Population dynamics, interactions and evolution of marine microbes using genomic approaches

Francisco Latorre Pérez

Instituto de Ciencias del Mar (ICM)

Directores:
Ramiro Logares Haurie
Olivier Jaillon

Tutor:
César Mösso Aranda

Septiembre 2022

Tesis doctoral presentada para la obtención del título de Doctor por la
Universidad Politécnica de Catalunya
Programa de doctorado de Ciencias del Mar

ABSTRACT

There is a myriad of microorganisms on Earth contributing to global biogeochemical cycles. In the surface ocean, the smallest microbes (picoplankton) are responsible for an important fraction of the total atmospheric carbon and nitrogen fixation. The ocean picoplankton encompasses both prokaryotes (bacteria and archaea) and tiny unicellular eukaryotes (protists). Despite their overall importance for the functioning of the biosphere, many questions remain unanswered on their biogeography, population dynamics, interactions, and evolution. Answering these questions is essential in the context of global change, as alterations of the ocean microbiome could impact the function of multiple ecosystems. In this thesis, we aim at reducing the knowledge gap on the above topics through the application of High-Throughput Sequencing (HTS) and genomic approaches, using data collected during the circumglobal *Tara Oceans* and *Malaspina-2010* expeditions, as well as at the Gulf of Maine (North Atlantic), and two Northwestern Mediterranean coastal microbial observatories (BBMO and SOLA stations).

Chapters 1 and 2 are dedicated to a small protistan group of heterotrophic flagellates (HF): the Marine Stramenopiles (MAST)-4, relevant during picoplankton grazing and nutrient remineralization. Due to its widespread distribution and relatively high abundance, MAST-4 has become a target group of microbes to study HF. Unfortunately, MAST-4 remains uncultured. We investigated four evolutionary-related species of MAST-4 (species A, B, C, and E) by reconstructing their genomes with Single-Cell genomics data.

In chapter 1, co-occurrence and biogeographic analyses in the surface global ocean indicated contrasting patterns driven by temperature. Although MAST-4 species were similar in terms of broad metabolic functions, they differed in the set of genes related to the food degradation machinery. We proposed that differential niche adaptation to temperature and prey type has promoted the evolutionary diversification in MAST-4. In chapter 2, we explore the intraspecies genomic divergence within each MAST-4 in the surface global ocean using metagenomic data. We found highly-differentiated populations among MAST-4A and C but lowly-differentiated populations in species B and E. Nevertheless, positive selection in specific genes pointed to niche

adaptation to different ocean basins driven by fluctuating temperature and salinity conditions.

Environmental factors also oscillate over time, but the effect they have over population remains a mystery. In chapter 3, we compared the genomic differentiation of 495 abundant prokaryotic metagenome-assembled genomes (MAGs) over 12 and 7 years in BBMO and SOLA stations, and across the surface global ocean. Our results suggested a stronger population differentiation at large-spatial scales, modulated by temperature and salinity, compared to long-temporal scales. However, underlying population structure was still detected in both time-series, with differential patterns of positive selection within the same genes.

Although protists and prokaryotes are very different in terms of cellular structure, feeding, diversity, and reproduction, they are interconnected through biogeochemical and food web networks. In chapter 4, we developed an approach to infer potential interactions between these two groups based on the annotation of functional genes within Single-Amplified Genomes (SAGs). From a collection of over 3,000 SAGs, we corroborated associations (potential interactions) reported in previous works and inferred new ones involving uncultured marine protists that may hold important roles in ecosystem functioning.

This thesis not only contributes to a better understanding of the biogeography, population dynamics, interactions, and evolution of marine microorganisms, but it also constitutes an expandable resource to test future hypotheses involving the ocean microbiome.

RESUMEN

Existen un sinfín de microorganismos que contribuyen a los ciclos biogeoquímicos globales. En la superficie oceánica, los microbios más pequeños (picoplancton) son responsables de fijar una gran parte del total de carbono y nitrógeno terrestre. El picoplancton agrupa tanto a procariotas (bacterias y arqueas) como a pequeños eucariotas unicelulares (protistas). A pesar de su importancia en el funcionamiento de la biosfera, existen aún muchas preguntas sin respuesta relacionadas con su biogeografía, dinámica poblacional, interacciones y evolución. Responder dichas preguntas es esencial en el contexto del cambio climático, ya que alteraciones en la microbiota marina podría impactar en el funcionamiento de múltiples ecosistemas. El principal objetivo de la tesis es llenar el vacío existente sobre los temas mencionados a través de la aplicación de técnicas de secuenciación de alto rendimiento (HTS) y de métodos genómicos, usando datos recolectados durante las expediciones globales de *Tara Oceans* y *Malaspina-2010*, el Golfo de Maine (Océano Atlántico Norte), y dos observatorios situados en la costa noroeste del mar Mediterráneo (BBMO y SOLA).

Los capítulos 1 y 2 están dedicados a un grupo de pequeños protistas flagelados heterótrofos (HF): los Stramenopilos Marinos (MAST)-4, relevantes en el consumo de picoplancton y la re-mineralización de nutrientes. Debido a su distribución generalizada y su abundancia relativa elevada, MAST-4 se ha convertido en un grupo microbiano modelo para estudiar a los HF. Hemos reconstruido los genomas de cuatro especies relacionadas evolutivamente de MAST-4 (A, B, C, y E) usando datos de genómica de célula única (SCG). En el capítulo 1, los análisis de coocurrencia y biogeografía en la superficie oceánica global indicaron patrones contrastantes relacionados con la temperatura. Aunque las especies de MAST-4 compartían funciones metabólicas, su contenido genético relacionado con la degradación de comida era diferente. Por lo tanto, propusimos que la adaptación a diferentes nichos promovió la evolución de MAST-4. En el capítulo 2, exploramos la divergencia genómica global dentro de cada especie individual usando datos metagenómicos. Encontramos poblaciones altamente diferenciadas en MAST-4A y C, pero poco diferenciadas en las especies B y E. La selección positiva de genes específicos señaló a una adaptación por zona oceánica en base a temperatura y salinidad.

Las condiciones ambientales oscilan en el tiempo, pero el efecto que tienen sobre las poblaciones en la escala temporal es poco conocido. En el capítulo 3, comparamos la diferenciación genómica de 495 genomas de procariotas abundantes en 12 y 7 años de datos temporales en las estaciones de BBMO y SOLA, y a través del océano global. Nuestros resultados sugirieron una diferenciación poblacional mayor en el espacio que en el tiempo modulada por la temperatura y la salinidad. Sin embargo, detectamos estructura poblacional subyacente en ambas series temporales con distintos patrones de selección en los mismos genes.

Aunque los protistas y los procariotas son muy diferentes en cuanto a estructura celular, alimentación, y diversidad, están conectados a través de las redes tróficas y biogeoquímicas. En el capítulo 4, desarrollamos un nuevo método para predecir interacciones potenciales entre estos grupos basado en la anotación de genes funcionales dentro de genomas de célula única (SAGs). Con una colección de más de 3000 SAGs, corroboramos interacciones predichas en trabajos previos y describimos otras nuevas involucrando protistas marinos no cultivados que podrían jugar papeles importantes en el funcionamiento del ecosistema.

En conclusión, la presente tesis no solo contribuye a un mejor entendimiento de la biogeografía, dinámica poblacional, interacciones, y evolución de microorganismos marinos, sino que constituye una fuente de referencia para testear futuras hipótesis que involucre a la microbiota marina.

KEYWORDS:

a) Population dynamics
b) Evolution
c) Interactions
d) Marine microbes
e) Protists
f) High-throughput sequencing
g) Single-cell Genomics
h) Marine ecology
i) Metagenomics
j) Marine environment

AKNOWLEDGMENTS

With sincere gratitude, I would like to acknowledge the support of the principal director, Dr. Ramiro Logares. Thank you for offering me the possibility to work in this project. This Doctoral Thesis at the Institut of Marine Sciences (ICM - CSIC) in the *Ecology Marine Microbes* group was supported by the **Spanish National Program FPI 2016 (BES-2016-076317, MICINN, Spain)**. A big thanks to Dr. Ramiro Logares, Dr. Olivier Jaillon, the co-director for this project, Dr. Ramon Massana, Dr. Pierre Galand and Dr. Rammunas Stepanauskas for providing access to the high-quality datasets used in this project (*Tara Oceans*, *Malaspina-2010*, BBMO and SOLA).

I am thankful for the valuable feedback and proofreading on preliminary and advanced sections of this manuscript from Dr. Pierre Galand, Dr. Ramon Massana, Dr. Ina M. Deutschmann, Dr. Anders K. Krabberød, Dr. Pierre Ramond, Dr. Célio Dias Santos and Dr. Marit F. M. Bjorbækmo.

I would also like to thank Céssar Mösso Aranda and Genoveva Comás, the tutor and secretary of the doctoral program in Marine Sciences at the Politecnic University of Barcelona (UPC), for their guidance and help throughout the years. Your assistance to the administrative works is enormously appreciated.

Many people supported, encouraged, and motivated me during this journey. A big thank you to all my colleagues at the ICM, Lidia, Ina, Sergio, Marta, Marina, Adrià, Pablo, Aleix, and Aurélie for the good memories and laughs.

I am eternally grateful for the magnificent people who supported me throughout the years, but specially at the end of the journey. Thank you, Lluis, Eva, Ina, Lidia, Oscar, my mother, and siblings.

Thank you! Gràcies! ¡Gracias!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

All acronyms and abbreviations presented in this thesis are correctly introduced in the main text. Find below the list of the most recurring abbreviations:

ASV – Amplicon Sequence Variant
BBMO – Blanes Bay Microbial Observatory
DNA – Deoxyribonucleic acid
FST – Fixation Index Distance
GH – Glycoside Hydrolase
GoM – Gulf of Maine
HF – Heterotrophic flagellate
HTS – High-throughput Sequencing
IO – Indian Ocean
LoCoS – Low Coverage Sequencing
MAST – Marine Stramenopile
MS – Mediterranean Sea
NAO – North Atlantic Ocean
NPO – North Pacific Ocean
OTU – Operational Taxonomical Unit
RNA – Ribonucleic acid
RS – Red Sea
SAO – South Atlantic Ocean
SCG – Single-Cell Genomics
SO – Southern Ocean
SOLA – SOMLIT Observatory Laboratoire Arago
SPO – South Pacific Ocean
TARA – Tara Oceans

List of units

ºC – Celsius degrees
PSU – Practical Salinity Unit
µm – Micrometer
km – Kilometer
bp – Base pairs
Kb – Kilobases
Mb – Megabases
Gb – Gigabases

INTRODUCTION

Marine ecosystems

Water is vital for all known forms of life on planet Earth. Many organisms rely on water for food, reproduction, and protection, spending most of their life cycles within aquatic habitats and are known as aquatic organisms. Within aquatic environments, these organisms interact with each other and with a broad variety of physicochemical factors. Aquatic ecosystems are dominated primarily by marine ecosystems (seas and oceans), cover about 71% of Earth's surface and connect the lithosphere to the atmosphere, transferring matter between them (1). Due to the chemical and physical properties of water, a wide range of chemical reactions occur, allowing for life to flourish and prosper (2). Marine ecosystems are also crucial to humanity, as important sources of ecosystem products and services (3), such as food, recreation, and job opportunities.

Oceans play a key role in regulating and maintaining the global climate; they mediate the Earth's temperature and are pivotal to the water cycle (4). The oceans are also the largest carbon storage of the planet (5). Over the last two centuries, the oceans have absorbed 1/3 of the carbon dioxide produced by humans and 90% of the heat trapped in the atmosphere by the increasing concentration of greenhouse gases (6). However, global temperature is still rising, melting the ice in the poles and increasing sea levels, modifying ocean currents and producing more extreme weather conditions (7). Understanding how marine ecosystems operate and how they are changing is of vital importance not only for humans, but for the sustainability of the planet.

Marine ecosystems are inhabited by a myriad of small, microscopic organisms (1), including microbial plankton, which hold important regulatory capabilities (8) and are responsible for approximately 50% of the primary production on Earth (9). Marine microorganisms also play crucial roles in the major biogeochemical cycles, such as nitrogen, phosphorus, sulfide, carbon and silica (10–12). Consequently, changes in the microbial plankton community can alter the planet's biogeochemical cycles at the local and global scale.

Understanding how changes in the biogeochemical cycles affects biodiversity, is of utmost importance for sustaining global ecosystem functioning (13). Because of this, the study of marine ecosystems has become a hot topic over the past decades; the number of research papers containing the words "marine ecosystems" published since

1970 has increased, reaching over seven times more manuscripts per year (*e.g.,* ca. 200 manuscript in the year 2000 and ca. 1,400 in 2013) (14). These articles comprise investigations of many different scientific topics: from biodiversity of aquatic organisms (bacteria, phytoplankton, fishes and mammals) and its functionality (biomass production, food webs dynamics and primary and secondary production) to basic environmental research and its impact (contamination, pollution, human pressure).

Of particular interest is the ecology of marine organisms. Ecology is defined as the scientific study of the distribution and abundance of organisms and the biotic and abiotic interactions that determine their distributions and abundances (15). Ecology aims to understand *where* organisms occur, *how many* of them appear there, and the reason *why*, focusing on four different levels: individual *organisms*, *populations* of the same species, species, and the *communities* that species form in habitats (15). Throughout this thesis, we studied the role of marine microorganisms belonging to different realms of life (Bacteria, Archaea, and Eukarya) along these four levels of ecological complexity. We aimed to explore the many nuances, behaviors, and patterns of marine microbes in the oceans.

Biogeography of marine microorganisms

Biogeography is the study of the distribution of biodiversity over spatiotemporal scales, aiming to unveil where organisms live, at what abundance, and why (16). Aquatic microbial communities are important for global biogeochemical cycles (12), as they include primary (photo- and chemoautotrophs) and secondary producers (heterotrophs) that allow the flux of carbon, nitrogen, oxygen and other elements across food webs. Furthermore, marine microbes can reintroduce into food webs dissolved organic carbon via the microbial loop and channel it into higher trophic levels (11,17,18). Consequently, characterizing the microorganisms present in the community is a crucial step to understand how the marine ecosystem works. Although biogeography as a topic has been studied in many types of animal and plant species before, it was only during the last two decades when biogeography studies of marine microorganism have flourished (8,19–22). The reason behind this is that marine microorganism biogeography studies have to deal with two main issues: challenging sampling environments and difficult microbial identification.

On the one hand, the marine ecosystem is an open and dynamic environment composed of a water column with a number of layers featuring different physical and chemical conditions. Mixing of these layers by oceanic currents and the up- and down-welling, constantly alters the environmental conditions over large and short spatiotemporal scales (23,24). When the environment changes, the microbial community normally changes too (25). Sample collection in these dynamic conditions comes with an underlying risk of under sampling certain organisms (26). On the other hand, even if we were able to sample the complete community at a certain location and timepoint, we would still have to deal with the task of identifying each organism.

Compared to plants or animals, microorganisms are difficult to study due to their small size and lack of morphological features to differentiate them apart. Furthermore, it is difficult to culture the vast majority of marine microorganisms in the laboratory due to the complexity of replicating marine conditions. Because of this, the majority of microorganism studied during the XIX and XX centuries were the ones which were easy to culture and characterize, such as the bacterial genera of *Photobacterium*, *Flavobacterium*, and *Vibrio* (27,28), among others. Identified microbial eukaryotes from this time period included the Radiolaria with their elaborate mineral skeletons (29) and phytoplankton with distinct morphological features such as dinoflagellates with their characteristic flagella, and diatoms with their distinctive silica cell walls (30).

With the advances in isolation technologies, such flow cytometry, and the improvement and reduced cost of High-Throughput Sequencing methods (HTS) and molecular biology, it is now possible to study marine microbial biogeography at a larger and more detailed scale (31–33). In an attempt to bring molecular biology closer to ecology, global sampling and data collection initiatives have been conducted, such as the Global Ocean Sampling Expedition (GOSE) (34) or the International Census of Marine Microbes (ICOMM) (35). Particularly relevant for this thesis are the *Malaspina* expedition (36) and the *Tara Oceans* expedition (8,37) (**Figure 1**). Both expeditions aimed to bring new insights on the biogeography and diversity of marine microorganisms across the global ocean. The *Malaspina-2010* expedition, which was conducted during 2010 and 2011, focused on assessing the biodiversity in the tropical and subtropical oceanic areas from the surface layer down to 4,000 meters deep. Simultaneously, the *Tara Oceans* expedition aimed to characterize fully planktonic ecosystems across all major oceans from 2009 to 2012, focusing in the sunlit waters.

These initiatives increased enormously the amount of available information on microbial diversity, genomics, morphology, and biogeography among others (36,38–40). Specifically relevant for this thesis, these initiatives produced millions of reads of HTS data that can be used to address ecological questions about marine microbial biogeography and diversity (37,41–44).

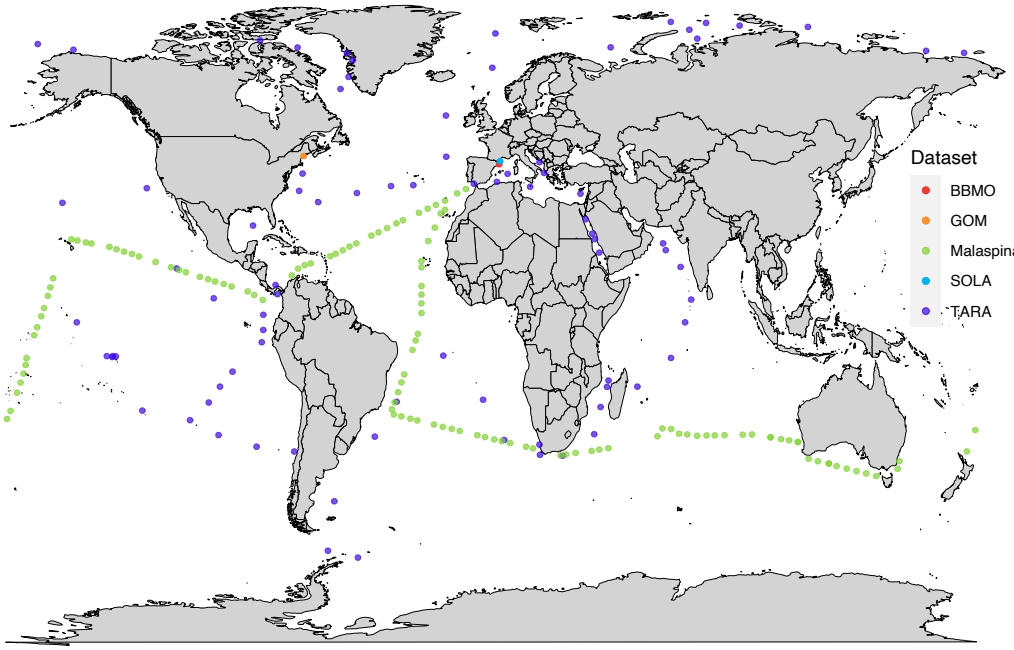

Figure 1. Location of all the sample sites for each of the datasets used in this thesis. Legend: BBMO – Blanes Bay Microbial Observatory; GOM – Gulf of Maine; Malaspina – *Malaspina-2010* expedition; SOLA – SOLA station at Banyuls Bay; TARA – *Tara Oceans* expedition.

The sequencing of the small subunit (SSU) of the ribosomal RNA gene, SSU 16S rRNA for bacteria and archaea and SSU 18S rRNA for eukaryotes, has revolutionized the field of microbiology. Over the last two decades, environmental surveys of SSU rRNA genes have provided a more refined picture of the diversity of both marine bacteria and archaea (45–47), and also revealed the diversity of unculturable marine eukaryotic microorganisms (48–50). This includes an important and relatively abundant group of phylogenetically distinct lineages of eukaryotic plankton in the open ocean (51–53); the Marine Stramenopiles (MAST). Most previous research on marine microbial biogeography attempted to taxonomically identify members of the microbial community, and to estimate their abundance and functionality based on SSU rRNA genes (54). Yet, despite the usefulness of the SSU rRNA, it is only a marker gene. During the last decade, genomic approaches, using the entire or a fraction of the genomes and transcriptomes of microbes have flourished (55–58). This

also includes metaomics studies, that investigate the genomic and transcriptomic information of many microbes simultaneously. During the same period, single-cell genomic approaches have become more popular as well (**Figure 2**) (59,60). These advances, fueled by the widespread use of HTS, have contributed enormously to expand the toolbox of microbial ecologists, generating a revolution in the field (See section "Sampling, molecular and bioinformatic approaches used in this thesis").

Half of this thesis is focused on expanding our knowledge about the biogeography and the processes structuring one particular group of MASTs species: MAST-4. To do so, we integrated together Amplicon Sequence Variant (ASV) data from the *Malaspina-2010* expedition with Single Cell Genomics (SCGs) (sequencing of individual cell DNA), Metagenomics and Metatranscriptomics data (sequencing of whole-community DNA and RNA) from the *Tara Oceans* expedition. We observed contrasting biogeographical distributions between MAST-4 species as a result of niche adaptation to different environments. In particular, through comparative genomics analyses, we suggest that temperature and food degrading capacity are factors shaping the evolutionary history of the MAST-4 linage (61).

For decades, it has been hypothesized that the marine environment is a large and connected ecosystem with few limits to dispersal. Under this assumption, unicellular microorganisms are able to travel through ocean currents and reach all major oceanic basins (62). Some marine microorganisms display wide-spread biogeographical distributions (63,64). The SAR11 clade is one of, if not the most, abundant bacteria in the oceans (45), yet it can be divided into subclades showing distinctive patterns and distributions in the global ocean (65). Although the eukaryotic MAST-4 is also a widespread linage of heterotrophic microorganisms, some species are more abundant than others in different habitats (Chapter 1) (52,61). With the introduction in marine ecology of HTS methodologies and the easy access to sequencing data, it has been reported that microorganisms that once were thought to be the same species are actually genetically different organisms (66) and that observed ubiquitous distributions could also be an artifact of not having enough resolution at the taxonomic level (67). Microorganisms with similar morphological traits and almost identical SSU rRNA genes can possess very different genetic traits and exhibit distinct biogeographical patterns and roles across environments. Consequently, ecotypes and populations need to be taken into consideration when studying the biogeography of a given organism.

Population genomics of marine microorganisms

The marine ecosystem is a very dynamic environment that experiences spatial and temporal variations (68). Temperature, salinity, nutrient and sunlight availability, and other physicochemical features, are highly different between locations, depths, and at different times of the year (69,70). Although the marine ecosystem is connected between the two poles and microbes can travel throughout the oceanic basins, changes in environmental conditions trigger the selection of better adapted organisms to specific environments (71). Traits that increase the fitness of microbial ecotypes will be passed down to their descendants, increasing survival and reproduction, and leading to adaptation (3,72,73), following the premises of the theory of evolution by *natural selection*, proposed by Charles Darwin in 1859. Although this theory may suggest that organisms are adapted to present environments, it is more accurate to say that natural selection shaped them to the past environments that exerted selection. At the same time, and because environments are in constant change, the present environment is providing new natural forces that will influence the future evolution of these organisms.

Compared to macro-organisms, microorganisms have a short generation time and can rapidly produce random mutations (74,75). Individuals from a particular species are not necessarily genetically identical, although sometimes they can appear so under the microscope, and they could have originated via asexual reproduction. When a microbe encounters a new environment, only those individuals that through random mutation developed a beneficial trait that increase their fitness in the new environment will survive and reproduce. These individuals of the same species that differ genetically are categorized into populations based on the showcased amount of genetic difference (Fixation Index, FST). In a nutshell, the FST allows quantifying whether different individuals of the same species belong or not to the same population based on genetic differences.

Populations are the desired taxonomical unit to assess diversity, ecosystem stability and flexibility, and ecological interactions (76,77). However, defining microbial populations is a difficult task, as most microbes do not provide enough resolution in morphological traits that would allow to define populations, and genotyping many microbial cells is a laborious task that only became feasible during the last two decades thanks to HTS. Furthermore, even when genotyping is available,

deciding when two populations are genetically divergent enough to be considered as two distinct entities is complicated due to a lack of standard FST threshold for microbes (77). In this thesis we extrapolate FST thresholds used in plant and animals (78,79) to define genomic populations of marine microorganisms (FST > 0.15).

The complexity of defining microbial populations increases with the lack of complete genomes for most microbes. Even though our knowledge of marine microbial diversity has increased with 16S and 18S SSU rRNA gene surveys, the truth is that in most cases, it does not provide enough resolution to identify populations (77). As random mutations can affect any gene of an organism, complete genomes are needed to investigate population genomics. Nowadays, with the improved and increased number of high-throughput genomic analyses and technologies (see section "Sampling, molecular and bioinformatic approaches used in this thesis"), more genomes for rare and uncultured microorganisms have become available (80–82).

The application of genomic technologies to the study of populations is what defines the field of *population genomics*. In particular, metagenomics, defined as a culture-independent method for the identification and characterization of all microorganisms and their genetic content (83), allows us to gain a more detailed knowledge of marine microbial ecosystems (31,32,82) and to assess population genomics in more depth (84). Specific metagenomic surveys have targeted microorganisms not only to study biodiversity and metabolic functions, but also to describe the role of populations in biochemical cycles and ecological processes. One of the first whole-genome shotgun sequencing studies in marine environments elucidated the gene content, diversity, and relative abundance of the organisms in the sampled locations. It discovered 148 unknown bacterial phylotypes, more than 1 million unknown genes and 782 rhodopsin-like photoreceptors, which altogether allowed for a better understanding of microbial photosynthesis in the ocean (85). Other metagenomic studies have shown how planktonic communities of marine eukaryotes, prokaryotes and viruses drive the carbon fixation in oligotrophic waters (86) and the sinking of organic matter into the deep ocean layers (87).

The search for genomic variants (alterations in the DNA of an organism) has become a standard procedure to determine genomic populations (88,89). These genomic variants, such as Single Nucleotide Polymorphisms (SNPs) or insertions and deletions

of genomic regions (INDELS), are the result of mutations due to errors during the replication of the DNA (90,91) and are used to measure genetic divergence, evolutionary history, natural selection, and gene flow (75). A few examples of SNP discovery using omic techniques in marine organisms include macrofauna such as the Pacific oyster (92) or the three-spined stickleback (93). Other genetic divergence studies based on genomic variants were successful in defining genomic populations within marine microorganisms using metagenomic samples. For example, the diatoms *Picea pungens* (94) and *Thalassiosira rotula* (95) show well defined populations with very contrasting genetic divergence patterns as a consequence of different limitations to gene flow. Similar studies analyzing population structure and its relationship with the environment have been carried out in prokaryotic communities (89) and in individual species (88), where temperature, salinity, and depth appeared to be the main drivers shaping population structure. Although some advances in population genomics of eukaryotic marine microbes have been made in recent years (96), our knowledge about their populations is still very shallow.

Environmental conditions are key to understanding population dynamics and structure. For this reason, global ocean metagenomics surveys (*e.g., Tara Oceans*) are particularly useful in population genomics studies since they include samples from a wide range of different conditions. However, the marine environment also fluctuates over time and along the different seasons of the year, and therefore genomic populations can emerge at this scale too. Along with spatial variation, analyzing how populations change over time is also a relevant topic to understand the marine ecosystem as a whole.

Over the past decades, there have been efforts to build long time-series of marine microbial and environmental data at specific locations to study how the environment and the microbial community change over the years. Some examples of long time-series include the Hawaii Ocean Time-series (HOT), which has investigated temporal dynamics in microbial ecology, chemistry and physics at the ALOHA station in the North Pacific Subtropical Gyre (NPSG) since 1988 (97); the Bermuda Atlantic Time-series Study (BATS) that collected data on the physical, biological and chemical properties every month since 1988 (98); or the MareChiara time-series that started to unveil aspects of zooplankton temporal evolution and recurrences in 1984 in the Gulf of Naples (99).

For this thesis, metagenomic data from two Mediterranean time-series (~ 130 km apart) have been used along with the large spatial metagenomic dataset from *Tara Oceans* (8,41) to assess the population structure and adaptations of marine microbes over spatiotemporal scales. The two time-series are located at the Blanes Bay Microbial Observatory (BBMO, Blanes, Spain) (100) and the SOMLIT Observatory Laboratoire Arago (SOLA, Banyuls sur Mer, France) (101) respectively (**Figure 1**). Microbial data from BBMO has been sampled periodically each month since 2001. A total of 140 surface-water metagenomes were produced between January 2009 and December 2020 (12 years). Similarly, 90 surface-water metagenomic samples (7 years) were produced in SOLA between January 2009 and December 2015. More than 1,500 bacterial and archaeal Metagenome Assembled Genomes (MAGs) were produced from 7 years of metagenomic data at the BBMO, which, along the Single-Cell Genomics data from *Tara Oceans* (58,61,81,102), were used to investigate different population dynamics at the global ocean and compared them to 12 and 7 years of monthly data. Results from this thesis show that genomic populations of prokaryotic MAGs were strongly differentiated and structured by temperature and salinity in the global ocean. In turn, population differentiation over long time scales was genome-specific and could either be strong or weak. Still, population structure over time was highly influenced by seasonal environmental changes (Chapter 3).

