583. https://doi.org/10.1073/pnas.95.12.6578

Wiegard, A., Dörrich, A. K., Deinzer, H.-T., Beck, C., Wilde, A., Holtzendorff, J., & Axmann, I. M. (2013). Biochemical analysis of three putative KaiC clock proteins from *Synechocystis* sp. PCC 6803 suggests their functional divergence. *Microbiology, 159*(Pt_5), 948-958. https://doi.org/https://doi.org/10.1099/mic.0.065425-0

Wiegard, A., Köbler, C., Oyama, K., Dörrich, A. K., Azai, C., Terauchi, K., Wilde, A., & Axmann, I. M. (2020). *Synechocystis* KaiC3 Displays Temperature- and KaiB-Dependent ATPase Activity and Is Important for Growth in Darkness. *Journal of Bacteriology, 202*(4). https://doi.org/10.1128/jb.00478-19

Williams, T. J., & Cavicchioli, R. (2014). Marine metaproteomics: deciphering the microbial metabolic food web. *Trends in Microbiology, 22*(5), 248-260. https://doi.org/https://doi.org/10.1016/j.tim.2014.03.004

Williams, T. J., Long, E., Evans, F., DeMaere, M. Z., Lauro, F. M., Raftery, M. J., Ducklow, H., Grzymski, J. J., Murray, A. E., & Cavicchioli, R. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *The ISME Journal, 6*(10), 1883-1900. https://doi.org/10.1038/ismej.2012.28

Williams, T. J., Wilkins, D., Long, E., Evans, F., DeMaere, M. Z., Raftery, M. J., & Cavicchioli, R. (2013). The role of planktonic *Flavobacteria* in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environmental Microbiology, 15*(5), 1302-1317. https://doi.org/https://doi.org/10.1111/1462-2920.12017

Wilmes, P., & Bond, P. L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environmental Microbiology, 6*(9), 911-920. https://doi.org/https://doi.org/10.1111/j.1462-2920.2004.00687.x

Wilmes, P., Heintz-Buschart, A., & Bond, P. L. (2015). A decade of metaproteomics: Where we stand and what the future holds. *Proteomics, 15*(20), 3409-3417. https://doi.org/https://doi.org/10.1002/pmic.201500183

Winter, C., Herndl, G. J., & Weinbauer, M. G. (2004). Diel cycles in viral infection of bacterioplankton in the North Sea. *Aquatic Microbial Ecology, 35*(3), 207-216. https://doi.org/10.3354/ame035207

Woelfle, M. A., Xu, Y., Qin, X., & Johnson, C. H. (2007). Circadian rhythms of superhelical status of DNA in *Cyanobacteria*. *Proceedings of the National Academy of Sciences of the United States of America, 104*(47), 18819-18824. https://doi.org/10.1073/pnas.0706069104

Wöhlbrand, L., Trautwein, K., & Rabus, R. (2013). Proteomic tools for environmental microbiology—A roadmap from sample preparation to protein identification and quantification. *Proteomics, 13*(18-19), 2700-2730. https://doi.org/https://doi.org/10.1002/pmic.201300175

Woodcroft, B. (2018). *Singlem. v0.11.0.* https://github.com/wwood/singlem/

Woodger, F. J., Badger, M. R., & Price, G. D. (2003). Inorganic Carbon Limitation Induces Transcripts Encoding Components of the CO2-Concentrating Mechanism in *Synechococcus* sp. PCC7942 through a Redox-Independent Pathway. *Plant Physiology, 133*(4), 2069-2080. https://doi.org/10.1104/pp.103.029728

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science, 347*(6223), 1257594. https://doi.org/10.1126/science.1257594

Wozniak, B., & Dera, J. (2007). *Light absorption by suspended particulate matter (SPM) in sea water*. Springer.

Wu, G., Tang, W., He, Y., Hu, J., Gong, S., He, Z., Wei, G., Lv, L., Jiang, Y., Zhou, H., & Chen, P. (2018). Light exposure influences the diurnal oscillation of gut microbiota in mice. *Biochemical and Biophysical Research Communications, 501*(1), 16-23. https://doi.org/10.1016/j.bbrc.2018.04.095

Yen, U.-C., Huang, T.-C., & Yen, T. C. (2004). Observation of the circadian photosynthetic rhythm in *Cyanobacteria* with a dissolved-oxygen meter. *Plant Science, 166*, 949-952. https://doi.org/10.1016/J.PLANTSCI.2003.12.005

Yoshida, T., Nishimura, Y., Watai, H., Haruki, N., Morimoto, D., Kaneko, H., Honda, T., Yamamoto, K., Hingamp, P., Sako, Y., Goto, S., & Ogata, H. (2018). Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *The ISME Journal, 12*(5), 1287-1295. https://doi.org/10.1038/s41396-018-0052-x

Yu, J., Liberton, M., Cliften, P. F., Head, R. D., Jacobs, J. M., Smith, R. D., Koppenaal, D. W., Brand, J. J., & Pakrasi, H. B. (2015). *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO2. *Scientific Reports, 5*(1), 1-10. https://doi.org/https://doi.org/10.1038/srep08132

Yurkov, V., & Csotonyi, J. T. (2009). New Light on Aerobic Anoxygenic Phototrophs. In C. N. Hunter, F. Daldal, M. C. Thurnauer, & J. T. Beatty (Eds.), *The Purple Phototrophic Bacteria* (pp. 31-55). Springer Netherlands.