Microbial interactions within marine communities

Above, we mentioned how abiotic factors could structure microbial species and populations. Yet, biotic interactions could also determine the biogeography of marine microbes, and despite their importance, they remain for the most part unknown. Microbial communities are composed of a wide diversity of species that are susceptible to changes in abundance based on microbe-microbe interactions (predation, symbiosis, parasitism, commensalism, amensalism) (103,104). As part of the microbial loop, microbes assimilate and process the dissolved organic carbon (DOC), which can then be channeled into higher trophic levels via predation (105). The mechanism that defines how carbon and nutrients flow between trophic levels is through the interaction of different species; photoautotrophs are eaten by heterotrophic microorganisms, which at the same time are a food source for higher trophic-level organisms. Recently, the Protist Interaction DAtabase (PIDA) was assembled with the objective to document all

published ecological interactions occurring between marine protists and other organisms down to the species level. PIDA collected protists interactions published between 1894 and 2017, including parasitism, predation, mutualism and commensalism (106). Still, most of the documented interactions in the database only reflect a small proportion of the actual interactions in marine environments and most remain unknown.

With the application of HTS technologies, it is now possible to perform co-occurrence analyses from detailed microbial community data in order to produce association networks that represent hypothesis of ecological interactions between microorganisms (103,104,107). Within a network, one organism or species is represented by a node, while an edge indicates a relationship between two nodes. These relationships can be positive (co-occurrence) or negative (exclusion) depending on their abundance patterns. A positive relationship might point to a mutualism or parasitism relationship, while a negative one might indicate competition or predation.

In marine environments, co-occurrence networks have been built using SSU 16S and 18S rRNA markers. For example, associations networks in the San Pedro Ocean Time-Series (SPOT) (California) (108) inferred negative associations between bacteria and protists pointing to predatory relationships, and between photosynthetic organisms (*Ostreococcus* and *Synechococcus*) suggesting competition. Moreover, SAR11 was described as a highly connected species. Another example of association networks are the temporal networks from the Blanes Bay Microbial Observatory in the Mediterranean Sea, which show more association partners during winter and spring compared to summer and autumn (104,109). Here, in Chapter 1, we constructed a network based on ASV data of MAST-4 species from the *Malaspina-2010* expedition (**Figure 1**). We detected a positive association between MAST-4B and C, suggesting mutualism, and a negative interaction of these two species with MAST-4A, possibly indicating competition (61).

Constructing interaction networks is also possible using metagenomic data. Within the *Tara Oceans* project, planktonic networks showed that most interactions are held by dinoflagellates and arthropods, such as the interaction between *Flavobacteria* and diatoms (110). Although correlation networks are useful tools to predict the dynamics and structure of marine microbial communities, they still have limitations (103). Correlation does not always imply an actual interaction, for example, when two

organisms' distributions are driven by the same environmental factor, their patterns of abundance might correlate, showing a positive association. However, this association does not necessarily mean symbiosis but likely represents similar niche (104). One approach to solve this issue is to use SCG techniques (59,60) to detect DNA from different organisms within one isolated cell, which would indicate a predatory, parasitic or symbiotic relationship.

Single-cell data has proven to be powerful for detecting bacterial (111) and viral interactions (112) within protists (113) and bacterioplankton (114) in marine environments. In order to explore ecological interactions involving eukaryotic marine microbes, in Chapter 4, we studied single-cell-based interaction networks between eukaryotic and prokaryotic microorganisms constructed with more than 3,000 eukaryotic Single Amplified Genomes (SAGs) from a few *Tara Oceans* stations in the Mediterranean Sea and Indian Ocean, the Blanes Bay Microbial Observatory, and the Gulf of Maine (**Figure 1**). We developed a new approach using functional genes to assess potential interactions. Our method successfully predicted interactions between protists and other microorganisms, including both prokaryotic (ca. 700 strong interactions) and eukaryotic species (ca. 500). In particular, common protist interactions with bacteria involved Alpha- and Gammaproteobacteria, Bacteroidota, Verrucomicrobiota, and Planctomycetota; while frequent interactions between protists and eukaryotes concerned Dinophyceae, Cryptophyceae, and Haptista, among others. A fraction of the observed potential interactions was common with other studies, either by being predicted by association networks or by culture studies. Yet, many inferred potential interactions, especially those involving uncultured protists (*e.g.,* MASTs), have not been reported previously.

Sampling, molecular and bioinformatic approaches used in this thesis

Amplicon Sequence Variants

High-Throughput Sequencing (HST) of a small region (~ 400 base pairs) of the 16S or 18 rRNA genes (amplicons) has become the method of choice for characterizing microbial communities (85). This approach consists of (a) extraction of DNA from environmental samples, (b) amplification of the target DNA region using the

Polymerase Chain Reaction (PCR) with a pair of primers that target the desired conserved sequence at both ends of the amplicon, (c) using unique labels (barcoding) in each sample for identification, and (d) high throughput sequencing of the amplicons (115).

Amplicon Sequence Variant (ASV) data in this thesis was obtained from surface waters (3 m depth) from a total of 120 globally distributed stations located in the tropical and sub-tropical ocean from the *Malaspina-2010* expedition. Water samples were collected with a 20 L Niskin bottle and filtered as explained in Chapter 1 (see section 1.2 Methods). DNA extraction was performed using standard phenol-chloroform protocols. Both the 16S region V4-V5, and the 18S region V4, of the rRNA gene were amplified from the same DNA extracts. Amplification was conducted with QIAGEN HotStar Taq master mix and amplicon libraries were paired-end sequenced on an *Illumina* MiSeq platform as explained in Logares *et al.,* (44). After sequencing, barcodes and low-quality amplicons are removed, and sequences are trimmed before downstream analyses. Amplicon Sequence Variants (ASVs) are then delineated using DADA2, which is a denoising algorithm that removes errors and can be used to remove chimeras (116).

Single-Cell Genomics

Single-Cell Genomics (SCG) emerged as a complementary tool to microbial cultivation by providing genomic information from individual and uncultured cells. Some powerful examples that demonstrate to which extent SCG is effective in obtaining genomes from uncultured microorganisms and unveiling their metabolic potential are the discovery of chemolithoautotrophy pathways in uncultured Proteobacteria (117); the identification of poorly understood Verrucomicrobia phylum and its capacity to degrade polysaccharides (118); or the many studies in which genomes from uncultured Marine Stramenopiles (MASTs) have been recovered and studied (58,61,81,102).

Overall, SCG consists of a series of integrated steps, beginning with sample collection and preservation, followed by physical cell isolation, lysis and whole genome amplification (WGA) of individual cells, and finishing with whole-genome sequencing and the consequent down-stream analyses (59,60) (**Figure 2**). Deep freezing with glycine betaine or glycerol is the most common approach for the preservation of

environmental samples with minimal loss of the cell integrity and its content (117,119). Then, cells are isolated by fluorescence activated cell sorting (FACS) into microwell plates and labeled in order to continue with cell lysis, which is usually performed with alkaline solutions, to make the genomic content of the cell available for WGA (59,119). Multiple displacement amplification (MDA) is the most widely used approach to produce long sequences suited for *de novo* genomic assembly. Despite being the most widely used method, MDA produces uneven genome coverage. More recent methodologies use a variation of MDA that utilizes a thermostable mutant of the phi29 polymerase that improves genome recovery (WGA-X) (120). In any case, the product of WGA is a Single Amplified Genome (SAG) that, after genomic assembly, can be passed to further down-stream analyses in similar ways to genomes from pure cultures.

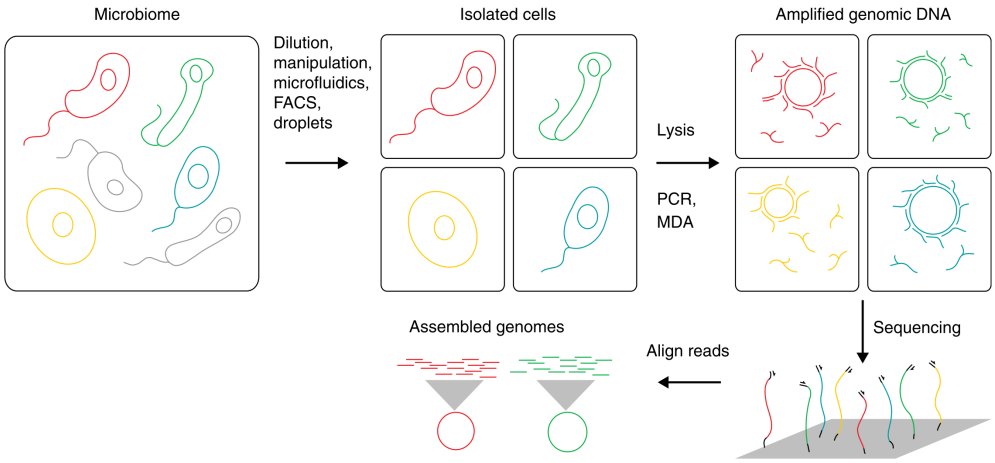

Figure 2.Single-cell genomics overview. Experimental steps include (upper half) isolation and lysis of single cells with subsequent amplification of their genomes, followed by (lower half) high-throughput sequencing and genome assembly. Legend: FACS - fluorescence-activated cell sorting; MDA - multiple displacement amplification; PCR - polymerase chain reaction. Adapted from (Tolonen and Xavier, 2017) *(121)*.

The SCG genomic data used in this thesis was collected from different sampling efforts: the *Tara Oceans* expedition (stations from the Mediterranean Sea and the Indian Ocean), the Blanes Bay Microbial Observatory (Northwest Mediterranean Sea) and the Gulf of Maine (North Atlantic Ocean). SCG genomic sequencing was performed as mentioned above using regular MDA amplification for deep-sequenced SAGs and WGA-X amplification for Low Coverage Sequencing (more details in the Methods section of Chapter 1 and 4).

Metagenomics and Metatranscriptomics

Metagenomics and metatransciptomics involve all the genetic material (DNA and RNA) recovered from environmental samples and have been widely used to explore diversity and structure of microbial communities (80,82,122,123). Recent advantages in HTS technologies and computational approaches have allowed for the reconstruction of metagenome-assembled genomes (MAGs) from metagenomic samples. A MAG is a group of DNA sequences that share similar characteristics and that represent a microbial genome. To achieve this, first, sequencing reads from metagenomic samples are assembled into contigs and scaffolds, and then, the scaffolds are grouped (binning) into MAGs based on tetranucleotide frequencies, abundances and codon usage (124). Often, this approach produces a great number of MAGs that require further quality checks and curation to select those with relatively high quality, as contamination from other microbials genomes can be introduced during the binning process (125). Similar to SCG, the produced MAGs can be analyzed in similar ways to genomes from pure cultures.

Global ocean metagenomes and metatranscriptomes used in this thesis were obtained from water samples collected during the *Tara Oceans* expedition for either the 0.22 – 3 µm fraction, the 0.8 – 5 µm fraction, or both. Temporal metagenomic data was collected at the BBMO and SOLA stations during 12 and 7 years of monthly samples (more details in the Methods section of Chapters 1, 2, and 3).

Aims of the thesis

The main aim of this thesis is to explore the relationships between microorganisms and the environment that they inhabit along the mentioned four levels of organization: organisms, populations, species, and communities. In particular, we aim to explore the population dynamics, interactions, and evolution of marine microbes across the global ocean and over long periods of time (12-7 years). We attempt to reach this objective by using state-of-the-art high throughput sequencing technologies, as well as bioinformatic and high-performance computing methods.

The achievement of this goal is structured in four chapters. The first chapter (*Niche adaptation promoted the evolutionary diversification of tiny ocean predators,* PNAS 2021) is designed as a first global description of the diversity of a monophyletic linage of uncultured marine protists (MAST-4) and the processes behind their contrasting distributions in the global ocean, focusing on the interspecies genomic divergence. This chapter functions as an introduction on how to use high throughput single-cell genomics to obtain curated genomes of individual protists and what biological answers can be assessed with them.

In the second chapter (*Global population structure of a unicellular marine predator*, unpublished), we further study the intraspecies divergence across the surface global ocean of the MAST-4 genomes reconstructed in the first chapter. This chapter serves as a first introduction on the usage of a large metagenomic dataset to assess population genomics and the environmental factors driving population structure and dynamics in a marine protist.

In the third chapter (*Microbial population structure over a decade and across the global ocean*, unpublished), we extrapolate the approaches used in chapter 2 to study population dynamics of abundant marine prokaryotes across the global ocean and over 12 and 7 years of temporal data in the Mediterranean Sea. This chapter aims at describing and comparing the general patterns of genomic differentiation over spatiotemporal scales and detecting what environmental factors shape the overall population structure of the studied marine microbes.

In the fourth chapter (*Investigating the marine protist interactome using Single-cell genomics*, unpublished) we aim at assessing the current state of the marine

interactome at some specific ocean locations considering marine protists and their interactions with other microorganisms, including bacteria, archaea and other eukaryotes. In particular, we focus on retrieving interactions related to predation, parasitism, and symbiosis, which are key relationships that shape food web dynamics in the marine environment.

The outline of the different topics studied can be linked to four general objectives:

Objective 1. Obtain and assess the interspecies genomic divergence of related uncultured protistan species, their adaptation to different environments and their evolutionary history that led to niche diversification.

Objective 2. Describe the intraspecies genomic divergence across the global ocean in an uncultured protistan species (same as objective 1). In particular, to describe the overall patterns of population structure and their links to adaptation to different environments.

Objective 3. Assess the amount of genomic diversity and its structure in selected marine prokaryotic species in the global ocean and over 12 and 7 years in the Mediterranean Sea. Specifically, to find main patterns of population structure and the environmental factors shaping them.

Objective 4. Describe and corroborate interactions that occur within marine protists using Single Cell Genomics data, focusing on identifying who are the microorganisms interacting and the established relationship between them.

CHAPTER 1

Niche adaptation promoted the evolutionary diversification of tiny ocean predators

Francisco Latorre[a], Ina M. Deutschmann[a], Aurelie Labarre[a], Aleix Obiol[a], Anders Krabberød[b], Eric Pelletier[c,d], Michael E. Sieracki[e], Corinne Cruaud[f], Olivier Jaillon[c,d], Ramon Massana[a], Ramiro Logares[a]

[a] Institute of Marine Sciences (ICM), CSIC, Barcelona, E-08003, Catalonia, Spain

[b] Department of Biosciences, Section for Genetics and Evolutionary Biology (Evogene), University of Oslo, Oslo, N-0316, Norway

[c] Metabolic Genomics, Genoscope, Institut de Biologie François Jacob, CEA, CNRS, Univ Evry, Université Paris Saclay, 91000 Evry, France

[d] Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022 / Tara Oceans GOSEE, 3 rue Michel-Ange, 75016 Paris, France

[e] National Science Foundation, 2415 Eisenhower Ave., Alexandria, VA 22314, U.S.A.

[f] Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, 91000 Evry, France.

Published at **Proceedings of the National Academy of Sciences** (**PNAS**) (2021)
DOI: https://doi.org/10.1073/pnas.2020955118

1.1. INTRODUCTION

Ocean microbes are fundamental for the functioning of the Earth ecosystem, playing prominent roles in the global cycling of carbon and nutrients (9). In particular, small phototrophic microbes are responsible for ~50% of the primary production on the planet (126). In turn, heterotrophic microbes have a fundamental role in nutrient cycling and food-web dynamics (127). Heterotrophic flagellates, along with marine viruses, maintain prokaryotic and eukaryotic picoplankton at relatively stable abundances (128). At the same time, they transfer part of the organic matter they consume from lower to upper trophic levels, thus being a key component at the base of the ocean's food web.

Among heterotrophic flagellates, Marine Stramenopiles (MASTs) play a prominent role in unicellular trophic interactions in the global ocean (129). MASTs are polyphyletic, including so far 18 subgroups (53). Except for a handful of strains, MASTs remain uncultured (130), which complicates the study of their cell physiology, ecology, and genomics. Studies using FISH (52,131,132) and metabarcoding (129,133) helped to determine MAST cell sizes (2-5 μm), vertical and horizontal distributions in the ocean, as well as metabolic activity. Further studies linked MAST's cell morphology with environmental heterogeneity, for example, MAST-1B cell size varies with temperature (131). Other studies provided insights into the predatory behaviors of some MAST groups. For instance, MAST-4 prey on *Synechococcus* (129) and SAR11 (111), two of the most abundant microorganisms in the ocean (45,103).

MAST-4 is a prominent clade within the MASTs, featuring small cells (2-3 μm), high relative abundance in comparison to other heterotrophic flagellates, and worldwide distribution (134). Due to these characteristics, MAST-4 can be considered as a model heterotrophic flagellate. MAST-4 is constituted by at least 6 recognized species: MAST-4A/B/C/D/E/F based on 18S rRNA gene phylogenies (53). The biogeography of specific MAST-4 species has been partially elucidated: MAST-4 A, B, and C occur in temperate and warm waters (17 – 30 ºC), whereas species E is typically found in colder waters (2 – 17 ºC) (102,135). This suggests that MAST-4 species have adapted to a different niche temperature. MAST-4 biogeography could also be controlled by bottom-up or top-down biotic factors, such as prey/food-availability (*e.g.,* bacteria, algae, Dissolved Organic Carbon) or predation respectively. Several studies have pointed to a positive correlation between the abundances of prokaryotic and heterotrophic flagellates

(128,136–138). Yet, it is unclear to what extent such biotic relationships can generate biogeography in MAST-4.

Biogeographic patterns of MAST-4 species can provide insights into the drivers that have promoted their evolutionary diversification. Identifying species-specific gene-functions, genes, or gene variants may point to differential adaptations conferring higher fitness in specific biotic or abiotic conditions. In a bacterivorous flagellate like MAST-4, a first approach for assessing species-specific adaptations is to analyze Ecologically Relevant Genes (ERGs), which are those that could reflect associations with environmental heterogeneity or different ecological roles. Candidate ERGs include the enzymes present in the lysosome that are involved in the digestive processes that follow phagocytosis, allowing the degradation of a wide variety of substances such as proteins, carbohydrates, or nucleic acids among others (139). In heterotrophic flagellates, lysosomal enzymes are of particular relevance because different suites could potentially be associated with the degradation of different food items. Among them, Glycoside Hydrolases (GHs), commonly found in lysosomes, catalyze the hydrolysis of glycosidic bonds in complex sugars, allowing the cell to digest other organisms. For example, lysozyme (N-acetylmuramide glycanhydrolase) is a well-known enzyme under the GH category that catalyzes the breakdown of the peptidoglycan cell wall found in bacteria (140). Other studies have shown that each MAST lineage may have a different functional profile in terms of organic matter processing (102).

Genomes are key to obtain ERGs from a species. Common genome sequencing protocols require thousands if not millions of cells, however recovering this number of cells from uncultured protists such as MAST-4 is an almost impossible task. This issue is circumvented with Single-Cell Genomics (SCG) (59,103). The principles of this method consist in isolating single cells using, for example, flow cytometry, lysing the cells, and amplifying and sequencing their genomes producing Single Amplified Genomes (SAGs). In previous work, Single-cell genomics allowed the recovery of ~20% of the genomes from individual MAST-4 cells, which increased to ~80% genome recovery when genomes from different cells were co-assembled (58,81,102). Here, we use the SAG collection produced by the *Tara Oceans* expedition (141), which generated 900 SAGs from 8 stations in the Indian Ocean and the Mediterranean Sea. We compiled the largest collection to date of MAST-4 SAGs, totaling 69 SAGs (23 MAST-4A, 9 MAST-4B, 20 MAST-4C, and 17 MAST-4E). Using this novel dataset, together with

other large metaomics datasets (metabarcoding, metagenomics, and metatranscriptomics) from the *Tara Oceans* and *Malaspina-2010* expeditions (36) we address the following questions: How different are MAST-4 species at the genome level? Did MAST-4 species diverge via niche adaptation? If so, is such adaptation reflected in their genomes and potential ecological interactions? Can ERG composition and expression provide insights on MAST-4 niche diversification?

1.2. METHODS

1.2.1. Geographic distribution of MAST-4 species and association patterns

The distribution of MAST-4 species as well as their association patterns were investigated using metabarcoding based on data from Logares *et al.,* (44). This dataset includes surface water samples (3 m depth) from a total of 120 globally-distributed stations located in the tropical and sub-tropical ocean that were sampled as part of the Malaspina 2010 expedition (36). Samples were obtained with a 20 L Niskin bottle deployed simultaneously to a CTD profiler that measured conductivity, temperature, oxygen, fluorescence, and turbidity. About 12 L of seawater were filtered to recover the smallest organismal size-fraction (0.2 - 3 μm; picoplankton). The concentration of inorganic nutrients (NO3-, NO2-, PO43-, SiO2) were included in our analyses (see Logares *et al.,* (44) for details on their measurement).

Both the 18S (V4 region (142)) and 16S (V4-V5 region (143)) rRNA-genes were analyzed. Operational Taxonomic Units (OTUs) were delineated as Amplicon Sequence Variants (ASV) using DADA2 (116) and OTU tables were generated. Amplifications were performed with QIAGEN HotStar Taq master mix (Qigen Inc., Valencia, CA, USA). Amplicon libraries were paired-end sequenced using Illumina MiSeq (2 x 250 bp) at the Research and Testing Laboratory facility (see Logares *et al.,* (44) for more details). We trimmed the 18S forward reads at 240 bp and the reverse reads at 180 bp, while for the 16S, forward reads were trimmed at 220 bp and reverse reads at 200 bp. Then, for the 18S, the maxEE was set to 7 and 8 for the forward and reverse reads respectively, while for the 16S, the maxEE was set to 2 for the forward reads and 4 for the reverse reads. OTUs were assigned taxonomy using the naïve Bayesian classifier method (144) together with the SILVA v132 database (145) as implemented in DADA2. Eukaryotic OTUs were also BLASTed against the Protist

Ribosomal Reference database (PR2, version 4.11.1 (146)). Streptophyta, Metazoa, nucleomorphs, chloroplasts, and mitochondria were removed from OTU tables.

To infer associations between OTUs we used eukaryotic and prokaryotic OTUs with total abundances >100 reads and occurrences >15% of the samples. All abundances were centered log-ratio (clr) transformed. Associations between OTUs were inferred using Maximal Information Coefficient (MIC) analyses as implemented in MICtools (147), which estimates the total information coefficient TICe and the maximal information coefficient MICe. TICe is used to estimate significant relationships, while their strength is calculated with MICe. TICe null distributions were estimated using 200,000 permutations and the significance level was set to 0.001 as suggested by Weiss *et al.,* (148). MICe = 0 indicates no association between OTUs, while MICe = 1 indicates strong association. Environmentally-driven associations between OTUs were detected and removed using EnDED (104,149), with the methods Interaction Information and Data Processing Inequality. Furthermore, to account for data sparsity and the consequential correlations between zeros in the dataset, we removed associations between OTUs that were not present in ≥ 50% of the samples, *i.e.,* less than half of the samples contained at least one of the two OTUs. We determined the Jaccard index for each association based on the presence of OTUs in the samples (intersection divided by union). We removed associations that featured a Jaccard index below 0.25. Moreover, only associations with MICe > 0.4 were considered. We used the Pearson and Spearman correlation coefficients to analyze the association type: positive Pearson or Spearman correlation coefficients point to co-occurrences, while negative values point to mutual exclusions. The distribution of OTUs across sea temperatures was explored using the *niche.val* function in the EcolUtils package (150). The abundance-weighted mean temperature was calculated for each OTU and used as an estimate of its temperature niche. We checked whether the obtained abundance-weighted mean temperature for each OTU was significantly different from chance ($p<0.05$) using a null model with 1,000 randomizations.

1.2.2. Genome reconstruction using Single Amplified Genomes

Plankton samples were collected during the circumglobal Tara Oceans expedition and cryopreserved as described elsewhere (60). Individual picoplankton cells were isolated from water samples and stained with 1x SYBR Green I (Life Technologies Corporation)

(42,81) using a MoFlo (Dako Cytomation Carpinteria, CA, USA) flow cytometer equipped with the CyClone robotic arm for sorting into plates of 384 wells. Cells were lysed and their DNA denatured using cold KOH. The genome from each single cell was amplified using Multiple Displacement Amplification (MDA) based on the Phi29 polymerase (RepliPHITM, Epicentre Biotechnologies, Madison, WI, USA) (111,151). All single-cell work was performed at the Single Cell Genomics Center (https://scgc.bigelow.org). The obtained SAGs were taxonomically screened by PCR amplification and Sanger sequencing of the 18S rRNA gene using universal eukaryotic primers. A total of 69 SAGs affiliating to MAST-4 species A/B/C/E were selected for downstream analyses. Each selected MAST-4 SAG was sequenced in 1/8 of a lane using either Illumina HiSeq2000 or HiSeq4000 at either the Oregon Health & Science University (USA) or the French National Sequencing Center (Genoscope, France). A total of 424.1 Gb of sequencing data was produced, averaging 6.1 (± 0.22) Gb per SAG. For each SAG, sampling location, depth, and date are reported in **Annex A Table 1**.

Each SAG was de novo assembled using SPAdes 3.10 (152) in single-cell mode "-sc" with default parameters. Contigs shorter than 1kbp were discarded. Quality control and general assembly statistics were computed with Quast v4.5 (153). Estimation of genome recovery was calculated with BUSCO v3 (Benchmarking Universal Single-Copy Orthologs) (154) using the Eukaryota_odb9 dataset (**Annex A Table 2**). SAGs were also co-assembled to increase genome recovery. Only SAGs belonging to putatively the same species were co-assembled. Thus, SAGs had to fulfill three conditions to be co-assembled: First, their 18S rRNA-gene amplicon needed to be >99.5% similar. Second, their Average Nucleotide Identity (ANI) had to be >95%; ANI was computed using Enveomics (155) with the full-length contigs of all SAGs within each species. Third, SAGs had to display a homogeneous composition in Emergent Self-Organizing Maps (ESOM) (156) based on tetranucleotide frequencies. Tetranucleotide frequencies were computed using a 4 bp sliding window and 1 bp step length in fragmented contigs between 2.5 and 5 kbp in size considering both DNA strands and were subsequently clustered using ESOM. Raw data were normalized using robust estimates of mean and variance ("Robust ZT" option) and trained with the k-Batch algorithm and Euclidean grid distance. If fragments from a given SAG were mixed with those from another SAG in tetranucleotide ESOM representations, it indicated that their genomes were similar. SAGs fulfilling the previous three criteria

were considered to belong to the same species and were subsequently co-assembled. Three MAST-4C SAGs (AB536_E17, AB536_F22, AB536_M21) showed more genomic divergence (ANI ~93%) compared to the others but were still included in the final co-assembly because the 18S and tetranucleotide frequencies passed the thresholds.

A total of 69 SAGs belonging to MAST-4 were co-assembled: MAST-4A (23 SAGs), MAST-4B (9 SAGs), MAST-4C (20 SAGs), and MAST-4E (17 SAGs). Prior to co-assembly, reads were digitally-normalized using BBNorm (157), considering a minimum coverage depth of 5x and a maximum target coverage depth of 100x. Normalized reads were co-assembled with SPAdes 3.10 using the single-cell mode ("--sc") running only the assembly module ("--only-assembler"). To extend contigs, they were re-scaffolded with SSPACE v3 (158). Repetitive regions were masked, along with tRNA sequences, using RepeatMasker (159) and tRNAscan-SE-1.3 (160). Quality and assembly statistics were computed with Quast (153) and are shown in **Annex A Table 2**. Parameters not mentioned were set to default. Co-assembled SAGs were carefully checked for foreign DNA. Based on the premise that sequences from the same species have virtually the same tetra-nucleotide frequencies, a second tetra-nucleotide ESOM map was built for the four MAST-4 co-assemblies with the same parameters as previously described. Contigs that did not cluster together with the majority of contigs from a given SAG co-assembly were removed. Subsequently, co-assembled contigs that were classified as prokaryotic were removed based on the 5-mer profiles using EukRep (161) with mild stringency. Lastly, eukaryotic contigs with extreme GC content values, *i.e.,* values outside the range of GC content mean ± 10 % (Standard deviation) in each SAG co-assembly, were removed as well (**Annex A Table 2**). Co-assembled genome completeness was estimated with BUSCO v3 (162). For each co-assembly, protein-coding genes were predicted de novo with AUGUSTUS 3.2.3 (163,164) using the identified BUSCO v3 proteins as the training set. Predicted genes were functionally annotated using 1) CAZy database from dbCAN v6 (165) and HMMER 3.1b2 (166) (e-value ≤ 10-5), 2) KEGG (Release 2015-10-12; (167,168)) and 3) eggNOG v4.5 (169), both using BLAST 2.2.28+ and considering hits with >25% identity, >60% query coverage, <10-5 e-value and amino acid alignment lengths >200. Gene sequences (nucleotides) were also mapped against the Marine Atlas of Tara Oceans Unigenes (MATOU) Version 1 (20171115) (43) using BLAST 2.2.28+ with the same thresholds as the ones above used for the amino acid sequences, except for the identity threshold,

which was increased to 75%, to consider nucleotide sequence variation instead of amino acid. MAST-4 genomes were clustered in terms of their GH composition with the hclust function in R based on "manhattan" distances.

1.2.3. Phylogenomics and genome differentiation

We used two approaches to analyze the phylogenetic vs. whole-genome differentiation among MAST-4 species. In the first approach, we randomly selected 30 conserved proteins (included in eukaryota_odb9, BUSCO v3) that were identified in all MAST-4 species (**Annex A Table 3**) as well as in other publicly available Stramenopile genomes: *Phytophtora sojae* (NCBI:txid67593), *Phytophtora infestans* (NCBI:txid403677), *Schizochytrium aggregatum* (JGI:Schag1), *Aurantiochytrium limacinum* (JGI:Aurli1) and *Cafeteria roenbergensis* (170). Genes were aligned individually with Mafft (171) using the '—auto' mode and concatenated with catfasta2phyml (172). Poorly aligned sequences and regions were removed using trimAl v1.4.rev22 (173) with "-automated1" mode and default parameters. The phylogenetic tree was built with RAxML version 8.0.0 (174) using the General Time Reversible model with a gamma-distributed rate variation among sites (GTR+G). Initial seed was "-p 666". In addition, we used the automatic bootstrap criterion (-autoMRE) and rapid Bootstrap mode (-f a). The second approach consisted of computing the Average Amino-acid Identity (AAI) for each pair of MAST-4 using Enveomics based on the predicted genes (amino acids). Genomes were clustered by similarity using the pvclust (175) package in R with "maximum" as the distance method.

1.2.4. Abundance and expression of selected MAST-4 ERGs in the ocean

We investigated the distribution, abundance, and expression in the global ocean of selected Ecological Relevant Genes (ERGs), in this case, lysosomal enzymes (glycoside hydrolases). For that, we mapped metagenomic and metatranscriptomic reads from Tara Oceans (a total of 52 surface water stations encompassing the 0.8 – 5 μm size fraction (total 104 samples), the organismal size range where MAST-4 is found) against predicted genes from each MAST-4 species (**Annex A Table 4**). Metatranscriptomic reads derived from sequencing polyA-enriched RNA (42,43). The mapping was done with BWA (176) and only hits with identity > 95% and an alignment length > 80 bp were considered. Reads that mapped to more than one target were discarded. Gene

abundance and expression estimates were normalized by dividing the Reads Per Kilobase (RPK) of each gene [number of mapped reads (counts) / gene length (kbp)] by the Scaling Factor (SF) [Sum of all considered RPKs in a sample / 10^6]. Hereafter, the abundance of genes and transcripts is expressed as Counts Per Million (CPM) or Transcripts Per Million (TPM) respectively. The comparison between the mean TPM values of the 20 selected MAST-4 GHs vs. the 152 single-copy housekeeping genes (from BUSCO v3's eukaryota_odb9 database) for each TARA Oceans station was performed using a two sample Wilcoxon test from the matrixTests R package (177) (**Annex A Table 8**).

1.2.5. Calculation of dN/dS ratios in homologous genes

Homologous MAST-4 genes were identified using reciprocal protein BLAST (v. 2.2.28+) with the following thresholds: >25% identity, >60% of query coverage, <10-5 e-value, and an alignment length >200 amino acids. Gene sequences (amino acid) were aligned using Mafft 7.402 with default parameters and then converted into a codon-based nucleotide alignment with Pal2nal (178). Alignments with one or more unknown nucleotides (Ns) were discarded. For each homolog, a nucleotide-based phylogenetic tree was built using RAxML 8.2.12 (174), with the model GTR+CAT, including bootstrap analyses, and a starting seed "-p 12345" as well as the optimization "-d" parameter. Positive selection was tested on each homolog with HyPhy 2.3.14 (179) using aBSREL (branch) (180) and MEME (site) (181) models considering the codon-based nucleotide alignment and the previous phylogenetic tree. Parameters included options for universal code and testing in all branches. A p-value of 0.1 (default) was used for the analysis with the MEME model.

1.2.6. Data availability

DNA sequences and metadata from the Malaspina expedition are publicly available at the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena; accession numbers PRJEB23913 [18S rRNA genes] & PRJEB25224 [16S rRNA genes]). DNA sequences from *Tara Oceans* are also stored at ENA with the accession numbers PRJEB6603 for the SAGs, PRJEB6609 for the metatranscriptomes, and PRJEB4352 for the metagenomes (**See Annex A Table 1** and **Table 4**). Genome co-assemblies, CDS

predictions, and amino acid predictions have been deposited in FigShare (doi: 10.6084/m9.figshare.13072322).

1.3. RESULTS

1.3.1. MAST-4 global distributions and associations

MAST-4A/B/C/E Operational Taxonomic Units (OTUs; "species" proxies) tended to display specific spatial distributions in the global ocean, in some cases markedly contrasting (**Figure 1.1**). Specifically, species A and C were abundant and widespread across the global ocean and even though both may appear in the same sample, they tended to exclude each other, as indicated by their association sign (**Figure 1.1**). For example, in the Pacific Ocean when moving from equatorial waters to the north, there was a partial replacement between MAST-4C and A (see arrows in **Figure 1.1**). Species B displayed a more restricted distribution and a lower abundance when compared to species C and A, being more prevalent in the tropical and subtropical Atlantic Ocean and in the tropical Pacific Ocean (**Figure 1.1**). Our analyses indicated that species B co-occurred with species C, with both species co-excluding from species A (**Figure 1.1**). Species E had a lower abundance than the other species in the tropical and subtropical global ocean, with a distribution being limited to a few locations, mostly coastal areas (**Figure 1.1**). Species E had a weak negative association with MAST-4B (**Figure 1.1**).

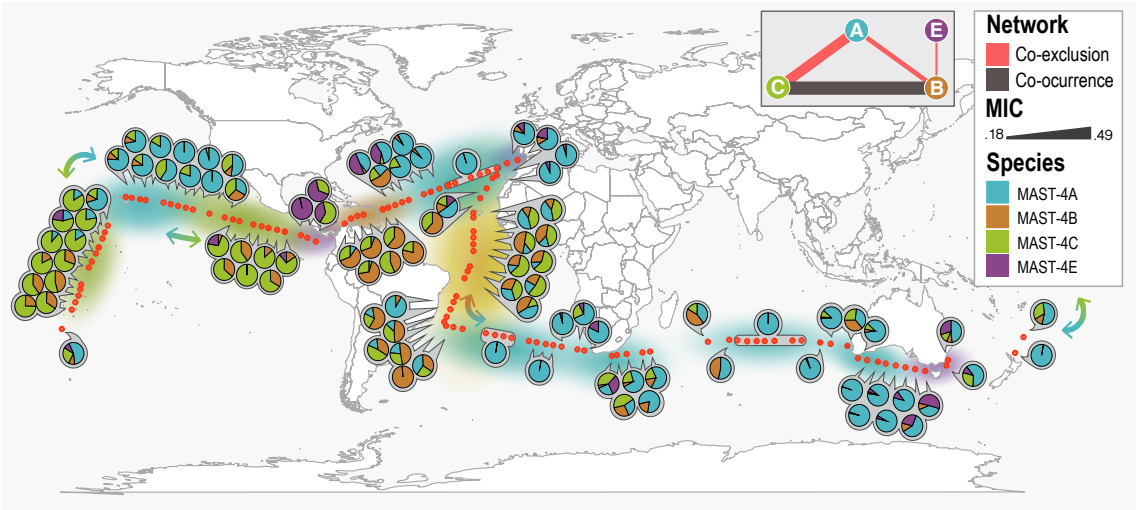

Figure 1.1. Distribution of MAST-4A/B/C/E species in the surface global ocean as inferred by OTUs based on the 18S rRNA gene (V4 region). Red dots show *Malaspina* stations while pie charts indicate the relative abundance of MAST-4 species at each station. The top-right inset network shows the association patterns between each MAST-4 species as measured using MIC analyses. The width of the edges in the network shows association strength as indicated in the legend (MIC). Background color shows the most abundant MAST-4 species in the region. Arrows point to areas with an important switch

of the abundant species: note that the most abundant species, A and C, alternate predominance in large oceanic regions.

We have also investigated the association patterns between MAST-4A/B/C/E OTUs with other picoeukaryotes and prokaryotes. We found a total of 258 associations with other picoeukaryotic and 18 with prokaryotic OTUs that cannot be explained by the measured environmental factors (**Figure 1.2A**). MAST-4C and MAST-4B displayed the largest number of associated OTUs, 191 and 174 respectively, while MAST-4A, despite being abundant and cosmopolitan, had only 23 associations. MAST-4E had only 3 associations to other taxa different from MAST-4 (**Figure 1.2A**). Most associated taxa were related to a unique (59.3%), or two (38.9%) MAST-4 species (mostly species B and C) [**Figure 1.2A**]. The co-occurring species B and C displayed the largest number of shared associated taxa (total 98 taxa), which in most cases (97%) were positively associated (**Figure 1.2A**). A lower number of associations (total 13) was shared by the mutually excluding species A and C and, as expected, had opposite signs (50% positive and 50% negative; OTUs positively associated with MAST-4A were negatively associated to MAST-4C and vice versa) [**Figure 1.2A**]. A similar trend was observed between OTUs associated with species A and B (**Figure 1.2A**).

The most represented eukaryotic classes in the network included parasites (Syndiniales; 40.7% of the OTUs) and other marine Stramenopiles (16.8%), including MAST-1/3/7/11/25 and other MAST-4 OTUs related to species B/C/E, which had different 18S-V4 sequences when compared to those from the SAGs. The most represented prokaryotic classes in the network included the heterotrophic species SAR86 (1.8%) and the small-sized marine Actinobacteria (Actinomarinales; 1.4%) (**Figure 1.2A**). Other ecologically relevant classes that were present but displayed fewer OTUs were the eukaryote Picozoa (2.14%), which have similar physiological characteristics to MAST-4 (182,183), or the prokaryotic SAR11 (0.71%), one of the most abundant bacteria in the ocean (129).

We analyzed the niche preference of individual MAST-4 OTUs as well as that of associated OTUs from other taxa in terms of temperature, salinity, NO2, NO3, PO4, SiO4, and fluorescence (**Annex A Table 5**). Adaptation to different temperature niches appeared as the main plausible driver explaining the co-exclusion between species A and species B & C (**Figure 1.2B**). The co-excluding species had different temperature preferences, with species B and C featuring a weighted mean temperature of 27.6 ºC,

while species A had a weighted mean temperature of 22.1 ºC. Both values were significantly different from chance. In contrast, species E did not show any preference

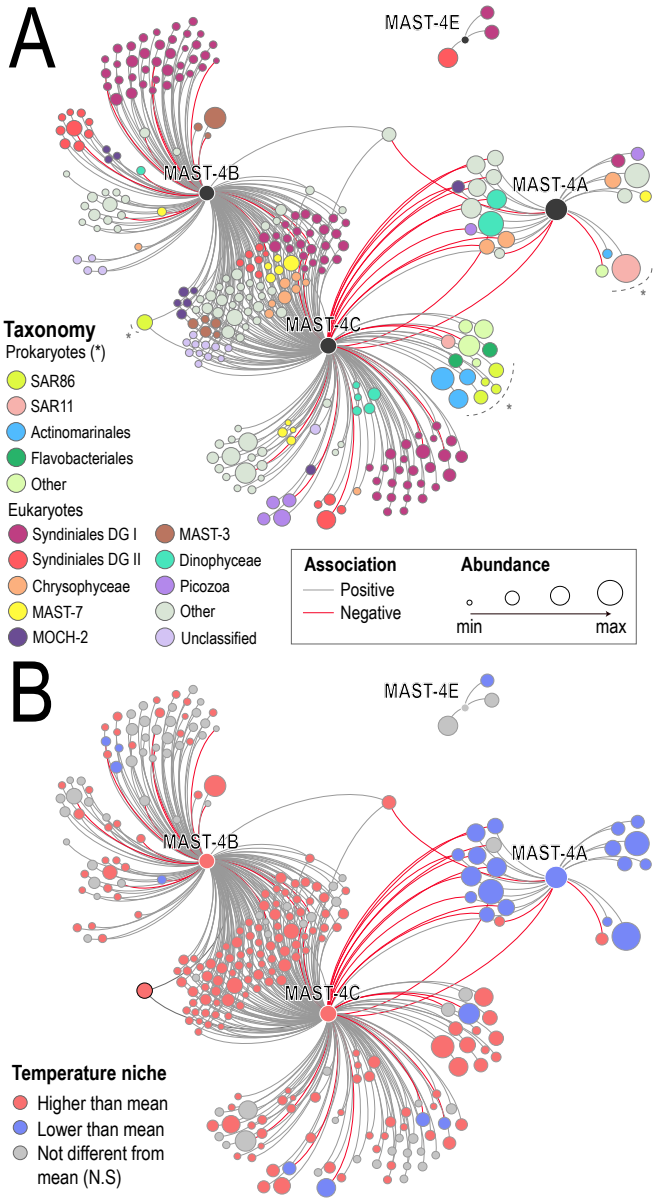

Figure 1.2. Association network including MAST-4 species, associated prokaryotes, and other pico-eukaryotes from the Malaspina expedition. Only OTUs with abundances >100 reads and occurrences >15% of the stations were considered in MIC analyses. A filtering strategy was applied to remove indirect (i.e., environmentally-driven) and weak associations (see Methods). Node size is proportional to the centered log-ratio (clr) transformed abundance sum (see Methods). **Panel A)** nodes are colored based on taxonomy. Legend: DG – Dino-Group. **Panel B)** node color indicates whether specific OTUs displayed weighted mean temperatures significantly lower or higher than the unweighted mean temperature (24.5 ºC), pointing to species with temperature distributions that differ from chance. Note that MAST-4A and both MAST-B/C tend to show co-occurrences with other OTUs that display coherent temperature preferences. N.S – Not Significant.

associated with temperature in our sample-set covering the tropical and subtropical ocean. A fraction of the taxa positively linked to MAST-4 species showed temperature

niche preferences that were coherent with those of species A, B, and C (**Figure 1.2B**; **Annex A Table 5**). For example, taxa positively associated with species A displayed an average weighted mean temperature of 22 ºC, while taxa positively associated with MAST-4B/C displayed an average weighted mean temperature of ~26 ºC. Both values differed when compared against the average unweighted mean temperature of the entire dataset: ~24 ºC. Note that detected associations reflecting only environmental preference were removed from the network, therefore remaining positive associations between microbes that prefer similar environmental conditions (*e.g.,* temperature) indicate cases where the links between microbes could not be explained by their comparable environmental preferences. Overall, water temperature explained up to 35% of the variance in the distribution of MAST-4 species (ADONIS, $p<0.05$).

1.3.2. Comparative genomics of MAST-4 species

A total of 69 single-cell genomes from MAST-4A (n = 23), MAST-4B (n = 9), MAST-4C (n = 20) and MAST-4E (n = 17) were analyzed. All MAST-4E cells were isolated from the same Tara Oceans station (station 23) at the same depth (Deep Chlorophyll Maximum - DCM) (**Annex A Table 1**). The other MAST-4 single-cells were isolated from different Tara Oceans stations located in either the Indian Ocean or in the Adriatic Sea. These cells originated also from different depths, including Surface or the DCM. Based on 18S rRNA-gene similarity, genome tetranucleotide composition, and average nucleotide identity, cells of MAST-4A/B/C/E were independently co-assembled (184). The two largest co-assemblies were MAST-4A (47.4 Mb) and MAST-4C (47.8 Mb), which contrasted in terms of size to MAST-4B (29 Mb) and MAST-4E (30.7 Mb). Accordingly, species A and C featured more predicted genes (15,508 and 16,260 respectively) than species B and E (10,019 and 9,042 respectively). MAST-4 multigene phylogenies based on 30 conserved single-copy predicted proteins (**Annex A Table 3**) as well as genome similarity based on Average Amino acid Identity (AAI) agreed with known phylogenetic relationships based on ribosomal RNA-gene sequences (185) (**Figure 1.3**). These results support our co-assembly and gene prediction strategy, suggesting also a substantial amount of evolutionary divergence between MAST-4 species A/B/C/E.

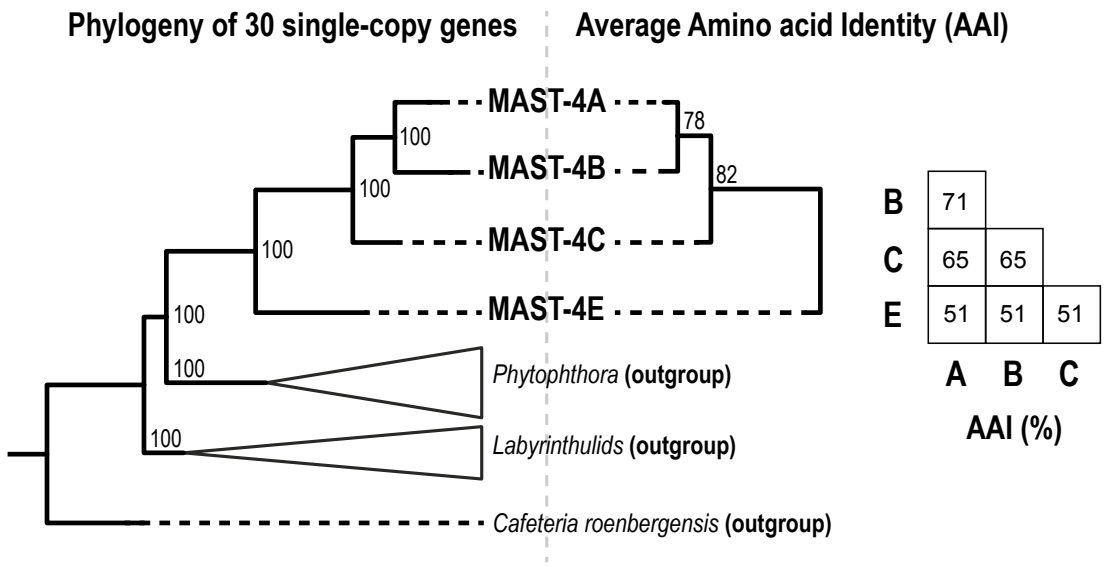

Figure 1.3. Evolutionary divergence between the studied MAST-4. Left-hand side: MAST-4 species phylogeny based on 30 single-copy protein genes from the BUSCO v3 eukaryota_odb9 database that were identified in the co-assemblies (see Methods; **Annex A Table 3**). Right-hand side: Clustering of MAST-4 co-assembled genomes and bootstrap support based on the Average Amino acid Identity (AAI) between predicted homologous genes. AAI values (%) between MAST-4 species are shown in the matrix.

All predicted MAST-4 genes were mapped against the Marine Atlas of Tara Oceans Unigenes (MATOU, a metatranscriptomics-based gene catalog of expressed eukaryotic genes clustered at 95% identity) (186) in order to: 1) assess whether predicted MAST-4A/B/C/E genes have been previously recovered in global-ocean metaomics surveys and 2) determine the presence of other environmental orthologs that could point to additional MAST-4 species that are prevalent in the ocean but were not considered in our work. We analyzed MATOU genes that had ≥ 75% nucleotide similarity to MAST-4A/B/C/E genes. This threshold was used to recover environmental orthologs belonging to both MAST-4A/B/C/E as well as other MAST-4 species. The number of orthologs detected in MATOU for MAST-4A/B/C/E was variable, with species A showing orthologs for ~25% of its genes, species B ~20%, species C ~33%, and species E ~13% (**Annex A Table 6**). Not a single MATOU unigene had orthologs present in all the analyzed MAST-4 species, while 81.9% of the MAST-4 orthologs present in MATOU were associated with a single MAST-4A/B/C/E species (**Annex A Figure 1**). This suggests that other MAST-4 species different from MAST-4A/B/C/E are not abundant in the tropical, subtropical and temperate open ocean, and that the recovered orthologs mainly represent population/ecotype variation. Yet, the MAST-4 group seems to have a limited representation in MATOU (only orthologs for ≤ 1/3 of

MAST-4A/B/C/E genes were found) and more environmental genes should be sampled over different spatiotemporal scales than that of *Tara Oceans* in order to support our findings. In any case, MATOU results were coherent with our previous AAI results indicating a substantial genome differentiation among MAST-4A/B/C/E.

Predicted amino-acid sequences were functionally annotated using the databases eggNOG and KEGG. eggNOG allowed the annotation of ~75% of the genes from the four species, while ~31% were annotated with KEGG. Considering that eggNOG includes environmental sequences, some with unknown functions, while KEGG is based on model or cultured organisms and annotated genes, these differences are not surprising. According to the broad eggNOG functional categories, MAST-4 species shared similar functional profiles (**Figure 1.4A**). Yet, about half of the eggNOG hits had no function associated, as the reference sequences were environmental. Nevertheless, the existence of these hits further supports our co-assembly and gene prediction approach. The most represented categories with known functions were 'Posttranslational modification, protein turnover, chaperones' and 'Signal transduction mechanisms', which group important genes for the proper functioning of the cell, along with 'Intracellular trafficking, secretion, and vesicular transport' and 'Carbohydrate transport and metabolism', which include pathways related to food ingestion and degradation (lysosomal reactions). Similarly, KEGG functional categories with the largest number of MAST-4 genes were 'Global Metabolism', 'Signal Transduction', and 'Transport and Catabolism' (**Annex A Figure 2**). The first two comprise broad housekeeping functions and pathways, while the third covers vesicular processes such as endo- and phagocytosis. As expected, the potential for grazing is represented in all four MAST-4 genomes

The amino-acid gene sequences were also annotated against the CAZy database, which targets functions affecting glycosidic bonds. A total of ~3% of the total MAST-4 genes had a match against the CAZy database (**Annex A Table 6**), and the group with the largest number of genes in MAST-4 species was the Glycoside Hydrolases (GHs) (**Figure 1.4B**). We have analyzed the GH composition of MAST-4, given that different GH repertoires in species could be linked to different capacities to degrade prey bacteria or microalgae (187,188). Most GH families were found in all MAST-4 species, but some were specific or missing in particular species (*e.g.,* GH23 specific to MAST-4B or GH22 missing in MAST-4E) [**Annex A Table 7**]. Clustering of MAST-4 species based

on GH composition generated two groups, species A – C and B – E (**Figure 1.4C**). Thus, MAST-4 genomes with contrasting geographic distributions (**Figure 1.1**) and contrasting potential ecological interactions (**Figure 1.2A**) were clustered together based on similar GH composition.

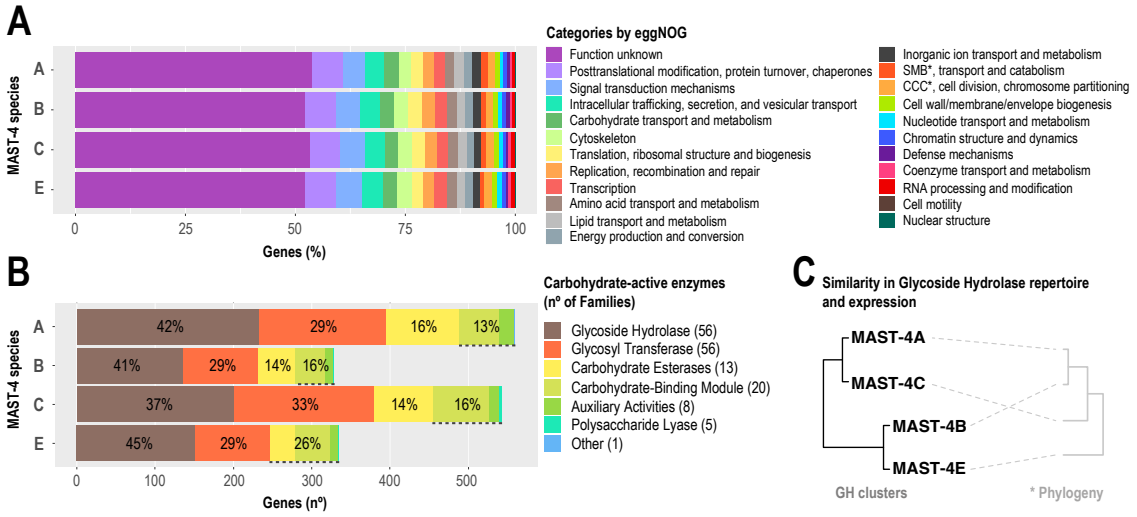

Figure 1.4. Functional profile of MAST-4 genes according to eggNOG and CAZy. Total MAST-4 genes analyzed were 15,508, 10,019, 16,260 and 9,042 for species A, B, C and E respectively. **Panel A)** eggNOG annotations indicated as percentage of genes falling into functional categories. SMB – Secondary Metabolites Biosynthesis, CCC – Cell Cycle Control. **Panel B)** Number of MAST-4 genes within CAZy categories and the corresponding percentage. The number of gene families considered within each CAZy category is indicated between parenthesis in the panel legend. **Panel C)** Clustering of MAST-4 species using Manhattan distances based on either their Glycoside Hydrolase (GH) composition or the GH expression (in TPM) results in the same clustering pattern. Note that MAST-4C and A are more similar in their GH content than E and B, which are more similar between themselves. ***** A schematic representation of the phylogeny of the studied MAST-4 is shown for comparison purposes (see **Figure 1.3** for more details).

1.3.3. Global expression of MAST-4 Glycoside Hydrolases

In MAST-4, Glycoside Hydrolases (GHs) are most likely involved in the machinery to digest food after phagocytosis. We used metatranscriptomic and metagenomic data from the Tara Oceans expedition to assess the expression and abundance of MAST-4's GH genes in the surface global ocean (**Figure 1.5A**). We found that there was no obvious relationship between GH gene-abundance and expression over the surface global ocean, indicating that differences in gene expression most likely represent up- or down-regulation of GH genes (**Figure 1.5B**; see also **Figure 1.5C** and **Annex A Figure 3B**). MAST-4's GH gene expression was highly heterogeneous in the surface global ocean (**Figure 1.5C**). The GH families with the highest expression were the lysozyme families GH22 and GH24, in charge of degrading the peptidoglycan in the bacterial cell wall (140,189), as well as the chitinase family GH19, involved in the degradation of chitin

(present in particulate detritus, crustaceans and several other organisms in the ocean) [**Figure 1.5C**]. These GH genes tended also to display a higher expression mean than single-copy housekeeping genes within the same Tara Oceans stations (**Annex A Table 8**). Interestingly, the South Pacific displayed low or absent GH expression in all MAST-4 species, despite GH gene abundances that were similar to those found in other regions displaying higher expression (**Annex A Figure 3B**). We found also clear differences in expression between species: for example, while species' A GHs were widely expressed in several regions, those GHs from species E were expressed only in specific samples, in particular in the North and South Atlantic. GH genes from species B and C were either not detected or had low expression in the South Atlantic samples, in contrast to specific GH genes from species A and E in the same region (**Figure 1.5C**). In turn, specific GHs from species B and C had higher expression than A and E in the Indian Ocean.

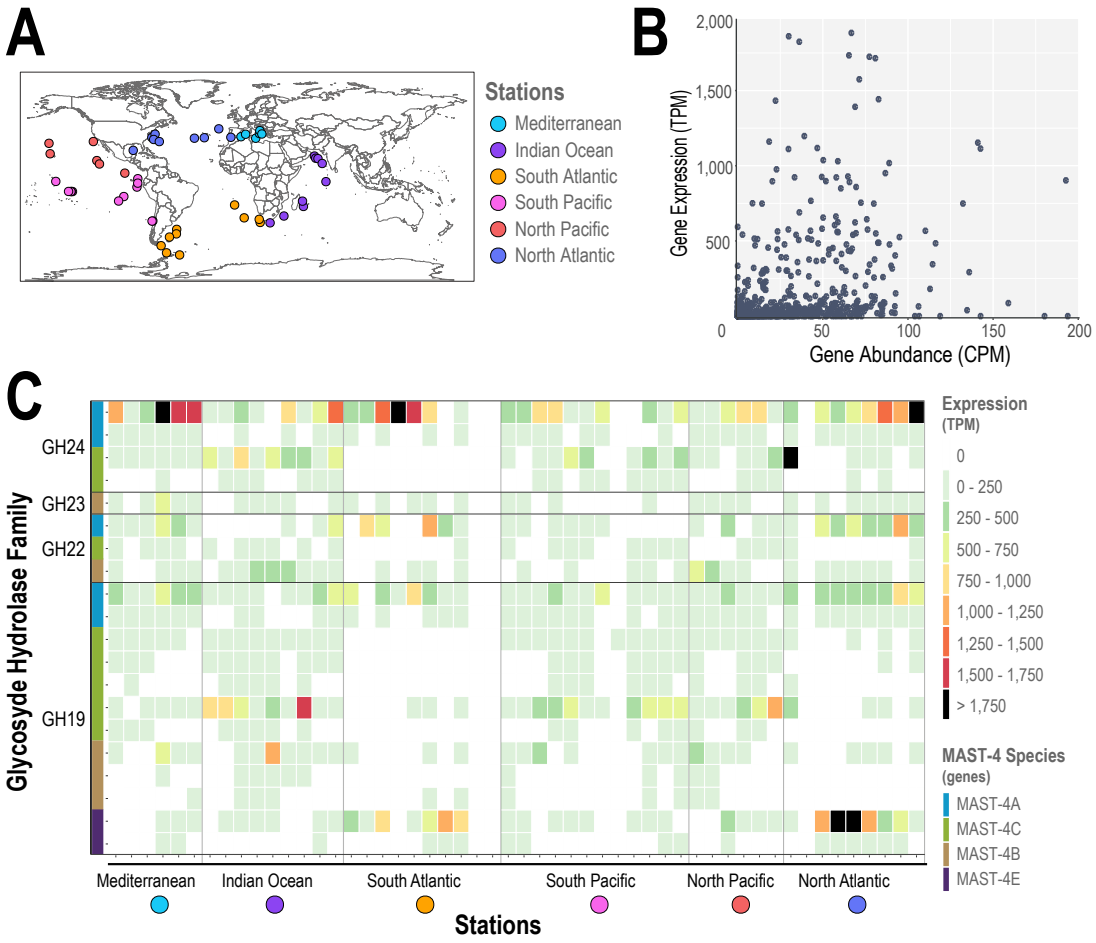

Figure 1.5. Expression and abundance of GHs in MAST-4A/B/C/E in the upper global ocean. **Panel A)** Geographic location of the metagenomic and metatranscriptomic samples from Tara Oceans. **Panel B)** Gene abundance vs. expression using normalized data for each gene and station. Note that the axes have different but proportional ranges of values. **Panel C)** Heatmap of the Glycoside Hydrolase families in

MAST-4 that had the highest expression. Samples are in the X-axis, grouped by ocean region and ordered following the expedition's trajectory. Genes in the Y-axis are organized by family and each species is indicated with a color. GH22, GH23 and GH24 are families of lysozymes and GH19 is a family of chitinases that can also act as lysozyme in some organisms.