Yutin, N., Suzuki, M. T., Teeling, H., Weber, M., Venter, J. C., Rusch, D. B., & Béjà, O. (2007). Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environmental Microbiology, 9*(6), 1464-1475. https://doi.org/https://doi.org/10.1111/j.1462-2920.2007.01265.x

Zehr, J. P., Mellon, M. T., & Zani, S. (1998). New Nitrogen-Fixing Microorganisms Detected in Oligotrophic Oceans by Amplification of Nitrogenase (nifH) Genes. *Applied and Environmental Microbiology, 64*(12), 5067. https://doi.org/10.1128/aem.64.12.5067-5067.1998

Zeng, Y., Selyanin, V., Lukeš, M., Dean, J., Kaftan, D., Feng, F., & Koblížek, M. (2015). Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium *Gemmatimonas phototrophica* sp. nov., and emended descriptions of the genus *Gemmatimonas* and *Gemmatimonas aurantiaca*. *International Journal of Systematic and Evolutionary Microbiology, 65*(8), 2410-2419. https://doi.org/10.1099/ijs.0.000272

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., ..., & Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research, 46*(D1), D754-d761. https://doi.org/10.1093/nar/gkx1098

Zhang, Y., Xiao, W., & Jiao, N. (2016). Linking biochemical properties of particles to particle-attached and free-living bacterial community structure along the particle density gradient from freshwater to open ocean. *Journal of Geophysical Research: Biogeosciences, 121*(8), 2261-2274. https://doi.org/https://doi.org/10.1002/2016JG003390

Zhao, C., Li, Z., Li, T., Zhang, Y., Bryant, D. A., & Zhao, J. (2015). High-yield production of extracellular type-I cellulose by the cyanobacterium *Synechococcus* sp. PCC 7002. *Cell Discovery, 1*(1), 15004. https://doi.org/10.1038/celldisc.2015.4

Zhao, X., & Drlica, K. (2014). Reactive oxygen species and the bacterial response to lethal stress. *Current Opinion in Microbiology, 21*, 1-6. https://doi.org/https://doi.org/10.1016/j.mib.2014.06.008

Zheng, Q., Liu, Y., Steindler, L., & Jiao, N. (2015). Pyrosequencing analysis of aerobic anoxygenic phototrophic bacterial community structure in the oligotrophic western Pacific Ocean. *FEMS Microbiology Letters, 362*(8), fnv034. https://doi.org/10.1093/femsle/fnv034

Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., Bender, R. A., & Kustu, S. (2000). Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: Scavenging as a defense against nitrogen limitation. *Proceedings of the National Academy of Sciences of the United States of America, 97*(26), 14674-14679. https://doi.org/doi:10.1073/pnas.97.26.14674

Zinser, E. R., Lindell, D., Johnson, Z. I., Futschik, M. E., Steglich, C., Coleman, M. L., Wright, M. A., Rector, T., Steen, R., McNulty, N., Thompson, L. R., & Chisholm, S. W. (2009). Choreography of the Transcriptome, Photophysiology, and Cell Cycle of a Minimal Photoautotroph, *Prochlorococcus*. *PLOS ONE, 4*(4), e5135. https://doi.org/10.1371/journal.pone.0005135

Zubkov, M. V. (2009). Photoheterotrophy in marine prokaryotes. *Journal of Plankton Research, 31*(9), 933-938. https://doi.org/10.1093/plankt/fbp043

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., Not, F., Massana, R., Ulloa, O., & Scanlan, D. J. (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology, 10*(1), 147-161. https://doi.org/10.1111/j.1462-2920.2007.01440.x

Appendices

Appendix A: *mPies* program features

- Project name: *mPies*: a novel metaproteomics tool for the creation of relevant protein databases and automatized protein annotation (DOI: 10.1101/690131).
- Project home page: https://github.com/johanneswerner/mPies
- Archived version: 10.5281/zenodo.2539710
- Operating system(s): Linux
- Programming language: Python 3
- Other requirements: Conda, other packages installed with *mPies*: CD-Hit AuxTools (10.1093/bioinformatics/bts565), Diamond (10.1038/nmeth.3176), ETE3 (http://mbe.oxfordjournals.org/content/early/2016/03/21/molbev.msw046), FragGeneScan (10.1093/nar/gkq747), MEGAHIT (10.1093/bioinformatics/btv033), Prodigal (10.1186/1471-2105-11-119), Snakemake (10.1093/bioinformatics/bts480), Spades (10.1089/cmb.2012.0021), Trimmomatic (10.1093/bioinformatics/btu170). At least 150 GB of disk space (depending on the used database), 64 GB memory (if assembly is necessary) and 16 CPUs are recommended for running *mPies*.
- License: GNU GPL-3.0
- Any restrictions to use by non-academics: none

Appendix B: Examples of functional consensus annotations using *mPies*.

Example 1: automatic *mPies* annotation

Candidate annotations from the top alignment hits	Frequency	Similarity (%)
Glutamine synthetase	16	80
Glutamine synthetase nodule isozyme	1	5
Glutamine synthetase root isozyme 3	1	5
Glutamine synthetase root isozyme A	1	5
Glutamine synthetase root isozyme B	1	5

Consensus functional annotation

Glutamine synthetase

Example 2: manual annotation

Candidate annotations from the top alignment hits	Occurrences	Proportion (%)
60 kDa chaperonin 1	9	45
60 kDa chaperonin 2	3	15
60 kDa chaperonin 3	3	15
60 kDa chaperonin	2	10
60 kDa chaperonin 4	1	5
60 kDa chaperonin 5	1	5
60 kDa chaperonin 7	1	5

Consensus functional annotation

60 kDa chaperonin

Example 3: no consensus annotation

Candidate annotations from the top alignment hits	Occurrences	Proportion (%)
3-hexulose-6-phosphate synthase	2	25
5,6,7,8-tetrahydromethanopterin hydro-lyase	2	25
5,6,7,8-tetrahydromethanopterin hydro-lyase	2	25
Bifunctional enzyme Fae/Hps	2	25

Consensus functional annotation

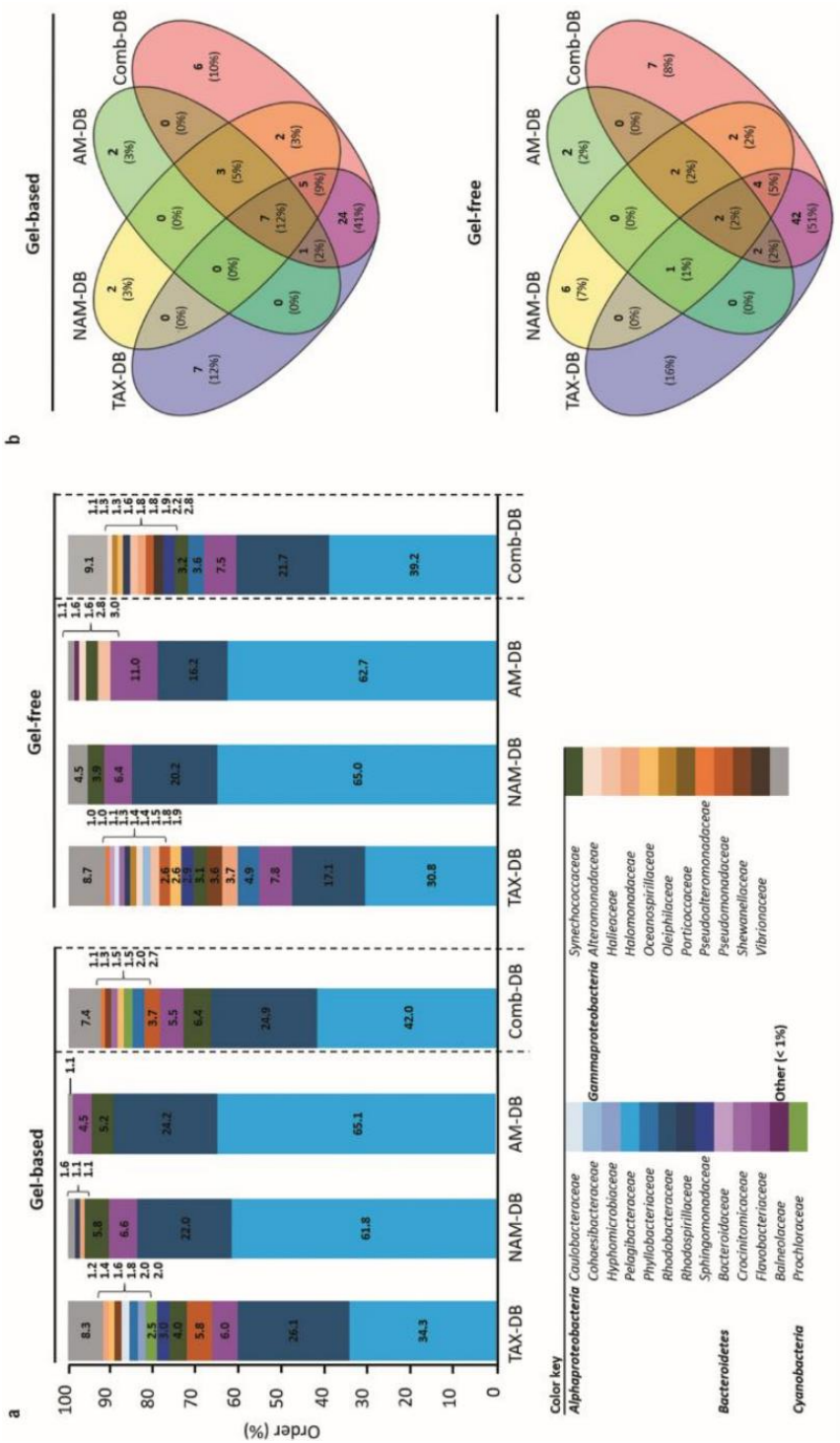
NA

Appendix C: Taxonomic and functional protein annotation depending on the methodology.

	Gel-based				Gel-free			
Taxonomic level	**TAX-DB**	**NAM-DB**	**AM-DB**	**Comb-DB**	**TAX-DB**	**NAM-DB**	**AM-DB**	**Comb-DB**
Domain	100	97	96	100	99	94	93	99
Phylum	95	92	91	94	95	87	88	92
Class	81	78	81	80	85	77	80	81
Order	70	41	45	51	71	42	44	52
Family	64	34	37	44	66	32	33	43
Genus	50	7	13	28	43	8	9	16
Function	66	50	61	59	77	54	54	67

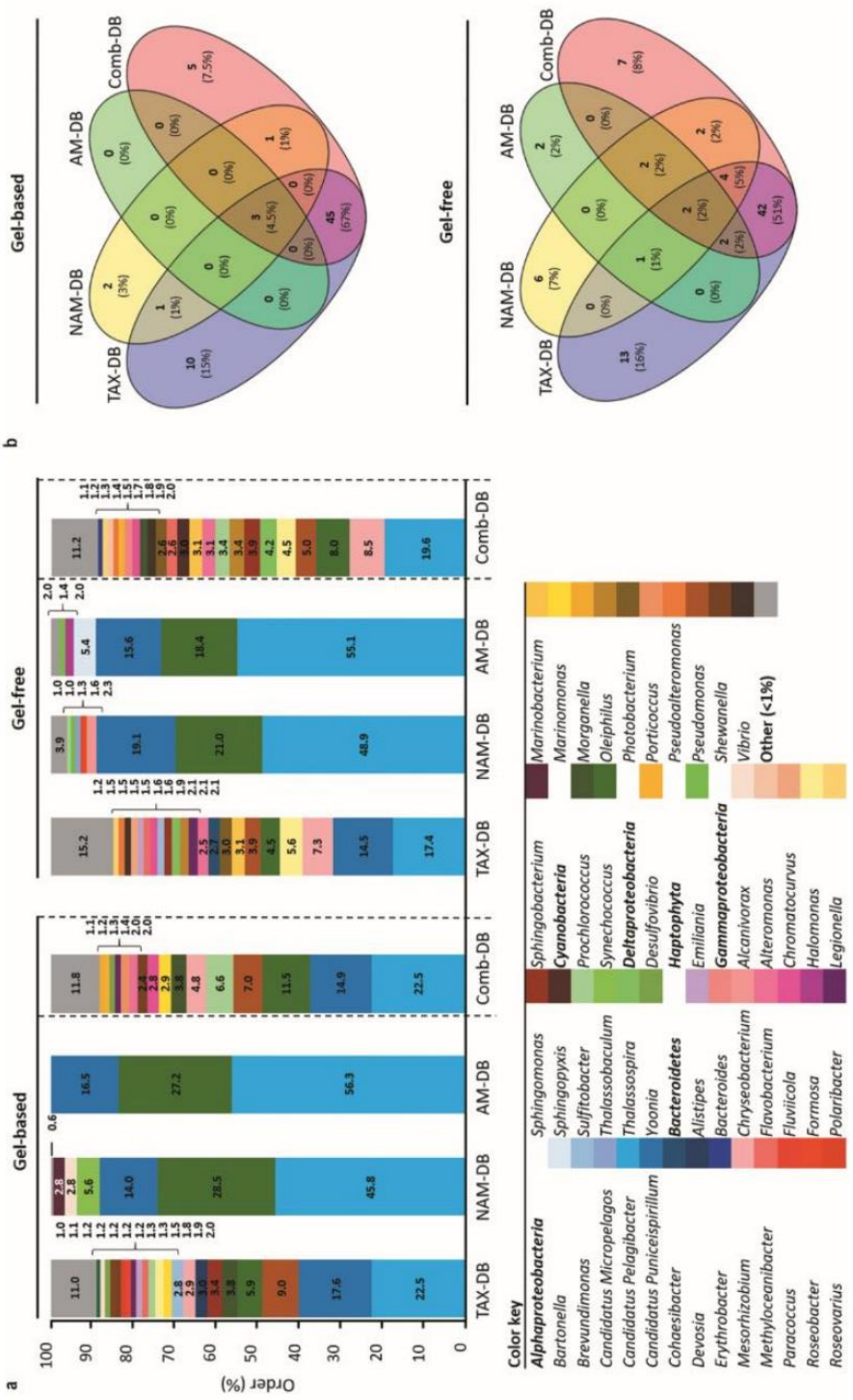
Appendix D: Relative taxonomic composition at family level depending on the methodology.