Differences in abundance and expression were also found in GH genes belonging to the same family and within the same MAST-4 species. For example, species A had two genes belonging to the GH24 family; one gene (631 bp) was more expressed than the other (1,465 bp), despite gene abundances being similar across all samples (**Figure 1.5C**; **Annex A Figure 3B**). These two genes shared 29.5% similarity at the amino acid level based on 73% coverage (153 amino acids) of the shorter gene. A similar pattern was observed in the two GH24 genes in MAST-4C: the shorter was more expressed than the longer (622 vs. 1,198 bp). In fact, the short and long GH24 genes from species A and C are homologs respectively: the short homologs have 79.4% identity (94% coverage) while the long homologs have 56.4% identity (87% coverage). In general, MAST-4 species with more than one gene belonging to the same GH family tended to express one particular variant over the others. One plausible explanation is that the under-expressed GHs are gene duplications. GH genes often undergo duplication, and thus several copies can be present in the form of paralogs (190–192). After gene duplication, a redundant copy is generated and freed from selective pressure, allowing it to accumulate mutations (193) and potentially lead to new functions (194,195).

1.3.4. Detecting positive selection acting on MAST-4 genes

We analyzed whether there is evidence of positive selection leading to niche adaptation in the different MAST-4 species. For that, we analyzed non-synonymous vs. synonymous substitutions (dN/dS) in selected homologous genes in MAST-4A/B/C/E. Normally, the ratio dN/dS is used to test hypotheses related to the action of selection on protein-coding genes, where dN/dS >1 indicates that substitutions generating changes in amino-acids are greater than substitutions that do not, suggesting the action of diversifying (i.e. positive) selection (196). A total of 692 alignments (homologous groups) were used for testing positive selection on both branch (whole sequence phylogeny) and codon analyses (gene site-specific) (180,181) (**Annex A Table 9**; see Methods). Overall, 60 gene alignments (8.7%) indicated positive selection in branch analyses, of which 57 alignments displayed selection in one branch and 3 in two branches (60 alignments, 63 total branches selected). MAST-4A and B appeared to be

the most selected branches, 22 (34.9%) and 25 (39.7%) times respectively, while MAST-4C and E had a low number of selected branches, 8 (12.7%) and 4 (6.3%) times respectively. In codon analyses, 478 gene alignments (69.1%) displayed positive selection in one or more positions, ranging from 1 to 15 positively selected codons per alignment. In Glycoside Hydrolases (GH), a key part of the predatory machinery of the MAST-4, 1 alignment (0.14%) showed positive selection in branch analyses for family GH74 while 14 alignments (2%) displayed positive selection in codon analyses that included GH3, GH13, GH16, GH19, GH28, GH30, G74, GH78, GH79 and GH99 (**Annex A Table 9**). Of all of them, only GH19 belongs to one of the most expressed families according to the metatranscriptomic analyses. Overall, these analyses suggest that adaptive evolution promoted the diversification of MAST-4 into species A, B, C & E, or at least that it promoted the diversification of specific genes.

1.4. DISCUSSION

Currents, waves, and wind promote the dispersal of plankton in the surface ocean. Given their typically large populations and small organismal sizes, microbial plankton species are expected to be widely distributed in the upper ocean. This is particularly relevant for the MAST-4 group, which features a moderate abundance (about 50 cells ml-1 in surface waters and ~10% of the heterotrophic flagellates (197)) and minute size. Such characteristics in combination would guarantee dispersal and widespread distributions (63), decreasing the potential effects of dispersal limitation (198). These characteristics would also promote a coupling between environmental heterogeneity (selection) and species distributions (199). Thus, we expected that MAST-4 distributions would reflect, to a certain extent, the abiotic and biotic conditions in the ocean. This is coherent with previous findings indicating that a) temperature is an important environmental variable driving MAST-4 distributions and b) that dispersal limitation does not seem to affect the distributions of MAST-4 species (135). We expanded previous knowledge by determining the temperature distribution of species A, B, and C. Specifically, we show that species B and C occur in warmer temperatures (weighted mean = 27.6 ºC), while species A is present in lower temperatures (weighted mean = 22.1 ºC). In contrast, we did not find evidence that the distribution of species E was affected by temperature in the tropical and subtropical ocean. This is coherent with reports indicating that MAST-4E inhabits cold waters (135).

Even though temperature is a key variable structuring the global ocean microbiota, including MAST-4 (8,44,200,201), biotic variables could also affect the distributions of MAST-4 species. We found that the number of associations between MAST-4 OTUs and bacterial OTUs was low. Actually, most associations were not considered as they were either weak (low correlation) or they just represented similar or different environmental preference (mainly temperature) between MAST-4 and bacterial OTUs. Altogether, this suggests that MAST-4 abundance and occurrence is weakly coupled to bacterial distributions and abundance in the upper ocean, which agrees with previous studies where changes in the overall heterotrophic flagellate abundances were related to water temperature (197). We detected a substantial number of taxa that were positively associated with either MAST-4B/C or MAST-4A but not to both. Even though associated taxa tended to reflect the temperature preference of the species to which they were associated (B/C or A), their association to different MAST-4 cannot be simply explained by similar niche temperature, since we also detected associations to OTUs without a significant temperature preference. The vast majority of associations were between species A or B/C with other picoeukaryotes, such as Syndiniales' Dino-group-I and II, which are known parasites (202), or MAST-3 and MAST-7, which are flagellates as well (53). These associations could either manifest a similar preference for an environmental variable different from temperature that covaries with MAST-4 distributions, or reflect real ecological interactions, including parasitism. For instance, there is evidence of MAST-4A having a predator-prey relationship with *Synechococcus* (131) and possibly with SAR11 (111), which was not only reflected in our networks from the Malaspina expedition, but also in previous studies from the *Tara Oceans* expedition (110). Results from *Tara Oceans* reported other taxa associated with MAST-4A that were corroborated by our results (MOCH-2, Chrysophyceae, MALVs, MAST-7). However, whether or not these associations reflect true ecological interactions needs to be proved with further experiments. Altogether, we did not find evidence that biotic interactions between MAST-4 and other microbes represent an important driver of MAST-4 biogeography.

Our results suggest that adaptation to different temperature niches and interspecific interactions between MAST-4 species (competition) are likely the main drivers determining MAST-4 biogeography. If so, differential adaptation should likely be reflected in the genomes of the MAST-4 species. Our analyses indicated that MAST-

4 species differ in genome size: two bigger genomes (MAST-4A and C) with a partial genome size of ~47 Mb and ~80% completeness and two smaller genomes (MAST-4B and E) of ~30 Mb and ~70% completeness, which correspond to ~59 and ~42 Mb full estimated genomes respectively. The observed differences in genome size need to be considered with care, as they may be reflecting incomplete genome assemblies. Nevertheless, our estimates of genome size were similar to those of *Cafeteria roenbergensis* (~40 Mb) (170), a heterotrophic flagellate in the same cell-size range of MAST-4, and other Stramenopile genomes, for example the diatom *Thalassiosira pseudonana* (~34.5 Mb) (203) or various *Phytophthora species* (*P. plurivora*, *P. multivora*, *P. kernoviae* and P. agathidicida with 41, 40, 43 and 37 Mb respectively) (204). This suggests that our partial genomes are likely large enough to be representative of the studied MAST-4 species. We found that differences in MAST-4 genome size were mirrored by the number of predicted genes in each species, which ranged between 9,042 and 16,260, even though larger genomes in eukaryotes do not always imply a greater number of genes (205). These differences in gene content between species may to some extent be linked to niche adaptation. Overall, none of the studied MAST-4 displayed any loss or gain of broad functional categories when compared to each other. In fact, they were similar in terms of the proportion of genes that belong to each functional trait, suggesting that MAST-4 metabolisms are broadly comparable, which agrees with other reported results in MAST-4 species A/C and E (102). Among the most represented functional categories in the MAST-4 genomes were those involved in phagocytosis and subsequent digestion. For instance, eggNOG's 'vesicular and carbon transport', along with KEGG's 'transport and catabolism', includes pathways for 'Endocytosis', 'Phagosome', 'Lysosome', 'Peroxisome' and 'Autophagy (animal and yeast)', all related to vesicular forms of transport and prey digestion. Thus, MAST-4's lifestyle as marine grazers (129,206) is in agreement with their broad genomic functions associated with phagocytosis. Yet, homologs among species were very different at the DNA or amino acid level. In particular, when comparing MAST-4A/B/C/E gene predictions against the Tara Oceans catalog of marine eukaryotic genes (MATOU) (33,43), the vast majority of homologs were unique to one MAST-4 species. In fact, we did not find a single gene in MATOU with homologs in all MAST-4A/B/C/E species, which manifests the interspecific differences of MAST-4 in terms of genomic composition. The substantial differentiation between homologs was reflected by the AAI and phylogenomic results as well (**Annex A Figure**

1), which altogether indicate that MAST-4 experienced substantial evolutionary diversification.

MAST-4 is not exclusively bacterivorous and can feed on other small organisms, for example *Micromonas pusilla* and *Ostreococcus sp.* (129), and perhaps complement its diet with non-infective viruses (113). A comparable diet has been observed in other heterotrophic flagellates (207). Such a variety of food items, which vary in quality and quantity, most likely require different metabolic machineries to digest them (102,208), in particular different carbohydrate-active enzymes. For example, studies in Fungi have shown that the number and composition of CAZymes may determine the degradation capacity of different plant biomass sources (209). Here, we analyzed the Glycoside Hydrolases (GHs), one of the most efficient known catalysts of organic substances in living organisms (210), and likely important for MAST-4's heterotrophic lifestyle. GHs genes accounted on average for 3% of the predicted genes in each MAST-4. Most of the GH gene families were found in the four species, but some were either exclusive of a single species or missing in others, which may be due to genome incompleteness. Similar patterns have been reported before, not only in a reduced number of MAST-4 species (102) but also in the fungal genus *Saccharomyces* (211), where the set of GH genes differs even in strains of the same species. Site (codon) analyses suggested positive selection in a few GH families in MAST-4, for example, within the GH19 gene family. Similarly, other GH families that are not lysozyme-like, such as GH3, GH30, or GH74 appeared to have experienced positive selection as well, even though they were not as much expressed in the global ocean as the lysozyme. Altogether, this suggests the action of adaptive evolution in the machinery that MAST-4 uses to digest food, and may reflect adaptations to the degradation of different compounds or prey.

The four MAST-4 species formed two groups based on GH composition (number of genes per family). One group consisted of species A and C and the other of species B and E. Interestingly, species A and C, with similar GH repertoires, showed spatial co-exclusion in the upper global ocean, while species C and B, with different GH repertoires, were co-occurring (**Figure 1.1**). These geographic distributions suggest that niche adaptation associated with different temperatures allowed MAST-4A and C to keep similar GH repertoires, while species adapted to similar temperatures that co-occur (C and B) were exposed to divergent selection diversifying their diets as a response to competition, which is reflected in their different GHs (208). We found that species A

and B/C have different niche temperatures (A= 22.1 ºC and B/C= 27.6ºC). Since temperature niche can be a phylogenetically conserved trait in specific microbes (212,213) it would have been expected that the closely related MAST-4A and MAST-4B share a similar temperature preference. However, species A had a temperature preference 5 ºC lower than that of B, suggesting that selection has promoted the adaptation of species A to lower temperatures perhaps to not compete with species C, or that species C is a superior competitor and excludes species A from warmer waters. Further, since MAST-4A, B, and C form a monophyletic group they are expected to share a comparable GH repertoire. But instead, our analysis showed that the GH repertoire of B was closest to E, suggesting that evolution promoted the divergence of MAST-4B's GH content.

The temperature distributions of the studied MAST-4 species, together with their different GH repertoires lead to two plausible evolutionary scenarios. MAST-4E, the deepest branching lineage, did not show a particular preference for either warm or cold waters in our data (**Annex A Table 5**), but other reports indicate it occurs in cold waters (135). Thus, during the MAST-4 diversification, species E would have either adapted to or remained in cold waters. Then, two evolutionary hypotheses emerge depending on whether the Last Common Ancestor (LCA) of MAST-4A/B/C originated in warm or cold waters: 1) The LCA of MAST-4A/B/C was adapted to warm waters and species C remained in warm waters. Then, the two most evolutionary derived species, A and B, diverged their niches as a result of competition with C; species A adapted to colder subtropical and temperate waters, while species B stayed in the tropics and avoided competition with C by changing its niche via diet modification, which is reflected in its GH composition, 2) The LCA of MAST-4A/B/C inhabited cold (subtropical) waters and then C and B adapted independently to warmer tropical habitats with B modifying its niche to avoid competition with C by changing its GH repertoire and consequently its diet. Even though both evolutionary scenarios are possible, our dN/dS results using homologous proteins of the four MAST-4 species are more coherent with the first evolutionary scenario by indicating that MAST-4A and MAST-4B appear to have diverged the most, as they displayed the effects of significant positive selection in 75% of the total alignments with branch selection.

We also analyzed MAST-4's GH distribution and expression in the surface global ocean, as this may shed light on whether species with similar GH composition

express similar or different genes when they co-occur, possibly indicating prey preference depending on the presence-absence of competitors. We found that the different species displayed a large heterogeneity in their expression patterns. The tropical species that co-occurred the most, C and B, showed dissimilar expression patterns, with some genes being highly expressed only in one species, which is coherent with their difference in GH composition as well as with a scenario proposing different food preferences. Furthermore, species C and B showed differences in expression over specific ocean regions, suggesting that despite their co-occurrence, their GH activity is modulated differently. In turn, the co-excluding species A and C, which display the most similar GH composition, appeared to express different GHs over the upper global ocean, suggesting that they regulate GH expression perhaps as an adaptation to different preys or that GH expression is affected by the different temperatures in which these species occur. Overall, our evidence suggests that species A, B, and C regulate GH genes differently, perhaps as an adaptation to different diets or prey, even though some differences in GH expression only reflect the presence or absence of MAST-4 species in specific ocean regions.

Altogether, our results suggest that the evolutionary diversification of MAST-4 was promoted by divergent adaptive evolution towards different temperature and/or diet niches possibly as a response to competition and that biotic interactions with other species did not have a major influence in MAST-4 diversification. The previous possibly led to the emergence of the species associated with tropical (MAST-4B and C), subtropical-temperate (MAST-4A), and subpolar-polar (MAST-4E) waters. Furthermore, species B may have diverged in its diet as a response to competition with C, and as a result, it has a different GH composition from its closest evolutionary relatives, A and C. If future cultures of MAST-4 species are established, the previous scenarios could be tested by determining the temperature range of species growing in isolation or with interspecific competitors. Our work represents a significant contribution to understand the evolution, diversity, biogeography, and function of the smallest predators in the ocean. This knowledge is fundamental to comprehend the base of marine food webs and the biotic and abiotic factors that may affect them, as well as the consequences in upper trophic levels.

CHAPTER 2

Global population structure of a unicellular marine predator

Francisco Latorre[a], Olivier Jaillon[b,c] , Michael E. Sieracki[d], Lidia Montiel[a], Corinne Cruaud[e], Ramon Massana[a], Ramiro Logares[a]

[a] Institute of Marine Sciences (ICM), CSIC, Barcelona, E-08003, Catalonia, Spain

[b] Metabolic Genomics, Genoscope, Institut de Biologie François Jacob, CEA, CNRS, Univ Evry, Université Paris Saclay, 91000 Evry, France

[c] Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022 / Tara Oceans GOSEE, 3 rue Michel-Ange, 75016 Paris, France

[d] National Science Foundation, 2415 Eisenhower Ave., Alexandria, VA 22314, U.S.A

[e] Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, 91000 Evry, France.

2.1. INTRODUCTION

Heterotrophic protists are common in all aquatic ecosystems (214), with marine heterotrophic flagellates (HF) alone representing around 20% of the total eukaryotic organisms in the sunlight zone of the ocean (215). Altogether, HFs have a crucial role in marine food webs by channeling nutrients and energy from primary producers to upper trophic levels. These organisms are active grazers, being important agents in the regulation of prokaryotic as well as small-eukaryotic abundances in the plankton (128,216). Traditionally, HFs have been combined into a single functional group, overlooking their organismal and evolutionary differences. Today we know that they are an assemblage including evolutionary divergent organisms affiliating with all major eukaryotic supergroups (217–220).

As the ocean is an interconnected medium, marine plankton, including HFs, can travel thousands of kilometers carried by currents, moving through cold polar regions and warm tropical waters, across regions with different abiotic or biotic features, or from the surface into the deep ocean. However, environmental conditions can be drastically different between the above regions and present low connectivity (*e.g.,* warm and polar regions or surface and deep ocean). These heterogeneous environmental conditions can exert disruptive selection and may lead to local adaptation and eventually different populations and new species (221). The large diversity of HF lineages may reflect the substantial heterogeneity that these organisms encountered in the ocean during their evolutionary diversification (44). Genomic diversity and population differentiation can be identified in the form of variants and have been used to study population genomics and the evolution of different species, from macrofauna (222–224) to microorganisms (88,96,225). Yet, the amount of population-level genomic variation in HF species and how it is structured has been poorly investigated. Considering the large population sizes and high reproduction rates of HFs, it is expected that their populations include a substantial genetic diversity, which is structured into locally or regionally adapted populations.

Among HF, Marine Stramenopiles (MASTs) with cell sizes ranging between 2-5 µm are the most frequent in the surface ocean (41,43,220,226). They can account for up to 50% of the abundance in metabarcoding analyses in specific locations (227), and the vast majority represent uncultured groups (58,220,228). So far, a total of 18 MAST

groups are recognized, branching at different basal areas of the stramenopile phylogeny (53). MASTs can inhabit oceanic waters from surface (MAST-1, -3, -4 and -7) (53,61,226,228) to deeper layers (MAST-23) (53), from tropical (MAST-1 and -4) (61,131) to polar areas (MAST-2) (132,227), and also freshwater environments (MAST-2 and -12) (53,229). MASTs play key roles in microbial food webs as active grazers on bacteria and small algae (58,129,131,133) and some have been described as symbionts of diatoms (MAST-3) (230).

Among MASTs, the MAST-4 clade shows a worldwide distribution and high relative abundance in surface marine waters in comparison to other HFs (9% of all HFs) (220). Along MAST-1 and -7, MAST-4 account for 10 – 20% of HF in marine ecosystems (52,131,197,231). As a result, MAST-4 is becoming a model organism to study HFs ecology in the ocean. MAST-4 shows a geographical distribution correlated with environmental heterogeneity (135), suggesting that its diversification in the surface ocean has been promoted by adaptation (61). The biogeography of some MAST-4 species has been partially elucidated. In a previous study (61) we found that MAST-4B and C co-occur in tropical waters, while simultaneously excluding themselves from MAST-4A, which predominantly inhabits subtropical waters. Although temperature and, to a lesser extent, salinity were defined as the main ecological drivers shaping the biogeography of MAST-4 species, gene expression and abundance data also suggested that competition for food (prey) played a key role in niche adaptation, contributing to the observed biogeography.

Despite these advances, the population genomic variation remains barely known for different species of MAST-4. A reason for this was the difficulty to retrieve genomic data from uncultured HF from the ocean. Yet, this has changed thanks to advances in metagenomics and single-cell genomics (SCG), which allowed us to collect genomic data from MAST-4 directly from the environment (58,81,102,232). Subsequent bioinformatics analyses permitted determining the genomic variation, which can occur as single nucleotide polymorphism (SNPs), insertions and deletions of genomic regions (indels) or gain or loss of complete genes. Specifically, we investigated the population genomics of the MAST-4 species A, B, C and E using genomes obtained via single-cell genomics as well as metagenomes from the *Tara Oceans* circumglobal expedition (43). Specifically, we ask: What amount of genetic diversity do the analyzed MAST-4 species show in the global ocean? Do they exhibit population structure? If so, which

environmental or geographic factors promote such structure? Can we detect genes or genomic regions featuring adaptations to different environmental conditions? Answers to these questions can bring new insights into the ecological drivers determining the diversity and functionality of marine HFs, expanding our knowledge about adaptive diversification, the establishment of new populations, and ultimately speciation in these keys, but poorly known, marine unicellular predators.

2.2. METHODS

2.2.1 Genome reconstruction using Single Amplified Genomes

Genomes from MAST-4 species A/B/C/E were reconstructed after co-assembling multiple Single Amplified Genomes (SAGs) obtained from plankton samples collected during the circumglobal *Tara Oceans* expedition (43). A total of 69 SAGs (over 424.1 Gb of sequencing data) were selected and processed to generate a single co-assembled genome for each species, from which genes were predicted and functionally annotated using different databases [KEGG (Release 2015-10-12) (167,168), eggNOG v4.5 (169) and CAZy database from dbCAN v6 (233,234)] as described in Chapter 1 (61).

2.2.2 Abundance of MAST-4 in the open ocean

We determined the abundance in the global ocean of the four studied MAST-4 genomes. To achieve that, we mapped 111 metagenomes from *Tara Oceans* (a total of 82 surface water stations encompassing the 0.8 – 5 μm size fraction; **Annex B Tables 1 and 2**) against the whole genome of each MAST-4 species. BAM files for each station and genome were generated with BWA 0.7.17-r1188 (176) and only reads with identity > 95% and an alignment coverage > 80% were kept. Raw read counts were used to estimate RPKG (mapped reads per Kb of Genome per Gb of metagenome) abundance values for each station. Genomic Horizontal Coverage, defined as the percentage of the genome covered by at least 1 filtered read was calculated using Samtools 1.8 (235).

2.2.3 Genetic divergence of MAST-4 in the open ocean

To assess the genetic divergence within MAST-4 species in the global ocean, we analyzed Single Nucleotide Polymorphisms (SNPs) and small insertions and deletions (indels) across different ocean regions. For each MAST-4 genome, a total of 82 BAM

files (one per station) were merged into a single BAM file using *Samtools 1.8 merge* function. Merged BAM files were used as input to *Freebayes* v1.3.1 (236) to perform variant calling, with ploidy set to 1 (-p 1) and minimum number of observations to support an alternate allele set to 4 (-C 4). The resulting variant call files (VCF) were used as input to a) *SnpEff 5.0e* (build 2021-03-09) (237) with default parameters to annotate and predict the effects of genetic variants on MAST-4 genomes, genes and proteins*;* and b) *POGENOM v.0.8.3* (89), with minimum coverage for a locus set to 10 (--min_count 10) and minimum number of stations that a locus needs to be present to 4 (--min_found 4) to compute the Fixation index (FST) values for all the pairwise comparisons between stations. Following Hartl and Clark (78), we established four groups of genetic differentiation based on FST pairwise values: FST < 0.05, little genetic divergence; 0.05 < FST < 0.15, moderate; 0.15 < FST < 0.25, high; FST > 0.25, very high. The amount of genetic differentiation (FST values) explained by selected environmental variables (Temperature, salinity, and density) was analyzed with PERMANOVA (*adonis* function in the *vegan* R-package) using environmental variables with Z-score normalization. Station 11 was removed from these analyses due to missing data.

2.2.4 Calculation of dN/dS ratios

Potential adaptive evolution in MAST-4 coding sequences was analyzed using the ratio of non-synonymous vs. synonymous substitutions (dN/dS). Overall, it is considered that a dN/dS > 1 implies positive or Darwinian selection, dN/dS < 1 stabilizing selection and dN/dS = 1 neutral selection. Here, we calculated the dN/dS ratios per gene and station following the indications of Nei and Gojobori (238) and Morelli *et al.,* (239).

2.2.5 Data availability

DNA sequences from *Tara Oceans* are stored at ENA with the accession numbers PRJEB6603 for the SAGs and PRJEB4352 for the metagenomes (**See Annex B Table**). Genome co-assemblies, CDS predictions, and amino acid predictions of MAST-4 are available in FigShare (doi: 10.6084/m9.figshare.13072322).

2.3. RESULTS

2.3.1 Variant detection and annotation in MAST-4

For each MAST-4 species, genomic variants were classified into Single-Nucleotide Polymorphisms (SNPs), Multiple-Nucleotide Polymorphisms (MNPs), DNA insertions and deletions (INDELs), and all the possible combinations (MIXED). A total of 864,009/131,091/668,613/137,357 genomic variants (18.2/4.5/14.0/4.5 variants per kb of genome) were predicted for MAST-4A/B/C/E, respectively. On average, 87% of the variants in each MAST-4 were SNPs, while the other 13% was distributed among MNPs, INDELs, and MIXED. Considering that one variant can have more than one effect on different genes (*e.g.,* an SNP in the downstream area of gene A can also be part of the upstream area of gene B) a total of 2,644,001/437,930/2,196,738/446,874 effects (3.06/3.34/3.29/3.25 effects per variant) were annotated for MAST-4A/B/C/E. On average, 78.4% were located in non-transcribed areas of the genome, 20.4% in coding regions and 1.2% in non-coding regions (**Annex B Table 3**). Variants located in coding regions were classified into missense and silent variants, depending on whether they change the resulting amino acid sequence or not, and nonsense variants if they truncate the resulting protein by introducing a stop codon. MAST-4A and E displayed an average of ~ 45% missense and ~ 55% silent variants, while MAST-4B and C showed an average of ~ 32% and ~ 68% respectively. Less than 1% of the effects were assigned as nonsense.

The effects of the variants were assigned to impact categories: HIGH, when the variant is assumed to have a disruptive impact on the protein (truncation or loss of function); MODERATE, when a variant can potentially change the protein effectiveness; LOW, when the variant is unlikely to change protein functionality; and MODIFIER, for non-coding variants or variants for which it is difficult to predict impact. On the one hand, the impact of variants on MAST-4A/B and C were proportionally similar, having on average 0%, 13.4%, 8.6%, and 78.0% for HIGH, MODERATE, LOW and MODIFIER categories respectively. On the other hand, MAST-4E showed proportionally more effects identified as MODIFIER and less as LOW in comparison to the other MAST-4 species (0.1%, 8.6%, 7.2% and 84.1% for HIGH, MODERATE, LOW and MODIFIER respectively) (**Annex B Table 3**).

2.3.2 Genetic divergence of MAST-4 populations

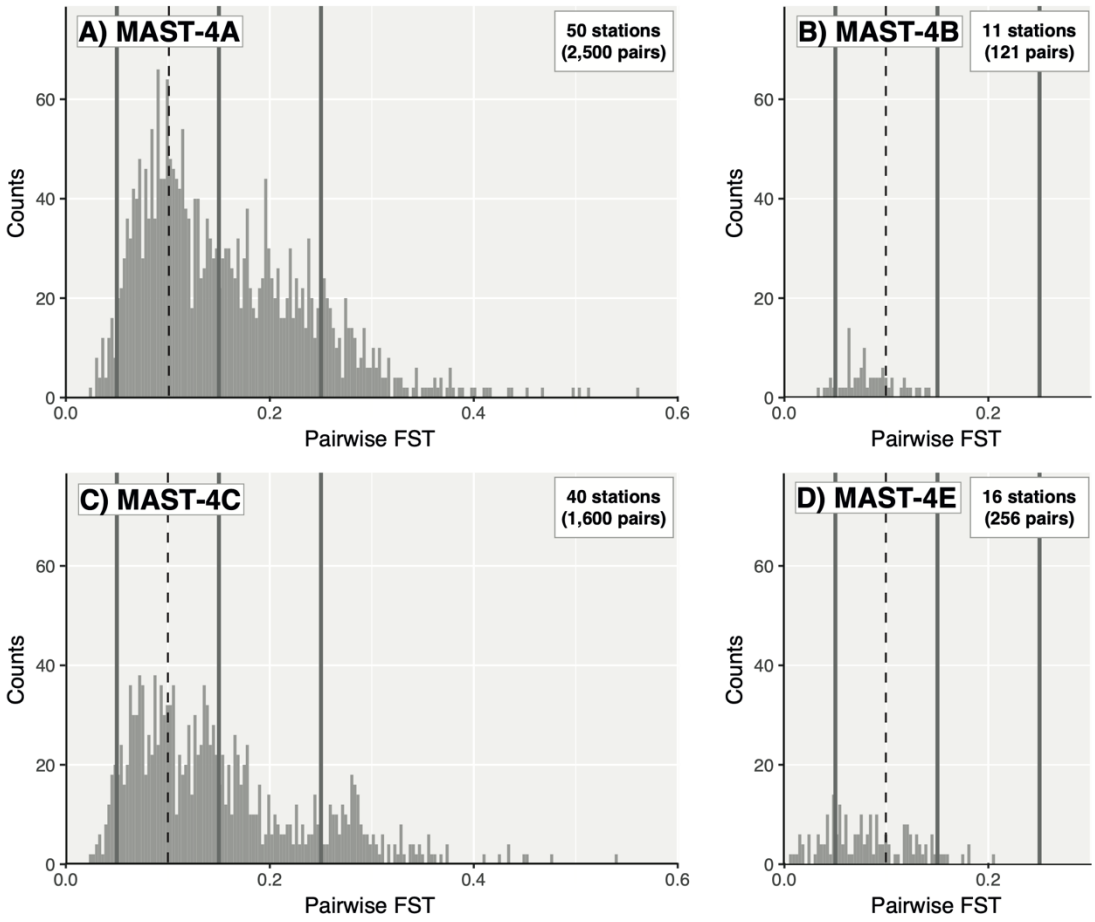

Figure 2.1. Fixation index distribution in the global ocean for MAST-4 species A, B, C & E. Histograms of all FST values among Tara Ocean stations (featuring horizontal coverage ≥ 25%) for **A)** MAST-4A, **B)** MAST-4B, **C)** MAST-4C and **D)** MAST-4E. Note that the four MAST-4 species had their FST distance peaks at the 0.05 – 0.15 range (dashed vertical line).