(**a**) Relative taxonomic composition at family level for each methodology. Values represent the proportion of proteins with identical taxonomy on total identified protein using TAX-DB, NAM-DB, AM-DB, or Comb-DB in both gel-free and gel-based approaches. The number of peptides detected for each protein was used as quantitative value. Protein isoforms and/or sub-units were grouped under the same function. Taxa displaying a proportion <1% were gathered into "Other" category. (**b**) Venn diagrams showing the number of common and unique taxa identified at family level.

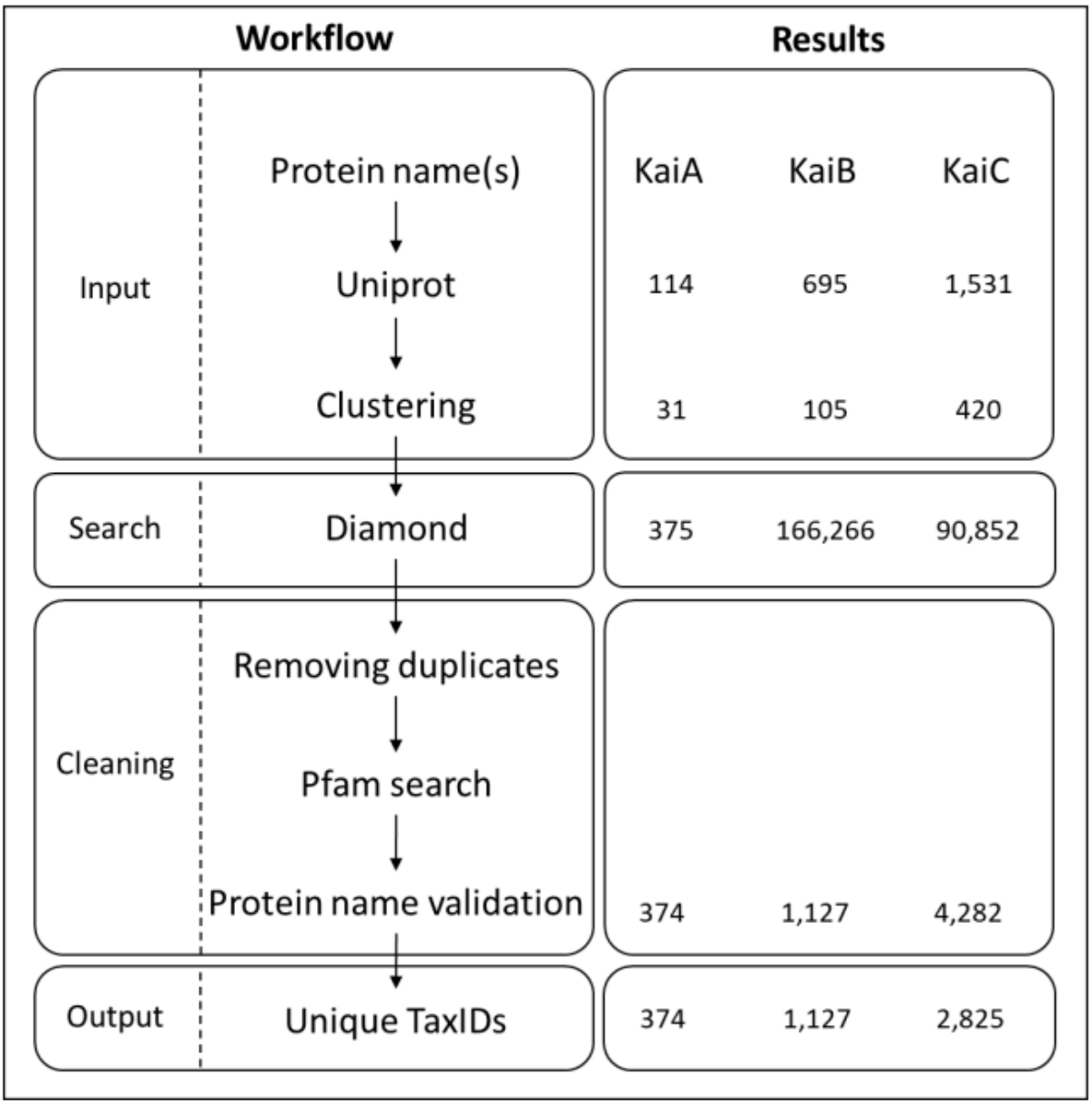


Appendix E: Relative taxonomic composition at genus level depending on the methodology.

(**a**) Relative taxonomic composition at genus level for each methodology. Values represent the proportion of proteins with identical taxonomy on total identified protein using TAX-DB, NAM-DB, AM-DB, or Comb-DB in both gel-free and gel-based approaches. The number of peptides detected for each protein was used as quantitative value. Protein isoforms and/or sub-units were grouped under the same function. Taxa displaying a proportion <1% were gathered into "Other" category. (**b**) Venn diagrams showing the number of common and unique taxa identified at family level.



Appendix F: Bioinformatical workflow of Kai protein homologs analysis in Prokaryotes. Numbers represent the number of protein sequences for each step of the workflow.



Appendix G: Concentration of RNA extracted from each sample in *R. rubrum* grown under light and dark cycle and continuous light.

	Light and dark cycle	Continuous light
Sample	[RNA] (ng/µl)	[RNA] (ng/µl)
16H_A1	31.501	25.429
16H_A3	31.962	17.968
16H_A5	29.486	33.833
00H_B1	97.081	56.628
00H_B3	85.78	51.576
00H_B5	87.648	48.401
08H_C1	73.886	191.68
08H_C3	153.718	95.776
08H_C5	182.46	141.14
16H_D1	370.961	339.105
16H_D3	383.374	250.747
16H_D5	477.855	380.776
00H_E1	692.797	828.615
00H_E3	146.308	617.402
00H_E5	1051.854	603.786
08H_F1	1225.78	691.672
08H_F3	717.75	1530.364
08H_F5	1566.83	2386.183
16H_G1	1028.058	586.754
16H_G3	1707.021	863.985
16H_G5	1070.296	1239.277

Appendix H: Up- and downregulated proteins in *R. rubrum* classified in protein regulation profiles. Up arrow and down arrow correspond to upregulation or downregulation, respectively, in LL compared to LD at each sampling time.

	Regulation profile	48h	56h	64h	72h	80h	88h	96h	Protein number
Upregulated	Single	↗	-	-	-	-	-	-	11
	Single	-	-	-	↗	-	-	-	8
	Single	-	-	-	-	-	-	↗	11
	Consecutive	↗	-	-	↗	-	-	-	1
	Single	↘	-	-	-	-	-	-	10
	Single	-	-	-	↘	-	-	-	19
	Single	-	-	-	-	-	-	↘	26
	Consecutive				↘			↘	2
Downregulated	Single	-	↗	-	-	-	-	-	74
	Single	-	-	-	-	↗	-	-	13
	Consecutive	-	↗	-	-	↗	-	-	1
	Single	-	-	↗	-	-	-	-	13
	Single	-	-	-	-	-	↗	-	19
	Consecutive	-	↗	↗	-	↗	↗	-	2
	Single	-	↘	-	-	-	-	-	107
	Single	-	-	-	-	↘	-	-	42
	Single	-	-	↘	-	-	-	-	10
	Single	-	-	-	-	-	↘	-	48
	Consecutive	-	↘	-	-	↘	-	-	12
	Consecutive	-	-	↘	-	-	↘	-	2

Appendix I: Regulated transcriptional regulatory proteins depending on light conditions in *R. rubrum*.

Transcriptional regulatory proteins

48h

Regulated proteins	FC
Transcriptional regulator, DeoR family	2.13
Transcriptional regulator, XRE family	1.74
Heat-inducible transcription repressor hrcA	0.57

56h

Regulated proteins	FC
Transcriptional regulatory protein	2.21
Transcriptional regulator, MucR family	1.55
Periplasmic binding protein/LacI transcriptional regulator	0.65
Two component transcriptional regulator, winged helix family	0.64
Transcriptional regulator, MarR family	0.62
Probable transcriptional regulatory protein Rru_A1086	0.57
Transcription antitermination protein nusG	0.56
Periplasmic binding protein/LacI transcriptional regulator	0.50
Transcriptional regulator, GntR family	0.29

64h

Regulated proteins	FC
Transcriptional regulatory protein	1.94
Transcriptional regulator, TetR family	1.85
Transcription-repair coupling factor	1.69
Transcriptional regulator, ArsR family	0.38

72h

Regulated proteins	FC
None	

80h

Regulated proteins	FC
Transcriptional regulatory protein	3.80
Periplasmic binding protein/LacI transcriptional regulator	0.61
Periplasmic binding protein/LacI transcriptional regulator	0.60
Periplasmic binding protein/LacI transcriptional regulator	0.56

88h

Regulated proteins	FC
Transcriptional regulatory protein	3.16
Two component transcriptional regulator, winged helix family	1.77
Transcriptional regulator, BadM/Rrf2 family	0.63
Transcriptional regulator BolA	0.63
Transcriptional regulator, XRE family	0.56
Transcriptional regulator, XRE family	0.42
Two component transcriptional regulator, winged helix family	0.30

96h

Regulated proteins	FC
None	

Appendix J: Regulated proteins involved in energy conversion depending on light conditions in *R. rubrum*.