For each MAST-4 genome, we computed the fixation index (FST) between pairs of Tara Ocean stations. FST measures the differentiation between two populations, with FST values ranging from 0 to 1, where a zero value implies that the two populations show no genetic differentiation, while a value of one implies a high population differentiation[51]. We followed the suggestions of Hartl and Clark[48], who delineated four groups of genetic differentiation based on F_{ST} pairwise values: FST < 0.05, little genetic divergence; 0.05 < FST < 0.15, moderate; 0.15 < FST < 0.25, high; FST > 0.25, very high. From global ocean metagenomes corresponding to a total of 82 locations (stations), we only considered FST values from those that mapped to at least 25% of a given genome (horizontal coverage). Thus, a total of 50/11/40/16 stations (st) were studied for MAST-4A/B/C/E respectively. Only MAST-4 species A and C showed FST

values over 0.25, with a maximum FST value of 0.56 and 0.54 respectively, indicating very high genetic divergence between MAST-4A or C populations in some locations. In contrast, MAST-4B and E always displayed FST values under 0.25, with a maximum FST value of 0.14 and 0.21 respectively. All four MAST-4 species exhibited their FST distance peaks at the 0.05 – 0.15 range (moderate genetic divergence) (**Figure 2.1**). All in all, these analyses demonstrated a substantial genetic divergence among the most abundant MAST-4 species in the global ocean, A and C.

2.3.3 Environmental heterogeneity and genetic divergence

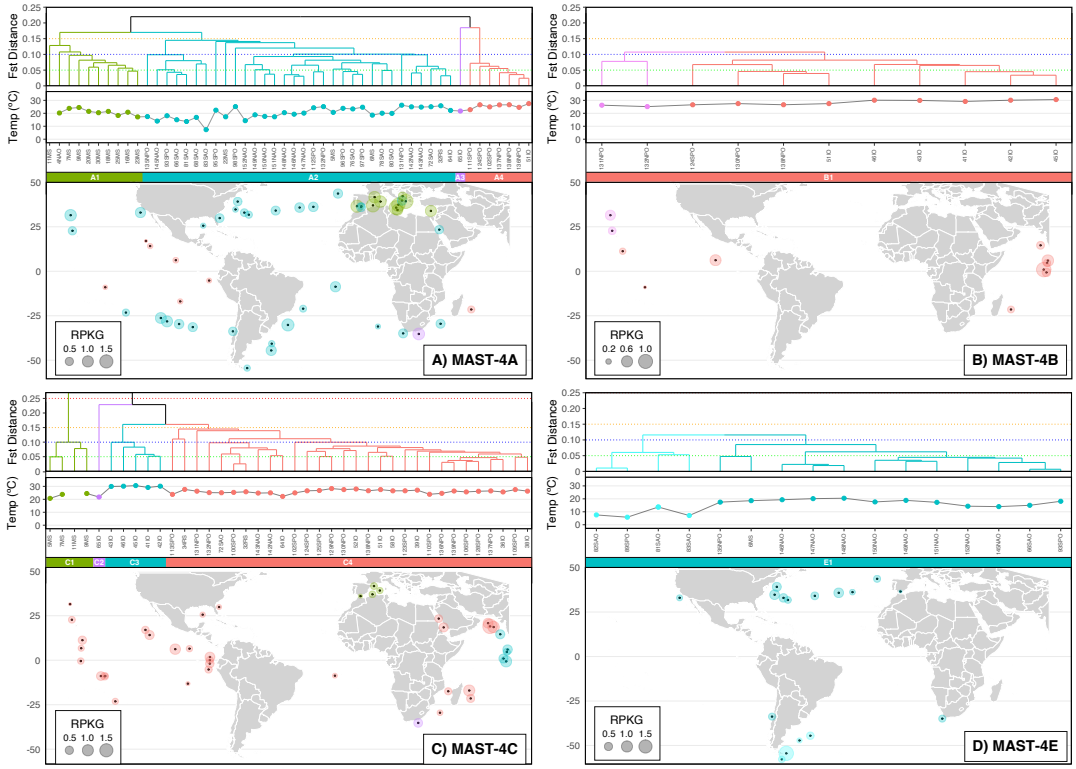

Figure 2.2. Genomic populations of MAST-4 species. Clustering of Tara Ocean stations based on FST values for A) MAST-4A, B) MAST-4B, C) MAST-4C and D) MAST-4E. For each species, a dendrogram of clustered (UPGMA) FST values is shown for metagenomes (stations) that mapped at least 25% of each genome, along with the corresponding surface water temperature. The colors in the dendrograms, temperature sub-panels, bubbles, and those in the horizontal bar in panels A and C indicate genomic populations delineated using an FST > 0.15 threshold, whole colors in panels B and E indicate genomic populations using an FST > 0.10 threshold. Each population is identified with a letter and number in the colored horizontal bar. Bubble size represents normalized species abundance (RPKG) for a given station. Station name tags include the Tara Ocean station number and an acronym of the ocean region to which they belong (MS – Mediterranean Sea; RS – Red Sea; IO – Indian Ocean; SAO – South Atlantic Ocean; SO – Southern Ocean; SPO – South Pacific Ocean; NPO – North Pacific Ocean; NAO – North Atlantic Ocean).

Overall, among the measured environmental variables, temperature and salinity were the most important in explaining the population-level differentiation within MAST-4 A, B, C & E (that is, FST values among TARA stations). Temperature was the main driver of MAST-4B, C, and E population differentiation, explaining 37%, 20%, and 60% (PERMANOVA, p-value < 0.001) of its variance in the global ocean. In turn, for species A, salinity was the main driver, explaining 30% of the population-level differentiation, while temperature explained 13% of it (p-value < 0.001).

MAST-4 A and C displayed differentiated populations in open ocean surface waters when clustering stations based on FST values and a threshold (average FST) of 0.15 (**Figure 2.2**). MAST-4A displayed a total of 4 populations. In sub-tropical waters, it showed one population in the Mediterranean Sea, A1 (10 st), and a second population encompassing the rest of the sub-tropical locations, A2 (32 st). A third population included only station 65, A3, near the South African coast (**Figure 2.2A**), while the last population was detected in tropical waters, A4 (7 st), where MAST-4A has low abundance. MAST-4C also showed 4 populations: one population in the Mediterranean Sea (4 st), C1; two populations in tropical waters, where MAST-4C is most abundant, including C3, present in the Arabian Sea (Indian Ocean; 5 st), and C4 present in the rest of the tropical stations (30 st). MAST-4C also showed one population, C2, present only in station 65 (South African coast) (**Figure 2.2C**).

In contrast to MAST-4A and C, MAST-4B and E appeared to have one single population each (B1 and E1) across the surface global ocean when considering an FST threshold of 0.15. However, MAST-4B was more abundant in tropical waters (**Figure 2.2B**), while MAST-4E was more abundant in sub-tropical and sub-polar waters, usually close to the coast except for some North Atlantic Ocean locations (**Figure 2.2D**). Nevertheless, if we decrease the FST threshold to 0.10, then two sub-populations for both MAST-4B and E emerge. Sub-population B1.1 was present in two stations of the North Pacific Ocean (2 st), while B1.2 was present in the North and South Pacific Oceans as well as in the Indian Ocean (9 st). Then, sub-population E1.1, encompassing South Atlantic and Pacific Ocean stations (4 st), and E1.2, including the Mediterranean Sea, Atlantic and Pacific Ocean locations (12 st). Summarized information of each MAST-4 population, including average temperature and salinity values, is available in **Annex B Table 4**.

2.3.4 Detecting population adaptation

When analyzing populations present in different environments, genes with a dN/dS > 1 (positive selection) may be representing the basis of adaptation. Therefore, we identified the genes with an average dN/dS > 1 in MAST-4 A and C populations. For MAST-4A, a total of 581, 189, 679 and 102 genes showed positive selection in populations A1, A2, A3 and A4 respectively. Yet, only 13 genes were exclusively selected in A1, 1 gene in A2 and 37 genes in A3. Similarly, for MAST-4C, populations C1, C2, C3 and C4 displayed a total of 93, 283, 195 and 47 positively selected genes, from which only 4, 20 and 10 were exclusive to populations C1, C2 and C3. Finally, the species that displayed a single population, that is MAST-4B and MAST-4E's populations B1 and E1 had 21 and 23 positively selected genes respectively. Even though many of these genes selected in specific populations were annotated with the eggNOG database, most annotations were environmental sequences with no function associated. The few genes that matched to a reference with a known function included housekeeping or metabolism-related functions and were not conclusive regarding positively selected functions in specific populations (**Annex B Table 5**).

MAST-4 genes were clustered based on their dN/dS ratio similarities across stations to detect broad patterns. For each MAST-4 species, genes were compelled into 50 clusters of variable size (**Figure 2.3**). Clusters were named following the CXSY formula, where X is the cluster number (1 – 50) and Y is the number of genes within the cluster. All dN/dS values for each cluster and station were computed from the average of all the genes within a cluster. These gene clusters were grouped into bundles based on the dN/dS patterns across samples; that is, from bundles with very low (ratio 0 – 0.2), low (0.2 – 0.5), and intermediate (0.5 – 0.8) dN/dS values across all samples, to bundles showcasing dN/dS values close to 1 or greater either across all samples or in specific locations (populations). Overall, a total of 5, 4, 4 and 6 groups of clusters were defined for MAST-4A, B, C, and E, respectively (**Figure 2.3**).

Specifically for each MAST-4 and gene cluster, MAST-4A featured on average the highest quantity of positively selected genes. Clusters C15S81, C34S113 and C36S122 had an average dN/dS > 1 in populations A1, A2 and A3 (316 genes total). Population A4 did not show clear patterns of positive selection in any of the genetic clusters, except for some isolated stations. In particular, the Mediterranean population

(A1) appeared to have specific patterns of positive selections for cluster C13S119 but also registered greater positive selection for clusters C46S112, C49S114 and C1S136 (total of 362 genes) in comparison to the other sub-tropical populations (A2 and A3). (**Figure 2.3A**). Even though MAST-4C had a similar number of predicted genes to MAST-4A, it displayed on average higher stabilizing selection (dN/dS < 1), contrasting with the overall higher positive selection detected in species A, which showed more gene clusters with a mean dN/dS > 1. Clusters C26S3488, C32S4237, C11S691 and C49S2003 had a mean dN/dS close to 0, accounting for 9,728 genes (64.1% of the total). Still, some genetic clusters seem to have experienced positive selection: clusters C36S42 and C23S92 (134 genes) showed positive selection across all genomic populations, but particularly in C2, C3 and C4, while clusters C7S58 and C44S55 (113 genes) were specific to the Indian Ocean (C3) and C47S76 to the Mediterranean (C1) and the Red Sea (stations 38 and 39 from C4) (**Figure 2.3C**).

MAST-4B population B1 showed a more erratic distribution of positive selection across stations. Except for clusters C32S13, C6S12, and C31S18 (a total of 43 genes) that showed clear patterns of selection not associated with specific basins, the remaining 47 clusters displayed dispersed peaks of positive selection across stations (**Figure 2.3B**). Lastly, MAST-4E, although displaying only one population according to our FST analyses, showed clear dN/dS differences between specific North Atlantic stations (from station 145 to 150) and the rest of the analyzed locations. Some examples were genetic clusters C5S31, C20S45, C37S42, C12S31 and C36S26, which displayed greater dN/dS values in the North Atlantic. In contrast, only cluster C7S18 showed an overall positive selection across the whole E1 population (**Figure 2.3D**).

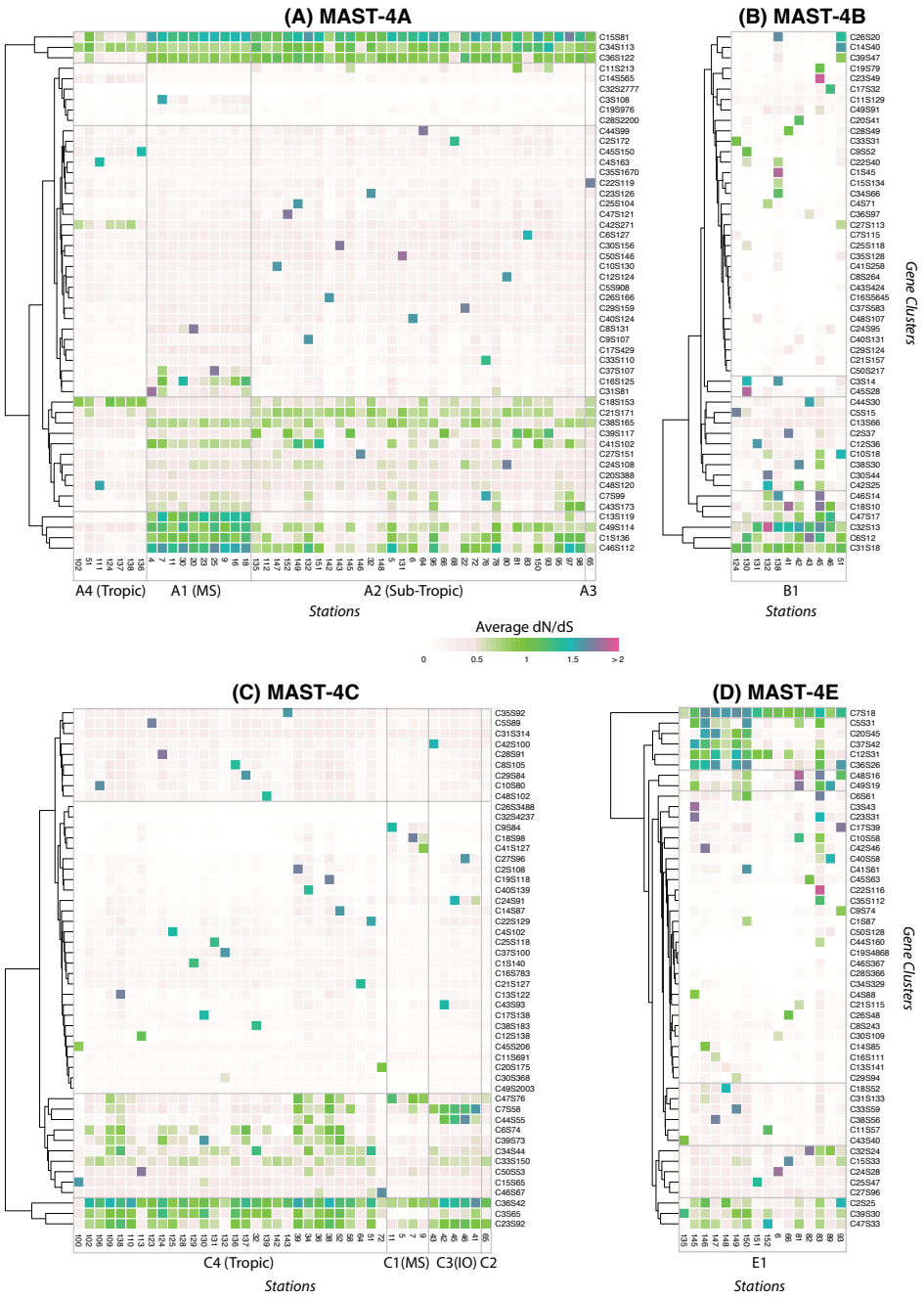

Figure 2.3. Distribution of MAST-4 gene clusters across genomic populations and Tara Oceans stations. Genes for each MAST-4 species were clustered based on similarities in dN/dS ratios across stations (UPGMA with "Manhattan" distance). For an easy representation, the resulting dendrogram was cut at 50 clusters. Colored tiles represent the average dN/dS values of the clusters per station. Gene cluster names are indicated as CXSY: where X is the cluster number (1 to 50) and Y is the number of genes within the cluster. Stations are grouped based on genomic populations; some genomic populations have a tag indicating the ocean region to which they belong (MS – Mediterranean, IO – Indian Ocean).

In terms of metabolic functions, none of the clusters analyzed for species A, B and E showed an enrichment of specific eggNOG functions as genes belonging to distinct functional categories were found in all clusters (**Annex B Figure 1**). Additionally, except for a handful of gene clusters, all of them had at least one gene

with an unknown function. In contrast, for MAST-4C, the cluster C36S42 (positively selected in all genomic populations) showed a larger proportion of genes within the *Chromatin structure and dynamics* category, while clusters C7S58 and C44S55 (positively selected in population C3) both displayed a larger proportion of genes within the *Carbohydrate transport and metabolism* category (**Annex B Figure 1C**). Still, although no particular enrichment was detected among species A, B, and E, some functional patterns were observed: a) MAST-4A gene cluster positive selected in the Mediterranean Sea (population A1) exhibited genes related to *Replication, recombination and repair*, *Transcription* and *Translation, ribosomal structure and biogenesis* categories (**Annex B Figure 1A**); b) MAST-4B's most positive selected gene clusters included genes from *Cell cycle control, cell division, chromosome partitioning* and *Cytoskeleton* categories (**Annex B Figure 1B**); and c) MAST-4E presented genes belonging to *Secondary metabolites biosynthesis* and *Amino acid and Inorganic ion transport and metabolisms* (**Annex B Figure 1D**).

2.4. DISCUSSION

The amount of genetic diversity in populations of marine microbes, and how that diversity is structured over space and time are still, for the most part, a matter of speculation. Here, we investigated the genomic diversity and population differentiation of four species of a key lineage of unicellular ocean predators, MAST-4, across the global ocean. We found that the number of variants per kilobase (Kb) in MAST-4 varied significantly among species, although on average, 87% of them were SNPs. The two most abundant MAST-4 species with the bigger genomes (A and C) had higher numbers of variants (18.2 and 14.0 variants per Kb of genome respectively) than the less abundant species with smaller genomes (B and E; 4.5 and 4.5 variants per Kb of genome respectively), and it could be hypothesized that the number of variants is correlated with either species abundance or genome size. Nevertheless, the number of genomic variants seems to be independent of genome size, as different microbial species with genomes in the same size range have been reported to present a contrasting number of variants. For example, the marine picoeukaryote *Bathycoccus prasinos*, featuring a genome (~ 15 Mb) that is half the size of MAST-4B and E (~ 30 Mb), displayed up to 200,000 SNPs (13.3 SNPs per Kb) (96). However, *Bathycoccus* is a relatively abundant species with a widespread distribution more comparable to those of

species A and C (53,61). Then, the archaea *Sulfolobus islandicus* with a genome size of ~ 2.6 Mb, and a maximum FST (between 97 strains) similar to those of MAST-4A and C (max FST > 0.5), displayed ~ 8,100 SNPs (~ 3.1 SNPs per kb) (240), but with a more restricted habitat (volcano springs and hot habitats) (241). Lastly, over 1 million genomic variants were predicted using 103 Tara Ocean metagenomes and 21 population genomes of SAR11, one of the most abundant bacteria on the planet with a genome size (~1.2 Mbp) smaller than those of MAST-4 (88). Overall, we could theorize that abundance and distribution appear to be important factors determining the number of genomic variants; abundant organisms may have more opportunities for mutations to generate variants, while wide-spread organisms have potentially adapted to more environmental conditions, resulting in a larger number of variants in the genome in comparison to other organisms with restricted habitat.

Our results pointed to strongly differentiated populations in some MAST-4 species, which is counterintuitive, considering the few limitations to dispersal in the surface open ocean. A previous population genetics study based on the 18S rDNA-ITS1 markers indicated a clear spatial structuring in MAST-4A and E. It already suggested the possibility of some sub-clades or populations, which were clearly driven by seawater temperature, with samples as far as the Norwegian Coast and the Pacific West coast showing a very high genetic flux (135). Here we went further, and instead of using specific markers to infer populations, we delineated them through the analysis of SNPs distributed over the entire genome. By using FST distances, we were able to identify four genomic populations for each MAST-4A and C. On the contrary, we were only able to predict one single heterogeneous population for both MAST-4B and E. Thus, our results based on genomics evidenced a larger number of MAST-4 populations than previous studies based on specific markers (135).

The FST index has been used in the past to delineate genomic populations in a wide range of marine life forms, from fish (223,242) to microbes (75,88)[13,55]. It can range from 0 to 1, and values above 0.25 indicate high genetic differentiation[8]. We found two distinct patterns based on the maximum FST divergence exhibited by the four MAST-4 species. *First*, we found a large maximum divergence for MAST-4A and C (maximum FST of 0.56 and 0.54 respectively), which were significantly above 0.25 but slightly below those observed among allopatric populations of the diatom *Picea pungens* (maximum FST = 0.76) featuring strong limitations to gene flow (94). *Second*,

we found moderate divergence patterns for MAST-4B and E (maximum FST of 0.14 and 0.21 respectively), which were similar to other cosmopolitan marine organisms with no geographical limitations to gene flow, such as the diatom *Thalassiosira rotula* (maximum FST = 0.139) (95). Thus, not only the obtained FST values for the four MAST-4 species were in the range of those reported for other marine microorganisms but also suggest that MAST-4A and C (greatest FST), in contraposition with MAST-4B and E (lower FST), have experienced larger population diversification due to a) adaptation to a broader range of niches, b) stronger dispersal limitation or c) both processes.

The emergence of population differentiation in the open global ocean could be explained by limitations to gene flow or adaptation to different environmental conditions. Physical barriers, such as oceanic currents, or geographic distance, are known to limit the dispersal of marine plankton and prevent gene flow, and these processes may have promoted population divergence in MAST-4 (75). Although it is assumed that MAST-4 has a high-dispersal capability in the open ocean (52,228), recent studies demonstrated that the global ocean surface picoplankton is strongly affected by dispersal limitation (44). In addition, local adaptation to different niches may also be promoting population differentiation in MAST-4 (61). Previous works have shown that temperature is the main driver structuring the biogeography of not only MAST-4 species, but also that of other MASTs in the open ocean (52,61,131)[24,25,31]. Yet, little was known about whether temperature affects MAST-4 population structure. We found that population-level genomic divergence in MAST-4A/B/C/E was significantly correlated with temperature and salinity across the surface global ocean, pointing to these variables as main factors structuring the populations of MAST-4. The two most abundant MAST-4 that we studied (A and C) displayed contrasting patterns: while the population-level variation of MAST-4A was mostly structured by salinity (explaining 30% of the FST variance), population variation of MAST-4C was mostly structured by temperature (20% of the FST variance). When translated into the delineated populations, we observed that the most abundant MAST-4A population, A1 in the Mediterranean Sea, had a mean salinity of 3 – 4 PSU greater than the other populations (both sub-tropical and tropical), while temperature was consistent between sub-tropical populations (~ 20 ºC in A1, A2, and A3). Meanwhile, MAST-4C population C3 in the

Indian Ocean had a mean temperature of 30.0 ºC, more than 3 ºC higher than the rest of the MAST-4C populations.

Even though we could not determine strong individual populations in MAST-4B and E, their population-level genetic variation was predominantly structured by temperature (explaining 37% and 60% of their FST variance respectively). Nevertheless, when using a lower FST threshold (0.10), two subpopulations could be identified for both MAST-4B and E. In MAST-4E and MAST-4B, the association between subpopulations and temperature was evident: subpopulation E1.1 in the southern sub-polar region encompassed stations with a mean temperature of 8.55 ºC, around 9 ºC lower than subpopulation E1.2 in the northern sub-tropical area; while subpopulation B1.1 in sub-tropical waters (North Pacific Ocean) showed a mean temperature 3 º C lower than B1.2 in the tropics. Salinity and temperature have been previously reported as variables structuring microbial populations. For example, in bacterioplankton populations in the Baltic Sea, where salinity is the main driver of population structure in many organisms (89); also, in *Prochlorococcus* ecotypes in the Atlantic Ocean, where temperature was significantly correlated with ecotypes abundances (243); and in SAR11 strains adapted to different current temperatures (88). Similarly, in the eukaryote domain, *Bathycoccus prasinos* also appear to be adapted to temperature in the North Atlantic at different depths (96). Overall, the population structure of MAST-4 in the global ocean is comparable to those found in other microbes.

Unlike temporal series where changes in the genomic variants can be studied over time, in a spatial dataset such as the Tara Oceans one, the variants only represent a snapshot in time. Thus, we cannot determine whether the detected variants are under positive or negative selection (that is, increasing or decreasing their frequencies) or not affected by it. Yet, we can estimate whether some variants have been experiencing positive ($dN/dS > 1$) or stabilizing ($dN/dS < 1$) selection. Originally, the dN/dS ratio was developed to quantify selection in orthologs. For microbial population genomic studies based on metagenomics, the relationship between selection and dN/dS ratios may be difficult to infer since it was originally designed for distantly diverged genetic sequences (244). Nonetheless, we calculated the dN/dS ratio for each MAST-4 gene aiming to identify those that may represent the basis of differential adaptation between populations. We detected positively selected gene clusters in all MAST-4 species that were associated with genomic populations and oceanic basins, for example in the

Mediterranean population (A1) of MAST-4A or the North Atlantic samples of MAST-4E (E1.2). Analyzing the functional role of selected genes can provide insights into the metabolic functions that have been the focus of selection.

There have been successful attempts of studying selection in functional pathways in marine microbial communities using dN/dS. Genes coding core functions showed more purifying selection compared to the average genes, while anti-microbial resistance genes had the highest dN/dS values (*i.e.,* diversifying selection) in bacterial Metagenome Assembled Genomes (MAGs) from Baltic Sea metagenomes (89). Although half the genes of MAST-4 had an unknown function, our results on positive selection of gene clustesr hinted at functional patterns that may be relevant to MAST-4 population structure. Genes involved in the *replication, transcription, and translation* of DNA and RNA in MAST-4A; *carbohydrate transport and metabolism* in MAST-4C; and *secondary metabolites biosynthesis* and *amino acid and inorganic ion transport and metabolism* in MAST-4E, are all functional categories that have been reported to be regulated by temperature changes in the bacteria *Sphingopyxis alaskensis* (245). Also linked to cold adaptation, mutations in a protein transporter catalyzing the export of cations are also found in *B. prasinos* (96).

Regardless of dN/dS, these measures only take into consideration mutations that occur in coding regions of the genome (exons) *i.e.,* variants that can result in a modified amino acid sequence, either altering their final function, its expression patterns or its tertiary structure (246). Yet, ~ 80% of MAST-4 variants are located in other regions, and those also have the potential to alter the expression of genes (247). Additionally, synonymous mutations have also been found to affect mRNA expression and alter fitness in yeast despite not changing protein sequence (248,249), implying that genes with a rather low dN/dS might be crucial for adaptation. Considering how relevant is differential gene expression in response to heat and salinity changes in the environment for other microorganisms (250,251), we can hypothesize that non-coding and synonymous mutations that change the expression of genes may also be contributing to shaping MAST-4 population structure in the surface global ocean. However, in-depth analyses of gene differential expression are needed to draw conclusions about it.

Our results expand our knowledge about the population genomics of a key unicellular predator in the ocean, MAST-4. In sum, MAST-4A emerged as a species

showing a high genetic divergence and featuring at least 4 genomic populations that adapted to different salinity and temperature optima to thrive in sub-tropical environments. In turn, MAST-4B showed a moderate degree of genetic divergence and a significant structure at the population level driven by temperature in tropical waters. MAST-4C is also dominating in tropical waters with a genetic differentiation significantly driven by temperature, but unlike species B, it displayed highly divergent genetic populations. Lastly, MAST-4E exhibited a moderate amount of genetic divergence, yet it showed evidence of including sub-populations adapted to different temperature ranges within sub-tropical and sub-polar waters. Overall, MAST-4 emerges as a group of unicellular predators that include some species with clearly defined populations that seemed to have emerged due to differential niche adaptation. A better comprehension of these populations can help to better understand marine food webs and their resilience to global change.