Proteins involved in energy metabolism

48h	
Regulated proteins	**FC**
Cytochrome c, class II	1.69
Ubiquinol-cytochrome C chaperone	1.62

56h	
Regulated proteins	**FC**
NADH-ubiquinone/plastoquinone oxidoreductase, chain 6	4.85
4Fe-4S ferredoxin, iron-sulfur binding	3.50
ATP synthase epsilon chain	1.95
ATP synthase subunit alpha	1.92
Ubiquinone biosynthesis protein COQ7	1.80
NADH-quinone oxidoreductase subunit B	1.67
NADH-quinone oxidoreductase subunit I	1.57
NADH:flavin oxidoreductase/NADH oxidase	1.57
ATP synthase gamma chain	1.50
PfkB	0.64
NADH:flavin oxidoreductase/NADH oxidase	0.61
NADH-plastoquinone oxidoreductase, chain 5	0.59
Cytochrome c, class II	0.59
Ferredoxin--NADP+ reductase	0.58
Chlorophyllide reductase iron protein subunit X	0.57
PfkB	0.54

64h	
Regulated proteins	**FC**
Photosynthetic reaction center, H-chain	2.84
Protoporphyrin IX magnesium-chelatase	0.48
4Fe-4S ferredoxin, iron-sulfur binding	0.17

72h	
Regulated proteins	**FC**
ATP synthase subunit c	9.99
4Fe-4S ferredoxin, iron-sulfur binding	2.03
Chlorophyll synthase	0.63
Photosynthetic reaction center, H-chain	0.37

80h	
Regulated proteins	**FC**
4Fe-4S ferredoxin, iron-sulfur binding	4.63
Cytochrome B561	2.12
Chlorophyllide reductase subunit Z	1.69
Photosynthetic complex assembly protein	1.51
Pyruvate ferredoxin/flavodoxin oxidoreductase	0.61
Ferredoxin	0.59
Cytochrome-c peroxidase	0.57
Cytochrome c2	0.46
Cytochrome c, class II	0.45
Cytochrome c'	0.39

88h	
Regulated proteins	**FC**
Cytochrome C oxidase, mono-heme subunit/FixO	0.66
NADH-quinone oxidoreductase subunit N	0.61
Ferredoxin	0.61
Ubiquinol-cytochrome c reductase iron-sulfur subunit	0.59
Cytochrome-c peroxidase	0.58
Phenylpyruvate decarboxylase	0.57
Chlorophyll synthase	0.56
4Fe-4S ferredoxin, iron-sulfur binding	0.56
Cytochrome c oxidase cbb3-type, subunit I	0.54
PfkB	0.53
2-desacetyl-2-hydroxyethyl bacteriochlorophyllide	0.52
Photosynthetic complex assembly protein	0.51
Cytochrome B561	0.32
4Fe-4S ferredoxin, iron-sulfur binding	0.05

96h	
Regulated proteins	**FC**
Cytochrome C oxidase, mono-heme subunit/FixO	0.66
4Fe-4S ferredoxin, iron-sulfur binding	0.66
Cytochrome c'	0.59
Cytochrome c2	0.58
Cytochrome c, class II	0.58
Cytochrome-c peroxidase	0.55
ATP12 ATPase	0.45
Photosynthetic reaction centre M subunit	0.24
4Fe-4S ferredoxin, iron-sulfur binding	0.10

Appendix K: Regulated stress response proteins depending on light conditions in *R. rubrum*.

Stress response proteins

48h	
Regulated proteins	**FC**
Transcriptional regulator, XRE family	1.74
Heat-inducible transcription repressor hrcA	0.57
Heat shock protein Hsp20	0.34

56h	
Regulated proteins	**FC**
Copper/Zinc superoxide dismutase	4.34
Heat shock protein Hsp20	2.12
Succinate semialdehyde dehydrogenase	1.96
Rubrerythrin	0.66
Glutathione S-transferase family protein	0.63
Stress protein	0.50
Superoxide dismutase	0.35

64h	
Regulated proteins	**FC**
Organic solvent tolerance proteintA-like	4.04
Transcriptional regulator, TetR family	1.85

72h	
Regulated proteins	**FC**
LexA repressor	3.89
Heat shock protein Hsp20	0.53

80h	
Regulated proteins	**FC**
LexA repressor	35.21
Glutathione S-transferase-like	0.60

88h	
Regulated proteins	**FC**
Transcriptional regulator, XRE family	0.56
Heat shock protein Hsp20	0.52
Transcriptional regulator, XRE family	0.42
Glutathione synthetase	0.38

96h	
Glutathione S-transferase-like	14.01
DNA mismatch repair protein MutL	2.24
Rubrerythrin	0.65
Aldehyde dehydrogenase	0.65
Heat shock protein Hsp20	0.53

Appendix L: Regulated proteins involved in viral injection/defense, motility, chemotaxis, nitrogen cycle, and circadian clock depending on light conditions in *R. rubrum*.