CHAPTER 3

Population structure over a decade and across the global ocean

Francisco Latorre[a], Lidia Montiel[a], Vanessa Balagué[a], Ramon Massana[a], Josep M. Gasol[a], Pierre E. Galand[b], Ramiro Logares[a]

[a] Institute of Marine Sciences (ICM), CSIC, Barcelona, E-08003, Catalonia, Spain

[b] Sorbonne Université, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques (LECOB), Banyuls-sur-Mer, France

3.1. INTRODUCTION

Accounting for approximately 10^{30} cells (252), the global biomass of microbes is dominated by prokaryotes (253). In the ocean, bacteria and archaea play key roles in the biogeochemical cycling of nutrients, matter, and energy (11,254). Marine microbes are highly diverse and encompass different lineages able to perform a wide array of complex chemical reactions (255), allowing them to colonize distinct habitats such as pelagic zones, subsurface open ocean waters and sediments (256). Despite their importance, key aspects of the ecology and evolution of marine microbes are poorly understood (257). In particular, we lack a clear understanding of the ecological and evolutionary processes occurring within microbial populations, that is, processes among closely related lineages that could be considered to belong to the same species. Improving our understanding of how prokaryotic populations are structured and become adapted to their environment is critical for developing better predictions of ecosystem dynamics in future scenarios of global change. Fundamental questions that remain partially answered are: how much genetic diversity is present within different microbial taxa? What are the drivers shaping the population structure of marine microbes? How do different populations adapt to environmental heterogeneity (*i.e.,* what is the genomic basis of population adaptation across species)?

Microbial population genomics aims at addressing the previous questions, disentangling the evolutionary history and adaptation of a given species and reconstructing the processes behind the emergence of population structure (258). So far, we know that major drivers that affect microbial communities (*i.e.,* different species) include oceanographic features such as currents, water masses, and the physicochemical characteristics of seawater (259–262), but little is known about the environmental factors shaping the structure of microbial populations (*i.e.,* variants within a species). At large spatial scales, these geographical features can both limit or promote dispersal between ocean basins (44,263), where physical and chemical factors such as temperature, salinity or nutrient concentration may be different, promoting the differential adaptation of microbial populations (264,265). The large population sizes and fast reproductive rates of microbes would support their rapid adaptation to local or regional environmental conditions compared to animals and plants (84). Yet, marine microbes may also display rapid adaptation in constant conditions (266,267) and may evolve faster through non-adaptive (neutral) processes over large spatial scales (268).

Investigating the genomics of microbial populations requires genomes, however these are difficult to obtain, given that most microbes are still uncultured (256,269). In the past decade, decreasing costs of DNA sequencing and increased throughput allowed us to obtain thousands of prokaryotic genomes from uncultured microorganisms directly from the ocean via metagenomics (32,124,270–272). The genomes, so-called Metagenome Assembled Genomes (MAGs), are typically far in terms of quality from gold-standard genomes obtained from cultures (273,274). Yet, they allow to access the genomic information of microbes that otherwise would be impossible to investigate. MAGs have thus greatly improved our understanding of the ocean microbiome. For example, Paoli and colleagues (272) recently demonstrated that a big portion of the biosynthetic potential in the global ocean microbiome belongs to poorly known microbial communities and is only accessible through the reconstruction of MAGs. Another metagenomic work (80) has found that mixotrophy (*i.e.,* the capacity to grow both auto- and heterotrophically) is an ecologically relevant trait found in several MAGs inhabiting the largest aquatic ecosystem of the Earth: the deep ocean.

Despite the increasing availability of marine MAGs and metagenomes (80,270,272), few studies have used them to investigate the structure of microbial populations and their potential adaptations to local environmental conditions. Among the few available studies, those involving the SAR11 group, *Prochlorococcus*, and *Synechococcus* stand out, showing that genomic population structure was not influenced by global dispersal limitation of water masses, but instead correlated with temperature, nutrient and light availability over different ocean regions (88,275–278). Additionally, SAR11 was found to be quite stable across niches due to high recombination rates between close and distant related linages (279). Another work detected genomic differentiation among SAR116 populations, some of which are considered endemic to the Mediterranean Sea and are believed to be adapted to specific environmental conditions, such as phosphate concentrations (280). Metagenomics is thus a powerful tool to study the population genomics of marine microorganisms, but the number of targeted species has remained very limited.

Here, we investigate the population structure of 495 marine microbial species (MAGs) in two long time-series and in the global ocean. These MAGs were recovered from the Blanes Bay Microbial Observatory (BBMO, Blanes, Spain) (100), a coastal site located in the Northwestern Mediterranean Sea. More precisely, we tested if their

genomic diversity varied at different spatiotemporal scales, and verified if environmental factors may be driving the diversity structure. To do so, we first compared the monthly genomic variation of the selected MAGs between two geographically close time-series sites: BBMO, during 12 years, and SOLA (Banyuls sur Mer, France) (281–283) during 7 years. We then investigated the spatial population patterns of the 495 microbial MAGs in the global ocean using 129 metagenomes from the *Tara Oceans* expedition (2009 - 2013) (200). Using these datasets, we ask: How much genomic diversity is found in the analyzed species at the investigated spatiotemporal scales? Is the genomic diversity structured? If so, which environmental factors may be driving it? Finally, what are the main differences in population structure between both time-series, and between the latter and the global ocean?

3.2. METHODS

3.2.1 Metagenomic datasets

Our spatiotemporal analysis includes 3 metagenomic datasets: (i) two long marine-coastal metagenomic time-series in the Mediterranean Sea separated by ca. 130 km [Blanes Bay Microbial Observatory (BBMO), Blanes, Spain (41º40'13N 2º48'0E) (100) and SOLA, Banyuls Bay, Banyuls sur Mer, France (42°29'3N 3°08'7E) (101,284)] and (ii) one metagenomic dataset from the surface global ocean generated during the *Tara Oceans* expedition 2009 – 2013 (200,285). Hereafter, these 3 datasets will be named BBMO, SOLA, and TARA respectively (**Annex C Figure 1**).

The BBMO dataset covers 12 years of monthly samples between January 2009 and December 2020 (140 metagenomes; **Annex C Table 1**) and SOLA comprises 7 years of monthly data between January 2009 and December 2015 (89 metagenomes, **Annex C Table 2**). All metagenomes were cleaned with *cutadapt 1.16* (286) using a quality threshold of 20 for both 5’ and 3’ ends, a minimum length corresponding to half the size of the metagenomic read length and Illumina-trueseq adapters (R1=AGATCGGAAGAGCACACGTCTGAACTCCAGTCA, R2=AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT). TARA and SOLA metagenomes from January 2012 until February 2015 were sequenced with a read length of 100 base pairs (bp), while all BBMO and the rest of SOLA metagenomes were sequenced with a read length of 150 bp. Lastly, all TARA metagenomes belonging to

the same station were concatenated together for simpler downstream analyses. The TARA dataset is composed of surface water samples collected from 82 stations encompassing the 0.2 – 1.6 and 0.22 – 3 μm size fractions during 2009 and 2013 (129 metagenomes; **Annex C Table 3**).

3.2.2 Metagenomic information content

The BBMO MAGs were delineated using a subset of the 12-year metagenomic dataset (7 years of monthly samples, a total of 84 samples; January 2009 to December 2015) due to the high computational demands of processing such a large amount of data. We aimed at co-assembling groups of metagenomes that had comparable information content. To define those groups, we determine pairwise metagenome similarity using simka v1.5.2 (287), with a k-mer size = 21, minimum shannon-index = 1.5, minimum read size = 70, and Bray Curtis dissimilarities. We clustered the Bray Curtis distances using *hclust* and UPGMA in R (288), which allowed us to define four groups (G) of samples corresponding to different times of the year: G1 (winter), G2 (spring), G3 (early summer) and G4 (late summer) with a total of 37, 9, 14 and 22 metagenomes each (**Annex C Figure 2**). Two samples were excluded as they did not belong to any group.

3.2.3 Co-Assembly and reconstruction of Metagenome-Assembled Genomes (MAGs)

Samples from each cluster were co-assembled together using MegaHIT v1.2.8 (289) with presets *meta-large* and 750 gigabytes of RAM. Before the binning step, and to obtain contig abundances across samples, the 84 BBMO metagenomes (years 2009-2015) were back-mapped to the four co-assemblies with BWA v.0.7.17-r1188 (176) in default mode. Unmapped reads, secondary hits, and reads with an alignment quality below 10 were removed using Samtools (235) v1.8 (for G2, G3 and G4) and v1.12 (for G1). The resulting BAM files were used as input to MetaBAT v2.12.1(290), ran in default mode with a minimum contig length of 2.5 kb, to generate four different sets of MAGs. The contig depth values per metagenome from MetaBAT, along with the original BAM files, were given as input to two other metagenomic binners to generate additional MAGs: Concoct v0.4.2 (291) and MaxBin2 v.2.2.5 (292), both ran in default mode and using a minimum contig length of 2.5 kb. Then, to further improve the quality and accuracy of the predicted genomes, all MAGs from the 3 binners were mixed and

refined with the MetaWrap v1.3-4bf5f8a pipeline (293) in default mode. Only refined MAGs with completeness ≥ 50% and contamination ≤ 10% were kept (294). A total of 2,311 MAGs were obtained (909, 347, 457, and 598 MAGs for G1, G2, G3, and G4, respectively), which were taxonomically annotated with GTDB-Tk v1.5 (295) in default mode with the *classify_wf* workflow. To remove redundancy, a genomic de-replication at 99% Average Nucleotide Identity (ANI) was done with dRep v2.3.2 (296) for all MAGs, independently of the original group of samples. In the end, a total of 1,505 high-quality non-redundant MAGs were generated. Gene predictions and functional annotation were carried out with Prokka v1.14.6 (297) and EnrichM v0.5.0's database v10 (298).

3.2.4 Abundance and Horizontal coverage across samples

In downstream analyses, we focused on a selection of 495 MAGs, 61 Archaea and 434 Bacteria, which belonged to the following highly abundant taxonomic groups in BBMO[14]: SAR11 (9 MAGs), SAR324 (5 MAGs), SAR116 (47 MAGs), SAR86 (44 MAGs), Balneolales (5 MAGs), Actinomarinales (7 MAGs) and Flavobacteriales (317 MAGs) (**Annex C Table 4**). To get the abundances across all datasets and to reduce the number of incorrectly mapped reads, we performed a competitive mapping approach. First, all MAGs with the same taxonomy at the order level were concatenated into a single fasta file, generating a total of 8 sets of MAGs (7 bacterial sets and 1 archaeal set; see above). Second, all the clean reads from TARA, BBMO, and SOLA were mapped with BWA against each set of MAGs. Only reads with identity > 95% and alignment coverage > 80% were kept. Third, individual BAM files for each MAG and sample were extracted from the BAM files generated in the second step using Samtools 1.8. Last, RPKG values (mapped reads per kb of Genome per Gb of metagenome) and Genomic Horizontal Coverage (percentage of the genome covered by at least 1 filtered read) were computed for each metagenomic sample with Samtools 1.8. Additionally, to investigate cyclic changes in MAG's abundance, periodicity profiles were calculated for each MAG at BBMO and SOLA using RPKG values. Results were plotted using the Lomb-Scargle Periodogram algorithm (p-value < 0.01) of the *randlsp* function from the *lomb* package (299)[62] under R version 4.0.5

3.2.5 Variant calling and genetic differentiation

To assess the genetic divergence of prokaryotic MAGs through our spatiotemporal datasets, we predicted Single Nucleotide Polymorphisms (SNPs), insertions and deletions across the three datasets. For each MAG and dataset, all BAM files were merged into a single file with the Samtools 1.8 *merge* function. The merged BAMs were then given as input to Freebayes v1.3.1 (236) to perform Variant Calling with ploidy set to 1 (-p 1) and a minimum of 4 observations to support alternate alleles (-C 4). The generated variant call files (VCF) were used in POGENOM v.0.8.3 (89) to compute the Fixation index (FST) values between all samples of each dataset and the non-synonymous vs. synonymous mutation (pNpS) ratios for each gene across all samples. Genes with a mean pN/pS ratio > 0.8 were selected for further analyses. The computation of the mean pN/pS omits NA values (when pS = 0), which results in some genes showing positive selection in a small number of samples (*e.g.,* 1 or 2 samples).

Genomic populations were defined by clustering FST values with the R function *hclust* (300) using the UPGMA algorithm, given a mean for each cluster (hereafter named Mean cFST). A second Mean FST (named Mean FST) was computed for all FST values within a genome and dataset. A Permutational Multivariate Analysis of Variance (Permanova) was carried out using the *adonis* function from the vegan 2.5.7 package (301) to assess what percentage of the variance of FST distances could be explained by changes in temperature and salinity. For this, temperature and salinity values were previously z-score normalized with the *scale* function.

3.3. RESULTS

3.3.1 Overall temporal and spatial genomic differentiation of the MAGs

We assessed genomic divergence (GD) for each one of the 495 MAGs across the two temporal datasets, BBMO and SOLA, and the spatial dataset, TARA (**Annex C Table 4**). We used four levels of genomic divergence based on Fixation Index (FST) distances: Little GD for FSTs < 0.05; Moderate GD for 0.05 ≤ FST < 0.15; High GD for 0.15 ≤ FST < 0.25; and Very high GD for FST ≥ 0.25 (78,79). For the 495 MAGs, genomic divergence in BBMO and SOLA was proportionally almost identical, with ~ 24%, ~ 46%, ~ 15%, and ~ 15% of the FSTs being classified into the Little, Moderate, High and Very high GD categories, while genomic divergence patterns in TARA were contrasting, with the percentages of the same categories being ~ 8%, ~ 15%, ~ 17% and

~ 63% (**Figure 3.1**). Thus, the 495 MAGs displayed moderate genomic differentiation over 7 to 12 years in two neighboring locations in the Mediterranean Sea, but high differentiation across the global ocean.

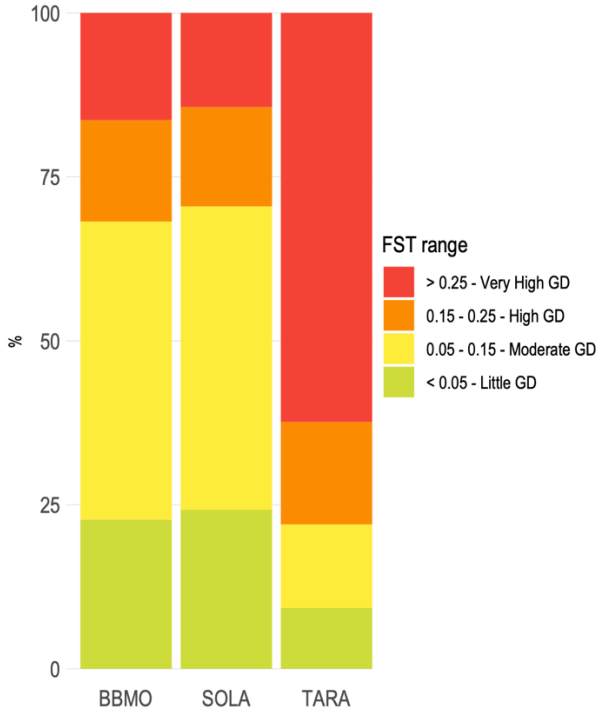

Figure 3.1. Temporal genomic divergence of the 495 MAGs at BBMO and SOLA, and spatial divergence at a global scale in TARA. FST values were classified into four levels of genetic divergence (GD): Little GD for FSTs < 0.05; Moderate GD for 0.05 ≤ FST < 0.15; High GD for 0.15 ≤ FST < 0.25; and Very high GD for FST ≥ 0.25. Only FST values for samples with at least 25% of horizontal coverage of the corresponding MAG were considered. BBMO and SOLA include both 12 and 7 years of monthly surface samples respectively, while TARA includes 82 surface stations from the global ocean from 2009 to 2013.

3.3.2 Individual MAG genomic differentiation

We investigated the genomic differentiation in MAGs that were well-represented in the metagenomic datasets. A total of 169 MAGs were selected out of 495, featuring a horizontal coverage > 25% in at least 20% of the BBMO samples and 10% of the 82 TARA stations (**Annex C Table 5**). SOLA samples were not included during the filtering process as horizontal coverage featured high similarity to BBMO samples. Then, we analyzed the FST patterns across the three datasets for each MAG individually. There was a moderate individual genomic divergence during 12 and 7 years at BBMO and SOLA respectively, within Archaea, Balneolales, Actinomarinales, SAR86, and SAR324 (**Figure 3.2**). In turn, SAR11 and SAR116 MAGs displayed little and moderate genomic divergence, while Flavobacteriales MAGs displayed all four levels of genomic differentiation (**Figure 3.2**). At the global scale (TARA), we found

contrasting patterns. All taxonomic groups displayed a very high divergence, except SAR11 and SAR116 which displayed enriched FST values in the moderate and high categories (**Figure 3.2**).

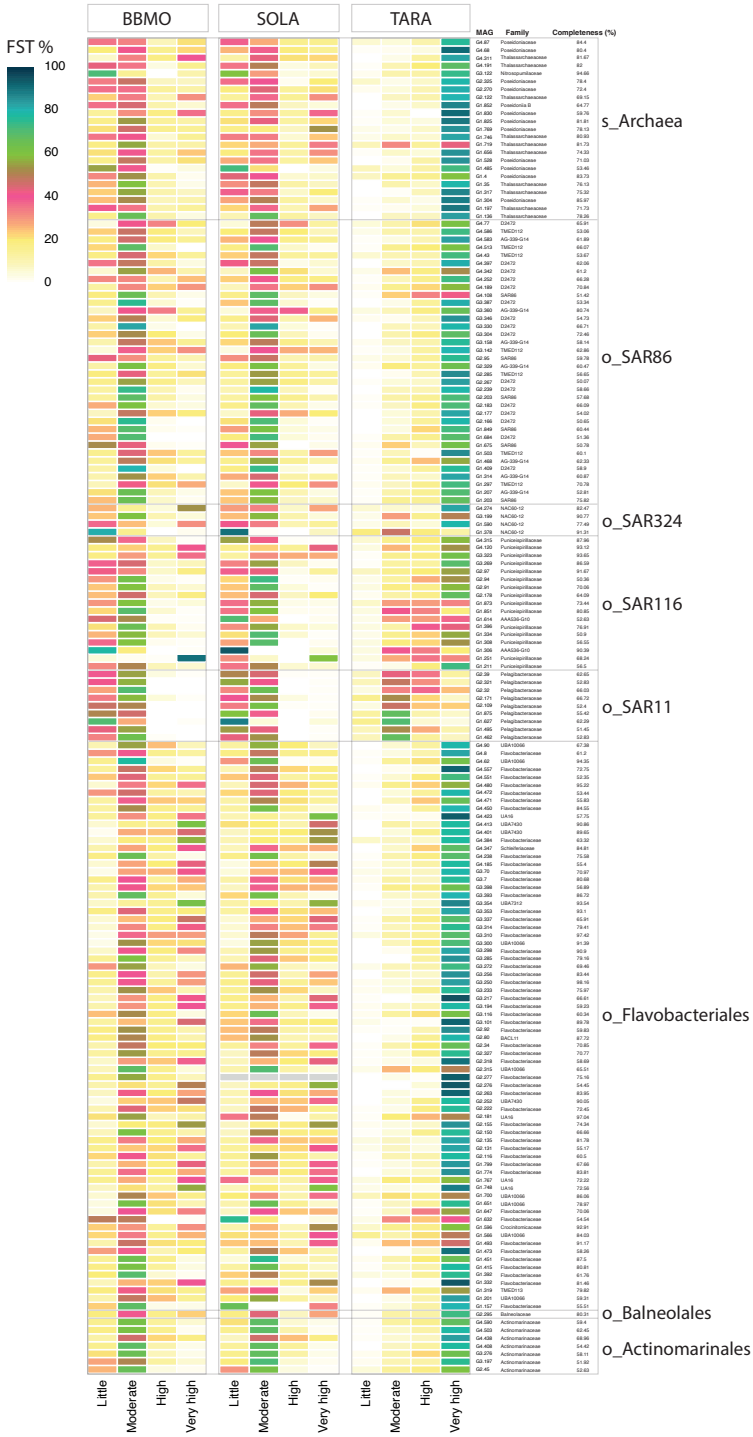

Figure 3.2. Individual MAG genomic divergence based on FST values across time (12 and 7 years at BBMO and SOLA) and space (global ocean TARA). FST values were classified into four levels of genetic divergence (GD): Little GD for FSTs < 0.05; Moderate GD for 0.05 ≤ FST < 0.15; High GD for 0.15 ≤ FST < 0.25; and Very high GD for FST ≥ 0.25. Only FST values for samples with at least 25% of horizontal coverage of the corresponding MAG were considered. BBMO and SOLA include 12 and 7

years of monthly surface samples respectively, while TARA includes 82 surface stations from the global ocean. Colors indicate the proportion of FST values that fall into each category (i.e., Little, Moderate, High, and Very high). For each row, MAG identification codes are included. Family indicates the taxonomic classification at the family level obtained using GTDB (note that for some MAGs, GTDB only provides strain codes and not a formal taxonomic name). Completeness indicates the percentage of genome completeness for each MAG as calculated with CheckM.

3.3.3 Population analyses

We detected genomic populations by clustering all samples of a dataset based on the pairwise FSTs for each of the 169 individual genomes. For that, we defined two main patterns of population differentiation based on the FST mean (named Mean FST to differentiate from the Mean cFST produced by the clustering algorithm): (i) weak, under 0.15, and (ii) strong, above 0.15 (79). In total, 33.1% (56) of the 169 selected MAGs showed strong population differentiation in BBMO, 31.4% (53) in SOLA and 95.3% (161) in TARA. The weak population differentiation over 12 and 7 years at BBMO and SOLA, and a strong one in the global ocean (TARA) was the common trend across the recovered MAGs (**Figure 3.1, Annex C Table 5**). We exemplified this trend using the archaeal MAG G3.122 (*Nitrososphaerales*), which showed weak population differentiation in both BBMO and SOLA (Mean FST of 0.06 and 0.05 respectively) but a strong one in the global ocean (TARA; Mean FST of 0.44) (**Figure 3.3A**). For MAG G3.122, we identified two populations (A and B) in both BBMO and SOLA (Mean cFST ~ 0.15), among which one was abundant in cold waters (winter and spring, population A). Temperature explained most FST variance for MAG G3.122 in BBMO (55% ADONIS; p-value < 0.05) but not in SOLA (ADONIS not significant). In the global ocean, a total of 5 highly divergent genomic populations (Mean cFST ~ 0.3) were defined across sub-tropical and sub-polar waters for MAG G3.122, which were significantly structured by temperature and salinity, explaining 29.4% and 13.8% of its variance, respectively (ADONIS; p-value < 0.05).

In contrast, other genomes exhibited strong population differentiation in both the time-series and in the global ocean (**Figure 3.2**). This behavior is exemplified by the SAR86 genome G1.297 (Mean FST of 0.19, 0.20, and 0.37 in BBMO, SOLA, and TARA, respectively; **Figure 3.3B**). For both time-series, one genomic population was differentiated from the others (FST > 0.25) (Population A, **Figure 3.3.3B**). It corresponded to cold waters (average temperature < 15 ºC) and showed low abundance. The other populations (B, C, D, E, F in BBMO and B, C, D, E, G, H in SOLA) displayed higher abundances and were present in samples of warmer waters (average

temperature ≥ 15 º C). There was a significant correlation between genomic divergence and temperature, which explained 59.2% and 72.1% of the variation in BBMO and SOLA, respectively (ADONIS, $p < 0.05$). In the global ocean, the SAR86 MAG G1.297 showed larger genomic differentiation (FST > 0.25) than in the time-series, with two populations present in the tropical and sub-tropical waters (Mediterranean Sea) **(Figure 3.3B)**. Temperature explained 51.9% of the variance in genomic differentiation (ADONIS, $p<0.05$), while salinity was not significant.

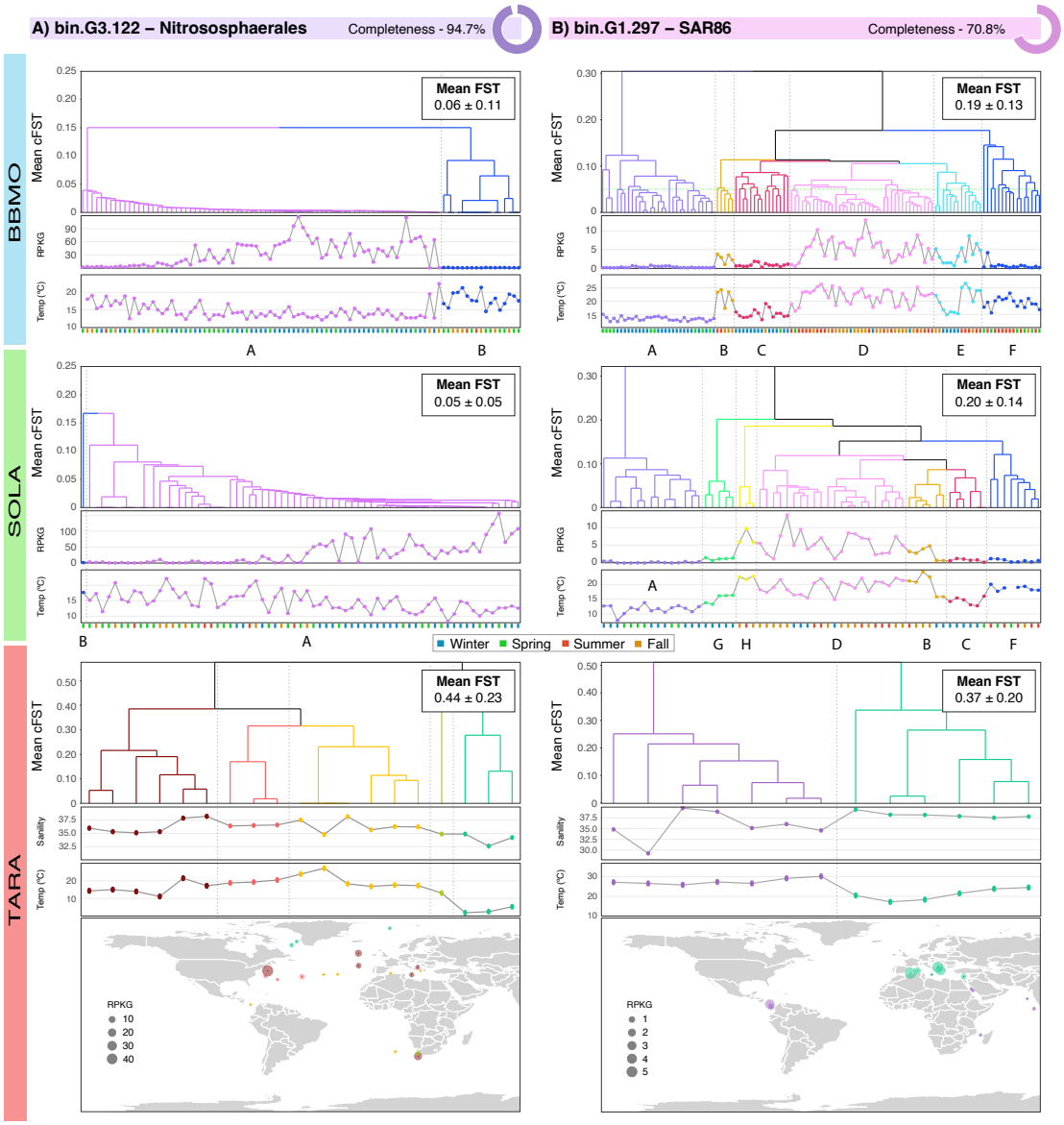

Figure 3.3. Population structure in the two time-series (BBMO and SOLA) and in the global ocean (TARA) for the A) archaea MAG Nitrososphaerales G3.122 and B) bacterial MAG SAR86 MAG G1.297. Populations are defined based on the Mean cFST computed by the clustering UPGMA algorithm (see dendrogram axis), giving the average FST for each cluster. A second Mean FST value indicates the mean (± standard deviation) of all FST values of a genome in each dataset. Populations are indicated with different colors and letters in each dataset. When two populations in BBMO and SOLA are the same, an identical color and letter are assigned to them. Temperature (ºC), salinity (PSU) and abundance (RPKG,

reads per kilobase of genome per gigabase of metagenome) are given accordingly. Note that salinity is not included in BBMO and SOLA due to limited variation. Colors in the x-axis in BBMO and SOLA indicate to which season each sample belongs. The color of the bubbles on the map indicates the presence of a given population in a specific geographic zone and the size of the bubble, its abundance (RPKG). Completeness refers to genome completeness as calculated with CheckM and is also visualized with the circle.

Another example of MAGs showing strong population differentiation in both, the time-series and the global ocean (**Figure 3.2**, **Annex C Table 5**), was Flavobacteriales G4.480 (Mean FST of 0.21, 0.16, and 0.31 in BBMO, SOLA, and TARA, respectively) (**Figure 3.4A**). It separated into a total of 7 and 5 genomic populations (Mean cFST ~ 0.15) in BBMO and SOLA respectively (**Figure 3.4A**). Populations corresponded to different seasons in the temporal datasets and were significantly related to temperature, which explained 52.9% and 49% of its variation in BBMO and SOLA, respectively (ADONIS, p-value < 0.05). Yet, this Flavobacteriales MAG was only abundant (RPKG > 1) in autumn and a few winter samples. In the global ocean, this MAG populated both tropical and sub-tropical waters with 10 detected genomic populations that were correlated with temperature and salinity, which explained 19.9% and 4.9% of its variation, respectively (ADONIS; p-value < 0.05). Some of these populations seemed to be predominant in the Mediterranean Sea, another in the Red Sea, the eastern region of the North Atlantic Ocean, or the South Pacific Ocean (**Figure 3.4A**).

An additional pattern that we detected among the studied MAGs was weak population differentiation in both the time-series and in the global ocean (**Figure 3.2**). This pattern was observed in 6 out of 9 SAR11 MAGs (**Annex C Table 5**). All SAR11 MAGs were relatively abundant throughout the years at both BBMO and SOLA, but displayed differentiated cold and warm water populations at both Mediterranean locations. In turn, in the global ocean, population differentiation in all SAR11 MAGs was either weak or barely strong (Mean FST < 0.2) with their presence restricted to sub-tropical waters (except for G1.495, which also appeared in sub-polar waters) (**Annex C Table 5**). These patterns are exemplified by SAR11 MAG G2.171, which showed Mean FST of 0.07, 0.07, and 0.11 for BBMO, SOLA, and TARA, respectively (**Figure 3.4B**). The population structure of SAR11 MAG G2.171 was significantly correlated with temperature in both time-series, which explained 66.1% and 32.9% of the variation in genomic differentiation (ADONIS, $p<0.05$). In the global ocean, temperature and salinity explained 16.7% and 25.2% of the population structure, respectively (ADONIS,

p<0.05). Although genomic divergence was low (< 0.15) we detected 4, 6 and 4 genomic populations for BBMO, SOLA, and the global ocean, respectively (**Figure 3.4B**). Populations B, C, and E in BBMO and SOLA were predominant in cold waters; populations A, F, and G predominated in warm waters, while population D predominated in both. Regarding TARA, all populations were found in sub-tropical waters, with clear differentiation between Mediterranean populations and the rest in the sub-tropic (**Figure 3.4B**)

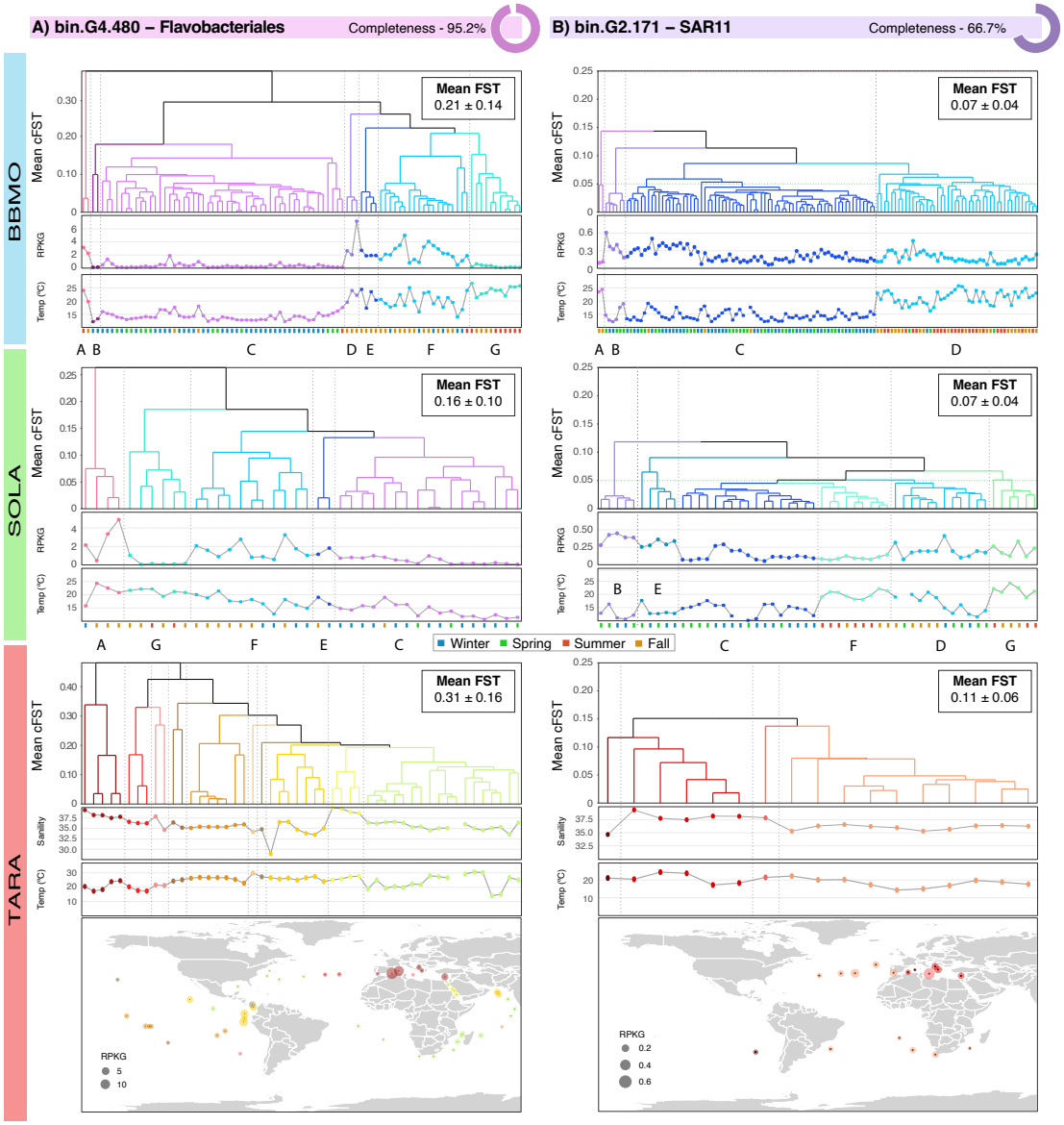

Figure 3.4. Population structure in the two time-series (BBMO and SOLA) and in the global ocean (TARA) for the A) bacterial *Flavobacteriales* G4.480, and B) bacterial SAR11 G2.171. Populations are defined based on the Mean cFST computed by the clustering UPGMA algorithm (see dendrogram axis), giving the average FST for each cluster. A second Mean FST value indicates the mean (± standard deviation) of all FST values of a genome in each dataset. Populations are indicated with different colors and letters in each dataset. When two populations in BBMO and SOLA are the same, an identical color and letter are assigned to them. Temperature (ºC), salinity (PSU) and abundance (RPKG, reads per

kilobase of genome and gigabase of metagenome) are given accordingly. Note that salinity is not included in BBMO and SOLA due to limited variation. Colors in the x-axis in BBMO and SOLA indicate to which season each sample belongs. The color of the bubbles on the map indicates the presence of a given population in a specific geographic zone and the size of the bubble, its abundance (RPKG). Completeness refers to genome completeness as calculated with *CheckM* and is also visualized with the circle.

3.3.4 Positive selection and population differentiation

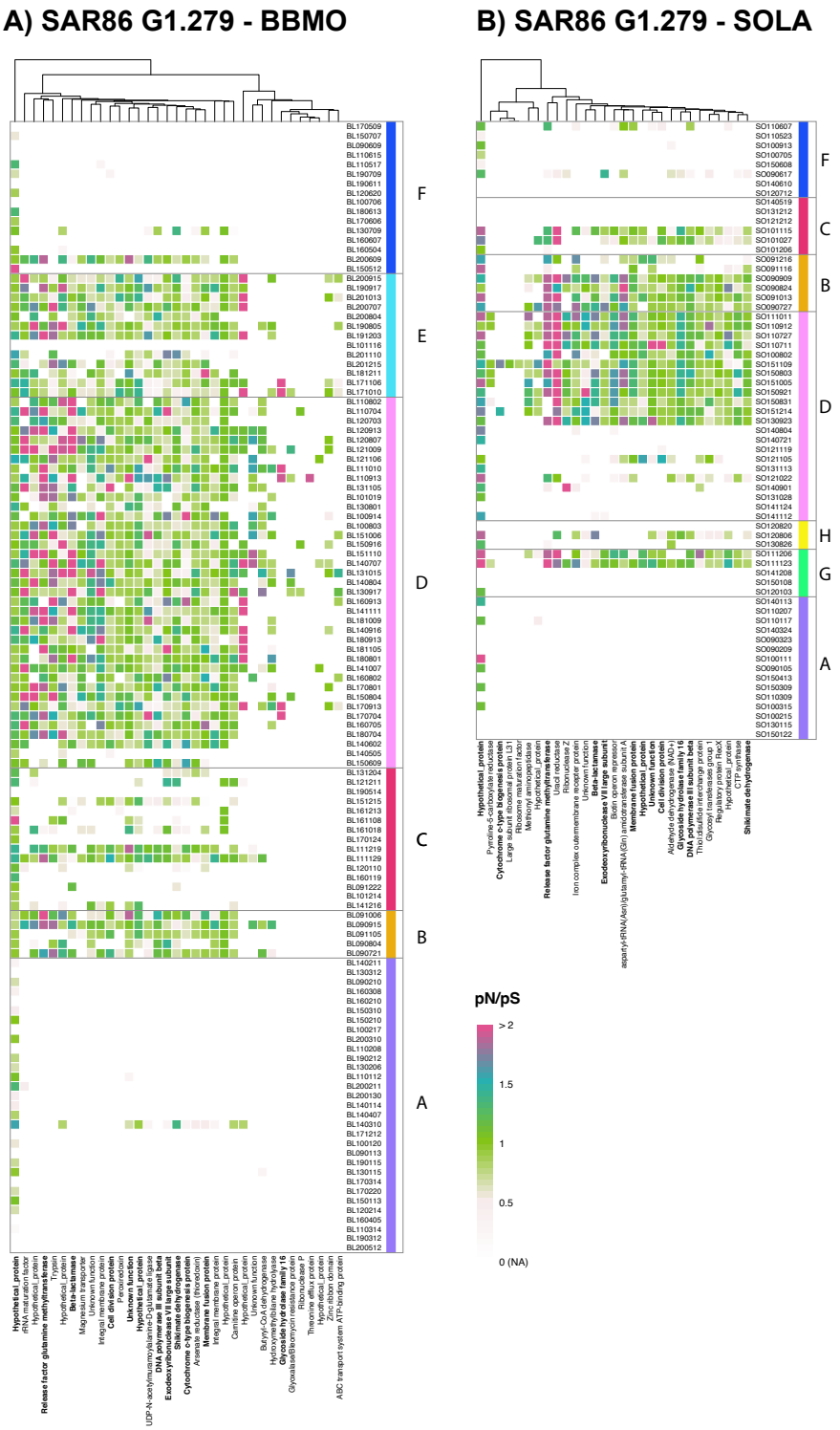

Figure 3.5. Positive selection patterns for genes with a mean pN/pS > 0.8 for MAG SAR86 G1.279 from A) BBMO and B) SOLA. The cell colors indicate the pN/pS of a gene in a given sample. White tails indicate NA values (pS = 0 in pN/pS calculations) due to lack of mapping in those samples. Mean

pN/pS is computed omitting NA values, resulting in genes found in a small set of samples (*e.g.,* 1 sample) having a mean pN/pS > 0.8. Genes in bold indicate those that are shared between BBMO and SOLA. Samples are grouped based on the population they belong to. The colors and letters of the bars grouping samples match those of the genomic populations from **Figure 3.3B**. TARA data is now shown due to only being 5 genes with exclusive positive selection in Mediterranean Sea samples.

We analyzed the potential action of positive selection in the 169 genomes to understand if the observed population differentiation is the result of adaptive processes. We analyzed the ratio of non-synonymous to synonymous mutations (pN/pS) for all genes of each MAG in both time-series and in the global ocean. We focused on genes with a mean pN/pS > 0.8 for each MAG across samples. The number of positively selected genes per Mega bases (Mb) of genome was higher in the time-series (BBMO and SOLA) than in the global ocean, despite the global ocean (TARA), despite the global ocean showing the highest genomic divergence. Among the genomes with the largest proportion of positively selected genes (> 13 genes/Mb of genome) across all datasets were the Flavobacteriales and SAR116 MAGs, followed by archaeal, SAR86 and SAR324 MAGs (**Annex C Table 6**). On the contrary, the SAR11 genomes displayed a low proportion of positively selected genes (< 6 genes/Mb of genomes) (**Annex C Table 6**). The genome with the overall highest proportion of positively selected genes was the Flavobacteriales G3.250, with 68.7, 59.8, and 28.5 genes/Mb in BBMO, SOLA, and TARA, respectively. SAR86 G1.297 (**Figure 3.3B**) and Flavobacteriales G4.480 (**Figure 3.4A**) were both among the MAGs with the largest proportion of positively selected genes, with 21.8/21.8/2.3 and 38.4/31.9/5.5 genes/Mb in BBMO, SOLA, and TARA, respectively. Other genomes showed no positively selected genes, such as SAR11 G1.627 and Flavobacteriales G1.157, or a small number of them (< 5 genes) (**Annex C Table 6**). In particular, SAR11 G2.171 (**Figure 3.4B**), had only 2 positively selected genes in both SOLA and the global ocean (3.23 genes/Mb each).

Even though the proportions of positively selected genes in the two time-series were different, some of the positively selected genes in both datasets were identical. In general, out of 169 MAGs, 49 (29 %) shared $\geq$ 50% of the positively selected genes in BBMO and SOLA (BBMO $\cap$ SOLA / BBMO $\cup$ SOLA), 76 (45 %) shared < 50% and 44 (26 %) showed no shared genes (**Annex C Table 6**). An example of a genome showcasing general patterns of positive selection (*i.e.,* several positively selected genes in both time-series with less than 50% of them shared) is SAR86 G1.297, which had a total of 35 and 29 positively selected genes at BBMO and SOLA, from which 12 were shared in both time-series (23.1% of the total of positively selected genes). These

included genes for the peptide chain release factor methyltransferase (*PrmC*), a beta-lactamase, a membrane fusion protein, a Glycoside hydrolase from family 16 (GH16), other proteins related to the cell cycle and division, and three hypothetical proteins with unknown functions (**Figure 3.5**). One gene, encoding for one of the hypothetical proteins, showed positive selection across all populations (A to H), while the rest showcased positive selection only in abundant populations (B, C, D, E, and G). Nevertheless, differences in selection patterns for specific genes were observed between the two time-series. On the one hand, at BBMO, the gene encoding for the GH16 was positively selected only in 6 samples taken in 2011 and 2017 in populations D and E, while the gene for the cytochrome c-type biogenesis protein showed an overall selection across samples encompassing populations B, C, D, E, and F (**Figure 3.5A**). On the other hand, in SOLA, GH16 displayed broader patterns of positive selection across samples taken in 2009, 2010, 2011, and 2015, while the cytochrome c-type biogenesis protein only showed positive selection in a single sample from November 2015 (**Figure 3.5B**). Regarding positive selection at a long-spatial scale, 5 genes displayed selection, including the GH16 gene, but only in Mediterranean samples (data not shown). Overall, these results suggest that natural selection has acted upon a specific set of genes over large periods of times (12 and 7 years) and in the global ocean. Natural selection appears to act even at small spatial scales, as sampling sites for BBMO and SOLA are ca. 130 km apart.

3.3.5 Seasonality and biogeography

Seasonal abundance patterns were detected across the 169 studied MAGs. Based on the Lomb-Scargle periodogram algorithm, we detected significant periodic signals of abundance over time in both temporal series, BBMO, and SOLA. MAGs' periodicities were categorized into four groups based on our results and according to their abundances: a) *Annual*, for periodicity every 12 months; b) *Biannual*, when such periodicity exists between 6 and 12 months; c) *No pattern*, when a significant signal of periodicity was detected without a clear period associated; and d) *Not significant*, for MAGs with no significant signal of periodicity. In this context, annual periodicity implies similar abundance patterns every season (*e.g.,* comparable RPKG in January of 2009, 2010, etc.) and biannual periodicity implies abundance repeating every 6 months (*e.g.,* in January and July of the same year, and again in next January). Annual periodicity was strongly linked to seasonal MAGs and biannual periodicity to MAGs

that appear throughout the year in different seasons. Out of 169 MAGs, 120 in BBMO (71%) and 139 in SOLA (82.2%) showed annual abundance patterns, while 24 in BBMO (14.8%) and 4 in SOLA (2.4%) exhibited a biannual behavior (**Annex C Table 7**). For the other groups, 5 (3%) and 19 (11.2%) MAGs in BBMO and 3 (1.8%) and 23 (13.6%) MAGs in SOLA were associated with the No pattern and Not significant categories, respectively. No periodicity was detected in between samples of less than 6 months apart (**Annex C Table 7**).

We analyzed the distribution patterns of the 169 MAGs in the global ocean (**Annex C Table 7**). A total of 164 (97%) MAGs were found to be abundant in sub-tropical waters, from which 37 (21.9%) were also found in tropical waters and 13 (7.7%) in sub-polar and polar waters. Only 1 SAR116 MAG (G2.34) was exclusive of tropical waters and 4 *Flavobacteriales* MAGs (G1.528, G1.627, G1.675, and G2.178) were exclusive of sub-polar and polar waters. Thus, despite all MAGs being recovered from a single location in the Northwestern Mediterranean Sea (BBMO), some of them displayed distribution preferences in the global ocean featuring environmental conditions substantially different from BBMO.

3.4. DISCUSSION

Understanding the genetic variation of populations and how that variation is structured over space and time is fundamental for comprehending the ocean microbiome and its adaptations. Yet, for most microbial species, this information is lacking. Our work represents a step forward toward the understanding of the genetic diversity and structure of marine prokaryotic populations as well as the identification of the genomic basis of population adaptation. Our analysis of population-level genomic differentiation of 495 Mediterranean genomes in two neighboring coastal time-series in the Mediterranean Sea, encompassing 12 and 7 years of monthly samples, and the global ocean, revealed contrasting trends of long-term population dynamics as well as population structure in the global ocean.

Genetic differentiation (FST values) was similar for the 495 MAGs in both time-series encompassing 12 (BBMO) and 7 (SOLA) years, suggesting that these locations in the Northwestern Mediterranean Sea, separated by 130 km, share populations that

follow similar seasonal dynamics. Most of the 495 MAGs showed significant seasonal abundance patterns where the populations reappeared every year during the same periods. Microbial communities in both BBMO and SOLA, investigated with metabarcoding of the rRNA gene pointed to strong seasonal trends (282,283,302–304)[40,41,68–70], yet whether or not both sites were inhabited by the same populations had never been tested. The main reason is that the boundaries of microbial populations are not known for most species. Our findings suggest that microbial populations may occupy ocean areas, or patches, of at least 10,000 km^2. This is coherent with previous studies indicating that patch sizes in the ocean, including relatively homogeneous microbial communities, have sizes ranging between a few to tens of kilometers (22,305). Nevertheless, there were some differences in FST between BBMO and SOLA that may be due to some seasonal differences in environmental conditions, provoked by the specific climatological and geographical context of each location, *i.e.,* SOLA suffers from occasional winter storms that bring nutrients from sediments to the water column and freshwater inputs from flooding nearby rivers (281,284). In addition, the distinct number of samples between datasets encompassing different periods (BBMO – January 2009 – December 2020, SOLA – January 2009 – December 2015) may influence FST computation.

We found that geographic population differentiation (global ocean scale) was larger than temporal differentiation in both time-series (scale of 12 and 7 years), probably reflecting the higher variability of surface-ocean environmental conditions at a large geographic scale compared to the seasonal variations at the two time-series in the Mediterranean Sea. Furthermore, in agreement with the idea that microbial populations inhabit oceanic patches that may range up to tens or a few hundred kilometers[71,72], we found a higher population differentiation at the scale of hundred kilometers within the Mediterranean Sea (TARA stations 7, 9, 18, 23 and 25) than in both time-series (**Annex C Table 5**). Overall, our results suggest that microbial population differentiation is stronger at large spatial scales (high and very high genomic differentiation) than at long temporal scales in the surface ocean (little and moderate genomic differentiation). Previous population genomics studies of bacterioplankton communities reported comparable patterns of genomic differentiation at smaller spatiotemporal scales in the Baltic Sea (spanning 1,700 km transect and 2 years of samples from the Linneaus Microbial Observatory) (89). Diverse differentiated populations of the ubiquitous

Prochlorococcus were linked to changes in temperature and nutrient availability across the global ocean at different depths (276). Additionally, vertical and regional diversity within populations of *Synechococcus* suggested adaptation to specific depths in the Atlantic and Pacific Oceans, mirroring those of *Prochlorococcus* (277).

Altogether, among the 169 MAGs well represented in the two time-series and the global ocean, we observed three main patterns based on how strong (Mean FST > 0.15) or weak (Mean FST < 0.15) (79) population differentiation was across datasets: (i) strong genomic differentiation in both temporal and spatial scales, (ii) weak genomic differentiation in both, or (iii) weak genomic differentiation in the temporal scale, but strong in the spatial. In general, for MAGs originating from the Mediterranean Sea, pattern (iii) was the most common case, while (i) and (ii) were specific to some genomes. We did not detect any cases of strong temporal and weak spatial differentiation in any of the analyzed genomes, suggesting that populations adapted to the heterogeneous environmental conditions of the global ocean do not require further adaptation to the seasonal environmental changes occurring over large periods. The studied archaeal genomes (23 MAGs) are a good representation of such general pattern (iii), with only a few (4 MAGs) deviating from it.

Here, we analyzed archaeal seasonality for both time-series during 12 and 7 years and their global ocean biogeography. The abundant Nitrososphaerales G3.122 is an example of an archaeal genome showing weak population differentiation over 12 and 7 years, appearing almost exclusively during the cold season (winter and spring, population A), but being strongly differentiated across the global ocean (**Figure 3.3A**). We observed significant underlying population structure linked to temperature in BBMO and the global ocean, but not in SOLA, probably due to the temperature range in SOLA [8.53 – 24.32 ºC] being colder, where Nitrososphaearales appears to be better adapted to, than in BBMO [12.16 – 26.72 ºC] (**Figure 3.3A**). Additionally, our data also pointed to salinity as a significant driver of population structure over large-spatial scales. These findings agree with previous studies that point to environmental factors such as salinity, light, temperature, ammonium, oxygen and sulfide as major drivers of ammonia-oxidizing archaea distribution (47,306). More examples of MAGs following the general trends of weak population divergence in the temporal scale and strong in the spatial were the analyzed SAR324 (4 MAGs). Previous population genomics studies in the time-series from the ALOHA station (North Atlantic) include SAR324 genomes.

They defined at least four ecotypes with specific depth and seasonal distribution across the year (307). Although SAR324 showed seasonal annual abundance patterns in BBMO and SOLA too, population differentiation at the temporal scale was weak instead. Overall, considering the results from SAR324 and the other analyzed MAGs with low genomic differentiation on the temporal scale but strong on the spatial, it appears that marine microbial populations are well defined between distant locations in the surface global ocean, but stabilized over large periods, with little variations between seasons.

In our case, although the majority of the MAGs (66%) displayed weak population differentiation over 12 and 7 years of temporal metagenomic data, many (33%) showed highly-differentiated populations in both spatial and temporal scales. Within our Mediterranean MAG collection, such patterns were represented by 34 Flavobacteriales genomes. In a previous work, amplicon 16S rRNA data revealed that the Flavobacteriales group is an ensemble of organisms able to inhabit warm oligotrophic waters, and cold and nutrient-rich water masses in the North Atlantic Ocean, with clear differences between populations (308). In the same line, our population genomics analyses showed contrasting distribution patterns of abundance and strong genomic differentiation during cold and warm seasons in BBMO and SOLA, as well as for subtropical, tropical and subpolar waters in the global ocean. A particular example is genome G4.480, which displayed a strong population structure driven by temperature in the global ocean (**Figure 3.4A**). In turn, salinity explained less than 5% of the variance of the population structure in the global ocean (22 – 40 PSU), which contrasts with salinity driving Flavobacteriales populations in the Baltic Sea (89), a region featuring wider salinity gradients (2 – 30 PSU). These 34 Flavobacteriales MAGs, along with 1 Actinomarinales, 1 Balneolales, 2 SAR116 and 7 SAR86 genomes from our MAG collection, suggest that highly-differentiated populations defined between distant locations in the global ocean can also change following seasonal patterns over large periods of times.

We also observed MAGs that displayed weak population differentiation (Mean FST < 0.15) in both time-series and the global ocean (8 out of the 169 analyzed MAGs, including 1 Archaea, 1 SAR324, and 6 SAR11). The other 3 SAR11 genomes (9 in total) showing strong differentiation had a mean FST in the global ocean barely over the threshold (0.15, 0.16, and 0.18). One possible explanation is that the large population

sizes of SAR11 in the ocean likely contribute to high dispersal rates that, coupled with high levels of intra- and inter-species recombination (279), tend to homogenize SAR11 populations, limiting their overall divergence. In comparison, Actinomarinales MAGs, which overlap with SAR11 in terms of habitat and distributions (45,309), displayed higher genomic differentiation, specifically at the spatial scale. The smaller population sizes of Actinomarinales may limit dispersal rates, promoting population differentiation. Still, even though population differentiation in SAR11 was limited, we observed that underlying population structure was correlated with temperature and salinity in both long time-series and the global ocean, which agrees with previous works where significant divergence of SAR11 populations was driven by global oceanic-current temperatures (88), suggesting that even for well-established and ubiquitous organisms with low genomic differentiation across the global ocean, the environmental selection is still relevant even at such minute level of diversity.

To understand whether the observed population differentiation is the result of adaptive processes and environmental selection, we analyzed the potential action of positive selection in the 169 genomes through the ratio of non-synonymous to synonymous mutations (pN/pS) for all genes over 12 and 7 years of temporal data and in the global ocean. In general, we did not observe any correlation between the proportions of positively selected genes and the amount of population differentiation. Specifically, proportions in BBMO and SOLA were higher compared to the global ocean, where population differentiation is overall stronger. It implies that the amount of selective pressure over longer time scales in one single location (up to 12 years) is higher than in smaller time windows (< 4 years) despite encompassing samples ranging thousands of kilometers. Moreover, a few taxonomic groups showed a tendency towards a larger proportion of positively selected genes across all datasets, such as Flavobacteriales, SAR116, and SAR86, compared to others that displayed smaller proportions, like SAR11. Nevertheless, this could be due to the overrepresentation of the former groups in our MAG collection, where a wide diversity of organisms within the same taxonomical order and different evolutionary histories might be present. Thus, our results support that the adaptive processes driving population differentiation are organism-specific and depend on the environmental context of each microorganism.

In terms of which specific genes were selected in each dataset, we observed a variable amount of them always present in the three datasets, specifically in the two

Mediterranean time-series, pointing to natural selection acting upon specific sets of genes. However, the patterns of selection within such set of positively selected genes were variable across datasets. We showed for example that in the SAR86 genome G1.297, which belongs to the second most common group of heterotrophic bacteria in the global ocean (310), a gene encoding for GH16, an enzyme that breaks the glycosidic bonds in various glucans and galactans (311), was widely positively selected in SOLA, but restricted to certain BBMO samples. Likewise, the Cytochrome c-type biogenesis protein, an enzyme involved in cellular energy transduction processes, biosynthesis of cofactors, lipidic signaling molecules and binding of gases, among other cellular processes (312), was only widely positively selected in BBMO. These contrasting patterns of selection could be a reflection of such genes having low abundance in specific datasets, by either being lost (biological reason) or not being recovered during read recruitment (technical issue). Nevertheless, our results suggest the existence of different selective pressures upon the same genes over periods of 12 and 7 years acting at a small distance (~130 km). Still, considering that SOLA experiences stronger and more frequent winter perturbations (281,284), we can hypothesize that the changes in environmental conditions (*i.e.,* nutrient availability) might be promoting the selection of genes related to substrate degradation and energy transduction in specific environments.

In conclusion, through the analyses of 495 MAGs, we showcased a high local genomic diversity within microbial species over up to 12 years, as well as spatially over the global ocean. In general, prokaryotic populations were highly differentiated across thousands of kilometers, but whether these populations remained genomically stable over the years or adapted to the environmental changes happening along season is specific to each microorganism. In particular, SAR11 genomes appeared as a very particular case, where low-differentiated populations were defined across the global ocean, likely as a consequence of its high dispersal capability and high intra- and interpopulation recombination. Moreover, our study indicates that environmentally driven population structure is not limited to specific species, but rather appears to be a general pattern for all prokaryotes, independently of how weak or strong the population differentiation is. Locally, temperature was a main driver structuring population differentiation over 12 and 7 years of temporal data at the Northwestern Mediterranean Sea, while globally temperature and salinity together shaped population structure.

Overall, our work indicates that metagenomic data is a powerful tool to determine and analyze microbial population structure, and therefore, explore the ecology and evolution of abundant key marine microorganisms. The methodology proposed here has the potential to be used to assess the niche adaptation and evolutionary history of other microbial species. It is a crucial topic to understand how microbes adapt to new and changing environments over long periods and large geographic distances, which is particularly relevant in the context of climate change.

CHAPTER 4

Investigating the marine protist interactome using Single-Cell Genomics

Francisco Latorre[a], Ina M. Deutschmann[a], Anders K. Krabberød[b], David López-Escardo[a], Michael E. Sieraki[c], Ramunas Stepanauskas[d], Olivier Jaillon[e,f], Ramon Massana[a], Ramiro Logares[a]

[a] Institute of Marine Sciences (ICM), CSIC, Barcelona, E-08003, Catalonia, Spain.

[b] Department of Biosciences, Section for Genetics and Evolutionary Biology, University of Oslo, 0316, Oslo, Norway

[c] National Science Foundation, 2415 Eisenhower Ave., Alexandria, VA 22314, U.S.A.

[d] Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States.

[e] Metabolic Genomics, Genoscope, Institut de Biologie François Jacob, CEA, CNRS, Univ Evry, Université Paris Saclay, 91000 Evry, France.

[f] Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022 / Tara Oceans GOSEE, 3 rue Michel-Ange, 75016 Paris, France.