Miscellanous

48h

Regulated proteins	FC
CRISPR-associated protein, Cse4 family	1.51

56h

Regulated proteins	FC
Methyl-accepting chemotaxis sensory transducer	2.48
Chemotaxis sensory transducer	1.53
Flagellin-like	0.66
Integration host factor subunit alpha	0.63
Phage shock protein A, PspA	0.60
CRISPR-associated protein, CT1974	0.59
Flagellar basal-body rod protein FlgC	0.52
Nitrogen regulatory protein P-II	0.18

64h

Regulated proteins	FC
Flagellin-like	0.63

72h

Regulated proteins	FC
CRISPR-associated protein, Cse4 family	1.57

80h

Regulated proteins	FC

88h

Regulated proteins	FC
Flagellin-like	0.30
Circadian clock protein KaiC	1.5
Circadian clock protein KaiC	0.6

96h

Regulated proteins	FC
Methyl-accepting chemotaxis sensory transducer	2.18
Chemotaxis response regulator protein-glutamate methylesterase 1	0.62

Diel cycle and circadian clocks in Prokaryotes

Diel cycle imposes daily oscillations in environmental conditions, which temporally structures most ecosystems. One of the most outstanding adaptations to this recurrent phenomenon is probably the emergence of time-keeping mechanisms – circadian clocks – that allow the organisms to synchronize their biological activities with daily variations. While circadian clocks are ubiquitous in Eukaryotes, they are so far only characterized in *Cyanobacteria* within Prokaryotes. However, growing evidence suggests that circadian clocks are widespread in the bacterial and archaeal domains. Studying the impact of diel cycle on Prokaryotes and deciphering their time-keeping mechanisms is of enormous biological importance as bacteria and archaea are crucial in all ecosystems and are essential to human health. Understanding their daily rhythms will undoubtedly provide numerous valuable insights in medical research, environmental sciences, and biotechnology. In this thesis, we assessed for the first time the impact of diel cycle on the functionality of marine picoplanktonic communities using a metaproteomic approach. Metaproteomics is a powerful tool that allows pertinent establishment of phenotype-genotype linkages, but despite its rapid development, this technology still faces many technical challenges that hamper its potential power. In this context, we developed *mPies*, a novel bioinformatic tool that addresses critical bioinformatic issues in the environmental metaproteomics, and we showed the importance of diversifying the experimental workflow for comprehensive metaproteomic studies. This methodological optimization helped us to compare the *in-situ* diel metaproteomic patterns of picoplanktonic communities sampled from the surface north-west Mediterranean Sea. Our data showed a taxa-specific response to diel cycle in the photoautotroph *Synechococcales* and in (photo)-heterotrophic bacteria such as *Flavobacteriales*, *Pelagibacterales* and *Rhodobacterales*. Next, we investigated the taxonomic distribution of circadian clock proteins in Prokaryotes, and we updated the list of bacteria and archaea of great ecological and industrial relevance that could potentially be clock-controlled. Based on this review, we choose to study the diel cycle impact and the circadian genes expression of the biotech promising *Rhodospirillum rubrum*. Our results showed that light and dark cycles induced broad regulation of its proteome and affected the regulation of *kai* genes at both transcriptional and translational levels. Overall, this work evidenced that diel cycle impacts the functionality of prokaryotes and reinforced the hypothesis of time-keeping mechanisms beyond *Cyanobacteria*.

Augustin Géron

Augustin Géron was born in 1994 in Belgium. He has graduated from the University of Mons with a bachelor's degree in Biology in 2015 and a master's degree in Biology of Organisms and Ecology in 2017. During his master's degree, he did an internship in the Chinese Center for Disease and Prevention in Chongqing (China). He received the award for the best master's thesis in Biology that year. Augustin started his doctoral thesis in November 2017 under a cotutelle agreement between the University of Stirling (Scotland) and the University of Mons. He spent two years in both universities, where he studied the diel cycle and the circadian clocks in Prokaryotes.

Université de Mons
20, Place du Parc, B7000 Mons – Belgique
Tél: +32(0)65 373111
Courriel: info.mons@umons.ac.be
www.umons.ac.be

University of Stirling
Stirling, FK9 4LA - Scotland UK
Tel: +44 (0) 1786 473171
Email: information.centre@stir.ac.uk
www.stir.ac.uk

UMONS Université de Mons

UNIVERSITY of STIRLING