4.1. INTRODUCTION

The ca. 10^{12} microbial species inhabiting planet Earth (313) have crucial roles in global biochemical cycles and food webs (11). In particular, marine microorganisms allow the recycling of nutrients and energy, and their flow from lower to upper trophic levels in the global ocean (127), which is a large integrated ecosystem that regulates and maintains the biosphere. The marine microbial food web is composed of species that are constantly interacting (via mutualism, competition, commensalism, amensalism, parasitism, and predation) (314). Furthermore, prokaryotes can acquire dissolved organic matter from the environment and incorporate it into their biomass, thus passing it to higher trophic levels via predation, also known as the microbial loop (17,105). Parasites and symbionts are known to increase the complexity and diversity of food webs, as they alter the number of species coexisting and the ecosystem structure (103,315). These ecological interactions among microbes underpin ocean ecosystem function. Yet, most microbial interactions remain unknown, representing one of the most extensive knowledge gaps in marine biology.

Global change may affect microbial interactions leading to unpredictable consequences (*e.g.,* changes in microbial communities that could impact ecosystem services or increase the spread of pathogens or toxic organisms). Thus, a primordial goal is to understand the marine microorganisms' role in the ocean, including molecular and ecological interactions (103). For this, two types of data are required: (i) a list of all species or populations, and (ii) their interactions in a spatiotemporal context (316). The implementation of High-Throughput Sequencing (HTS) in microbial ecology contributed to the identification of marine microbes, their diversity, and spatiotemporal distributions (52,54,122,131,132,201,262). Still,–marine microbial diversity remains poorly characterized at the genomic level, because most microbes are unculturable, leading to a lack of accessible genomes (269). As mentioned, very little is known about microbial interactions (103,106).

An interaction can be beneficial (positive), detrimental (negative), or neutral for one or both of the interacting organisms. Thus, a variety of dynamics can be established between two interacting organisms, including loss-loss (*e.g.,* competition), neutral-loss (*e.g.,* amensalism), win-loss (*e.g.,* predation and parasitism), win-neutral (*e.g.,* commensalism) and win-win (*e.g.,* mutualism) relationships (103,104,314). These

interactions may require physical contact with others (*e.g.,* symbiosis) or not (*e.g.,* amensalism). In the latter case, the release of metabolites by one species may affect another triggering a particular cell response (*e.g.,* cell growth) (317,318). The majority of previous studies on marine microbial interactions have targeted specific relationships, such as competition (319), parasitism (320), or predation (321); yet the understanding of the overall microbial interactome in the global ocean remains limited (103).

Association networks based on microbial abundance data have become a key approach for studying complex interactions in natural systems (104,106,109). However, networks still require the characterization of most microbial species or populations, a daunting task for most natural ecosystems (54). A widely used solution to overcome this limitation is the use of metabarcoding and operational taxonomic units (OTUs), which allows for the characterization of most lineages present in a microbial community (54). Then, association networks can be constructed based on OTU abundances. In a typical association network, nodes (OTUs) are interconnected through edges, representing an association or potential interaction between organisms, which can be either positive or negative (314). Several studies have constructed marine microbial association networks. For example, ten years of monthly data from the Blanes Bay Microbial Observatory (BBMO) in the Northwestern Mediterranean Sea showcased several seasonal associations between Alphaproteobacteria OTUs and major protistan groups, such as Dinoflagellates, Diatoms, Cryptophytes, Mamiellophyceae, and Syndiniales (109). Large-scale spatial analyses from the *Tara oceans* expedition predicted more than 80,000 potential interactions in the global ocean based on OTU abundances (110), the most common being between the parasitic Syndiniales and Dinoflagellates. The previous study also reported associations between Flavobacteria and diatoms, and between Dinoflagellates and Rhodobacterales, which were already known from cultures (322,323). Interestingly, among the predicted associations in the *Tara Oceans* study were previously confirmed interactions by Single-Cell Genomic (SCG) techniques (111).

High-throughput SCG methodologies first isolate individual cells using Fluorescence Activated Cell Sorting (FACS), and proceed with cell lysis, genome amplification, and sequencing, resulting in a Single Amplified Genome (SAG) (59). However, if two individual cells are physically interacting in a symbiotic, parasitic, or predatory relationship, or are simply attached, they may be isolated together leading to

both genomes being sequenced. For instance, previous studies found within the same SAG the eukaryotic flagellate MAST-4 along the planktonic bacterium SAR11 (111), the protist picobiliphyte with bacteria (324), and viruses within their hosts (112,113,324). Thus, SAGs help infer physical interactions between marine microorganisms.

Here, we inferred potential microbial interactions using 3,015 eukaryotic SAGs, one of the biggest collections of eukaryotic SAGs to date. These SAGs were isolated from distinct marine locations, including cells from the *Tara Oceans* expedition (stations in the Mediterranean and the Indian Ocean), the Blanes Bay Microbial Observatory in the Northwestern Mediterranean Sea, and the Gulf of Maine in the North Atlantic Ocean. We aim at answering the following questions: How many interactions can we detect and which ones are the most common? Who are the organisms interacting? Can we establish which kind of relationship they hold?

4.2. METHODS

4.2.1 Sample collection and Low Coverage Sequencing

To acquire single cells, water samples from the Gulf of Maine (GoM) and the Blanes Bay Microbial Observatory (BBMO) were collected. The GoM sample was collected in Boothbay Harbor, Maine, United States (43◦50′39.87″N 69◦38′27.49″W) at one meter depth on the 19th of July, 2009 (**Annex D Figure 1**). The sorting of plastidic (phototrophic) cells was done based on their chlorophyll autofluorescence, and the sorting for aplastidic (heterotrophic) cells was done with LysoTracker Green DND-26 (75 nmol L−1; Invitrogen, Carlsbad, CA, United States) as described in (Brown *et al.,* 2020) (113) at the Single Cell Genomics Center, Maine, United States. Two water samples from BBMO at the North Western Mediterranean Sea (41◦40′N, 2◦48′E; http://bbmo.icm.csic.es/) (100) were collected in the winter and summer of 2016 (19th of January and 5th July) at 1 m depth and ~1 km offshore (**Annex D Figure 1**). Water samples were pre-filtered *in situ* with a 200 µm nylon-mesh and transported to the laboratory, where they were treated with 6% glycine betaine (SigmaAldrich), frozen in liquid nitrogen, and stored at –80ºC. Plastidic eukaryotic cells were sorted based on chlorophyll autofluorescence from cryopreserved winter BBMO samples as described above. Aplastidic eukaryotic cells were sorted from cryopreserved winter and summer BBMO samples with an SYBR Green DNA stain (42,81). Cellular DNA was obtained

after cell lysis with KOH and amplified with either multiple displacement amplification (MDA) using phi29 polymerase (Thermo Fisher) or WGA-X using phi29mut8 (Thermo Fisher) as described in (Brown *et al.,* 2020) (113).

Next, low Coverage Sequencing (LoCoS), a cost-effective approach to obtain limited genomic data from a maximal number of individual cells, was performed on both phototrophic and heterotrophic eukaryotic cells from GoM and BBMO as described in (Stepanauskas *et al.,* 2017) (120). The obtained Single-Amplified Genomes (SAGs) were grouped into six datasets: one dataset for GoM SAGs for both phototrophic and heterotrophic protists (GoM dataset, 912 SAGs), and five datasets for BBMO; two BBMO datasets were sequencing replicates for winter phototrophic cells (WA170123 and WA170125, 378 SAGs each), two BBMO datasets were pseudo-replicates of the same summer heterotrophic cells under distinct sequence coverage conditions (SH171117 and SHp170809, 382 and 307 SAGs) and one final dataset for heterotrophic winter cells (WH180222, 372 SAGs) (**Annex D Table 1**)

4.2.2 Sample collection and deep SAG sequencing

Planktonic water samples were collected during the circumglobal *Tara Oceans* expedition (**Annex D Figure 1)** and cryopreserved as described in (Heywood *et al.*, 2011) (60). Individual cells from the picoplankton fraction (0-8 – 5 µm) were isolated and stained with SYBR Green stain using a MoFlo flow cytometer as described elsewhere (58,61,81). A total of 205 SAGs were obtained from individual cells using a phi29 polymerase-based MDA reaction (111,151). All single-cell work was carried out by the Single Cell Genomic Center. Sequencing was done in 1/8 of a lane using either *Illumina* HiSeq2000 or HiSeq4000 at either Oregon Health & Science University (USA) or the French National Sequencing Center (Genoscope, France).

Additional single-cell samples for deep sequencing were collected at the BBMO on May 8th 2018 using the same protocols as described above, including single-cell sorting of pigmented and unpigmented small protists, and whole genome amplification by MDA. KAPA or NextEra preparation kits were used for 81 BBMO cells in different Illumina platforms and sequencing services as described in (Labarre *et al.,* 2021) (58). The resulting SAGs were taxonomically screened by PCR amplification and Sanger

sequencing of the 18S rRNA gene using universal eukaryotic primers (**Annex D Table 2**).

4.2.3 Assembly, gene prediction, and taxonomical assignation

All SAGs were individually assembled into contigs using SPAdes 3.13.0 in single-cell mode (--sc) and default parameters (152). Contigs of > 3,000 base pairs [bp] were taxonomically classified as eukaryotic, bacterial, archaeal, plastidic, and unknown origin with Tiara 1.0.2 (325) in default mode. Only eukaryotic and bacterial contigs were used in downstream analyses. Prokaryotic genes were predicted in bacterial contigs with prodigal v2.6.3 (326) in default mode and taxonomically assigned with the GTDB gene database r89 (327,328) using MMseqs2 version 15c77624453c757c15790b9c3511212caec870b0 (329). Eukaryotic exons were predicted on eukaryotic contigs using the *predictexons* function from Metaeuk version ea903e554a71285b95da54029fe288d7b7867bba (330) with the following parameters: --metaeuk-eval 0.0001, --min-length 40, --slice-search, --min-ungapped-score 35, --min-exon-aa 20 and --metaeuk-tcov 0.6. Moreover, redundancy of the predicted exons was reduced with *reduceredundancy* function in default mode. Metaeuk was run using a custom database combining MERC (proteins assembled from eukaryotic Tara Oceans metatranscriptomic datasets (331)), uniclust90 seed proteins (332) and MMETSP proteins (333,334). Lastly, eukaryotic exons were taxonomically annotated using MMseqs2 against the EukProt database (335) in default mode.

4.2.4 Interaction prediction and network construction

To predict interactions from SAGs, cells sequenced from the LoCoS datasets (GoM, WA170123, WA170125, SHp170809, SH171117, and WH180222) were assigned eukaryotic taxonomy based on the predicted exons using the Last Common Ancestor (LCA) approach implemented in MMseqs2. Only exons with an LCA at the level of "genus" were considered. SAGs were assigned to the "genus" with the highest number of annotated exons (minimum of 10% of total exons within the SAG). In turn, SAGs with high sequencing depth from the BBMO and TARA datasets were assigned a taxonomy based on the 18S rRNA gene, which was already available.

Within each SAG, we tested for the presence of other microorganisms. Specifically, we tested for prokaryotes (some archaea genes were detected despite

filtering) by assigning taxonomy from bacterial genes, which was cut at the taxonomical "order" level for simplicity. Additionally, in deep sequenced TARA and BBMO SAGs, we tested for eukaryotes by assigning taxonomy from 18S rRNA genes and exons, which were also cut at the taxonomic "order" level. Exons with an assigned taxonomical level matching the 18S rRNA were manually filtered out to avoid noise.

Next, we predicted potential eukaryote-prokaryote and eukaryote-eukaryote interactions. We related microbes to each other when at least one prokaryotic/eukaryotic gene was found within a SAG (A total of 3,035 and 916 interactions for euk-prok and euk-euk, respectively). We required strong evidence for potential interactions by only considering microbes that occurred in two SAGs within the same dataset (A total of 698 and 484 for euk-prok and euk-euk). We visualized strong potential interactions with Gephi 0.9.2 (336).

4.3. RESULTS

4.3.1 Eukaryote – prokaryote interactions from Low Coverage Sequencing SAGs

LoCoS was performed on environmental SAGs from BBMO and GoM water samples. BBMO SAGs were divided into five sets based on the time of the year they were isolated (winter and summer) and their tropism (photo and heterotrophy). MDA and sequencing replicates were performed for winter phototrophic cells (WA170123 and WA170125) under the same methodological conditions (382 obtained SAGs each). Additionally, SAGs were produced for heterotrophic cells from the summer sample (SH171117 and SHp170809) under different sequencing coverage conditions, resulting in a distinct number of SAGs (382 and 307, respectively). Overall, the number of potential interactions based on the assigned taxonomy of the predicted genes was similar between replicates (**Table 4.1**).

On the one hand, datasets WA170123 and WA170125 recovered 34 and 42 interactions respectively, from which 8 and 11 showed stronger evidence (*i.e.,* they were found in at least two SAGs) between Pelagophyceae and Chlorophyta with mainly Alphaproteobacteria, Bacteroidota, and Planctomycetota (**Figure 4.1A**). In particular, potential interactions between the eukaryotic *Micromonas* with Pelagibacterales, Flavobacteriales, Caulobacterales, and Nisaeales were observed in both replicates, while interactions with bacterial Burkholderiales and archaeal Pacearchaeales were only found

in WA170125. Similarly, the potential interactions between the eukaryotic *Pelagomonas* and Pelagibacterales, Flavobacteriales, Caulobacterales, and Planctomycetales were observed in both replicates (**Annex D Tables 3 and 4)**.

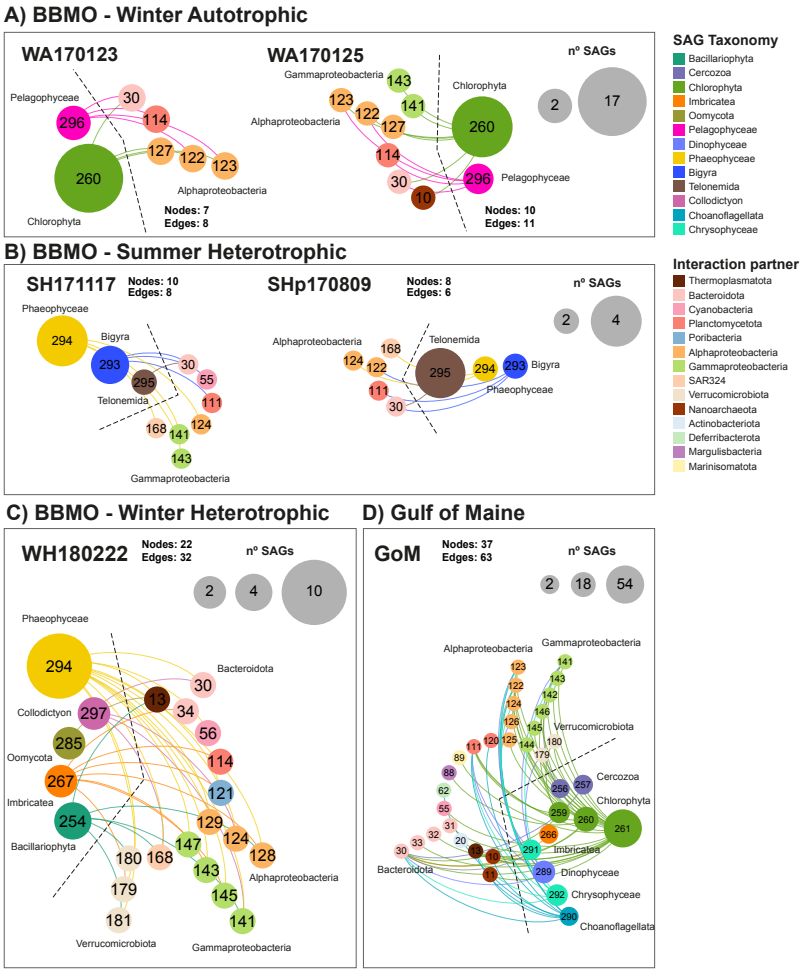

Figure 4.1. Networks of potential eukaryote – prokaryote interactions from LoCoS SAGs for **A)** BBMO winter phototrophic (plastidic) cells; **B)** BBMO summer heterotrophic (aplastidic) cells; **C)** BBMO winter heterotrophic cells; and **D)** GoM both heterotrophic and phototrophic cells. Prokaryotes are connected to a eukaryote if they appear in at least 2 SAGs. Edge color coincides with the taxonomy of the eukaryote, *i.e.,* the main taxonomic assignment of the SAG. Eukaryotic nodes are separated from the prokaryotic nodes by dashed lines and their sizes represent the number of SAGs in the dataset. Nodes are grouped to different axes by taxonomical class, and ordered along the axis based on the number of potential interactions, *i.e.,* the most connected nodes within a class are further away from the center of the circle. The node labels (numbers within the nodes) link to the complete taxonomy of the microbes, shown in **Annex D Table 11**. The number of nodes and edges is provided.

Table 4.1. The number of potential eukaryote-prokaryote interactions across all datasets. Values only consider strong potential interactions, that is, those that appear in at least 2 SAGs.

SAG Taxonomy	nº occurrences	%
TARA		
Chrysophyte-G	36	7.9
Chrysophyte-H	56	12.3
Dictyochophyceae	38	8.3
MAST-11	14	3.1
MAST-1D	10	2.2
MAST-3A	9	2.0

MAST-3F	37	8.1
MAST-4A	115	25.2
MAST-4B	47	10.3
MAST-4C	34	7.4
MAST-4E	44	9.6
MAST-7	17	3.7
TOTAL	*457*	*100.0*
BBMO		
Chlorarachniophyta-sp1	1	0.9
ChrysophyceaeG-sp2	1	0.9
MAST-1C-sp1	13	11.5
MAST-1D-sp2	82	72.6
MAST-4A-sp1	9	8.0
MAST-8B-sp1	1	0.9
Micromonas-sp1	1	0.9
Picozoa-sp1	3	2.7
Prymnesiophyceae-sp1	2	1.8
TOTAL	*113*	*100.0*
LoCoS - GoM		
Abollifer	1	1.6
Amoebophrya	6	9.5
Bathycoccus	4	6.3
Didymoeca	9	14.3
Lotharella	1	1.6
Mataza	1	1.6
Micromonas	8	12.7
Ostreococcus	23	36.5
Paraphysomonas	3	4.8
Unknown-Chrysophyceae	7	11.1
TOTAL	*63*	*100.0*
LoCoS BBMO - WH180222		
Collodictyon	4	12.5
Euglypha	7	21.9
Nemacystus	14	43.8
Pythium	1	3.1
Thalassiosira	6	18.8
TOTAL	*32*	*100.0*
LoCoS BBMO - SH171117		
Incisomonas	3	37.5
Nemacystus	4	50
Telonema	1	12.5
TOTAL	*8*	*100.0*
LoCoS BBMO - SHp170809		
Incisomonas	3	42.9
Nemacystus	2	28.6
Telonema	1	14.3
TOTAL	*7*	*100.0*
LoCoS BBMO - WA170123		
Micromonas	4	50
Pelagomonas	4	50
TOTAL	*8*	*100.0*
LoCoS BBMO - WA170125		
Micromonas	7	63.6
Pelagomonas	4	36.4
TOTAL	*11*	*100.0*

On the other hand, pseudo-replicates SH171117 and SHp170809 showed 61 and 99 total interactions, from which 8 and 6 had strong evidence, including potential interactions found in both datasets between Bigyra, Telonemida, and Phaeophyceae with Alphaproteobacteria, SAR324 and Bacteroidota (**Figure 4.1B**), such as *Telonema* with Flavobacteriales; *Incisomonas* (MAST-3) with Flavobacteriales and

Phycisphaerales. A strong potential interaction between *Incisomonas* and Synechococcales was exclusive of SH171117 (**Figure 4.1B, Annex D Tables 5 and 6**).

A total of 372 heterotrophic SAGs from winter collected at BBMO were sequenced (WH180222). From a total of 287 potential interactions, 32 strong potential interactions were detected featuring Phaeophyceae, Collodictyon, Oomycota, Imbricatea, and Bacillariophyta with Alphaproteobacteria, Gammaproteobacteria, Bacteroidota, and Verrucomicrobiota (**Figure 4.1C**). The most relevant potential interactions were between *Thalassiosira* (18.8% of the total interactions) and SAR324 (found four times), and *Euglypha* (Cercozoa; 21.9%) with Planctomycetales (three times) (**Table 4.1**, **Annex D Table 7**).

From GoM, 912 SAGs were sequenced, 595 aplastidic and 317 plastidic. A total of 179 potential interactions were predicted, from which only 63 were considered strong potential interactions. The most common potential interactions included Cercozoa, Chlorophyta, Imbricatea, Chrysophyceae, Dinophyceae, and Choanoflagellata with Alphaproteobacteria, Gammaproteobacteria, Bacteroidota, and Verrucomicrobiota (**Figure 1D**). Specifically, *Ostreococcus* appeared as the genus with the greatest number of potential interactions (36.5%), followed by *Didymoeca* [Choanozoa] (14.3%), *Micromonas* (12.7%), and an unknown Chrysophyceae (11.1%), among others (**Table 1**). The most prominent potential interactions were between *Ostreococcus* and Phycisphaerales (26 times), Flavobacteriales (21 times) and the archaeal Poseidoniales (11); *Micromonas* with Flavobacteriales (8 times), Phycisphaerales (6 times), Pelagibacteriales (4 times), and Poseidoniales (4 times); and the unknown Chrysophyceae with Flavobacteriales (7 times) and Pelagibacterales (4 times) (**Annex D Table 8**).

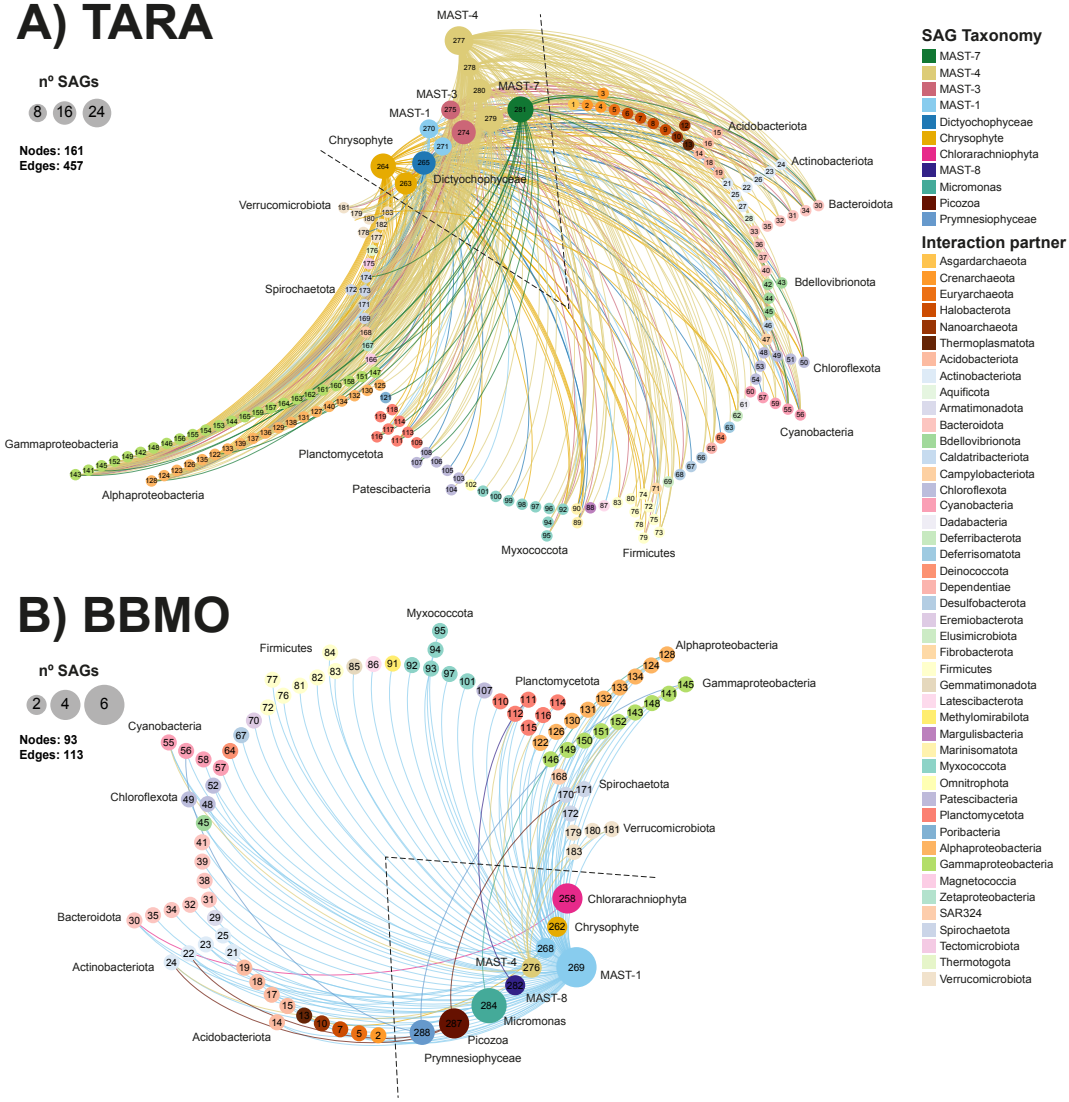

Figure 4.2. Networks of potential eukaryote – prokaryote interactions from deep-sequenced SAGs for **A)** Tara Oceans and **B)** BBMO. Prokaryotes are connected to a eukaryote if they appear in at least 2 SAGs. Edge color coincides with the taxonomy of the eukaryote, *i.e.,* the main taxonomic assignment of the SAG. Eukaryotic nodes are separated from the prokaryotic nodes by dashed lines and their sizes represent the number of SAGs in the dataset. Nodes are grouped to different axes by taxonomical class, and ordered along the axis based on the number of potential interactions, *i.e.,* the most connected nodes within a class are further away from the center of the circle. The node labels (numbers within the nodes) link to the complete taxonomy of the microbes, shown in **Annex D Table 11**. The number of nodes and edges is provided.

4.3.2 Eukaryote – prokaryote interactions from Deep Sequencing SAGs

A total of 205 and 81 SAGs obtained from the Tara Oceans expedition and the BBMO, respectively, were sequenced deeply. SAGs from TARA displayed 1,130 potential interactions, from which 457 were strong potential interactions. Among these, the stramenopile MAST-4A showed the largest number of potential interactions (25.2%), followed by Chrysophyte-H (12.3%) and MAST-4B (10.3%) (**Table 4.1**). Across all cells, potential eukaryotic interactions with Alphaproteobacteria, Gammaproteobacteria,

Acidobacteriota, Actinobacteriota, Bacteroidota, Chloroflexota, Cyanobacteria, Firmicutes, Myxococcota, Patescibacteria, Planctomycetota, Spirochaeota, and Verrucomicrobiota were common (**Figure 4.2A**). Specifically, potential eukaryotic interactions with Flavobacteriales, Synechococcales, Cyanobacteriales, and Pelagibacterales. MAST-4A, MAST-4E, MAST-11, Chrysophyte-H, and Chrysophyte-G potentially interacted with SAR86. Also, MAST-4A, MAST-4B, MAST-3F, and Chrysophyte-G potentially interacted with SAR324 (**Figure 4.2A**; **Annex D Table 9**).

Table 4.2. The number of potential eukaryote-eukaryote interactions for deep sequencing SAGs. Values only consider strong potential interactions, that is, those that appeared in at least two SAGs with the same taxonomy.

SAG Taxonomy	nº occurrences	%
TARA		
Bathycoccus prasinos	6	2.0
Chrysophyte-G	39	12.8
Chrysophyte-H	15	4.9
Dictyochophyceae	25	8.2
MAST-11	32	10.5
MAST-1D	15	4.9
MAST-3A	27	8.9
MAST-3F	21	6.9
MAST-4A	23	7.6
MAST-4B	15	4.9
MAST-4C	23	7.6
MAST-4E	21	6.9
MAST-7	34	11.2
MAST-9	7	2.3
Pelagomonas calceolata	1	0.3
TOTAL	*304*	*100.0*
BBMO		
Chlorarachniophyta-sp1	6	3.4
ChrysophyceaeG-sp2	2	1.1
MAST-1C-sp1	6	3.4
MAST-1D-sp2	42	23.7
MAST-3C-sp1	11	6.2
MAST-3C-sp2	14	7.9
MAST-4A-sp1	2	1.1
MAST-8B-sp1	39	22.0
Micromonas-sp1	3	1.7
Picozoa-sp1	45	25.4
Prymnesiophyceae-sp1	7	4.0
TOTAL	*177*	100.0

Among BBMO SAGs, 1,203 potential interactions were predicted, from which 113 were strong potential interactions. The strong potential interactions were similar to those found in TARA SAGs, which included Alphaproteobacteria, Gammaproteobacteria, Acidobacteriota, Actinobacteriota, Bacteroidota, Chloroflexota, Cyanobacteria, Firmicutes, Myxococcota, Planctomycetota, Spirochaeota and Verrucomicrobiota (**Figure 4.2B**). MAST-1D dominated in the number of potential interactions (72.6%), followed by MAST-1C (11.5%) and MAST-4A (8%) (**Table 4.1**). Strong potential interactions of MAST-1D with Flavobacteriales, Synechococcales,

Cyanobacteriales, and SAR324 were observed, as well as two potential interactions between Picozoa and Actinobacteria (**Figure 4.2B**; **Annex D Table 10**).

4.3.3 Eukaryote –eukaryote potential interactions from Deep Sequencing SAGs

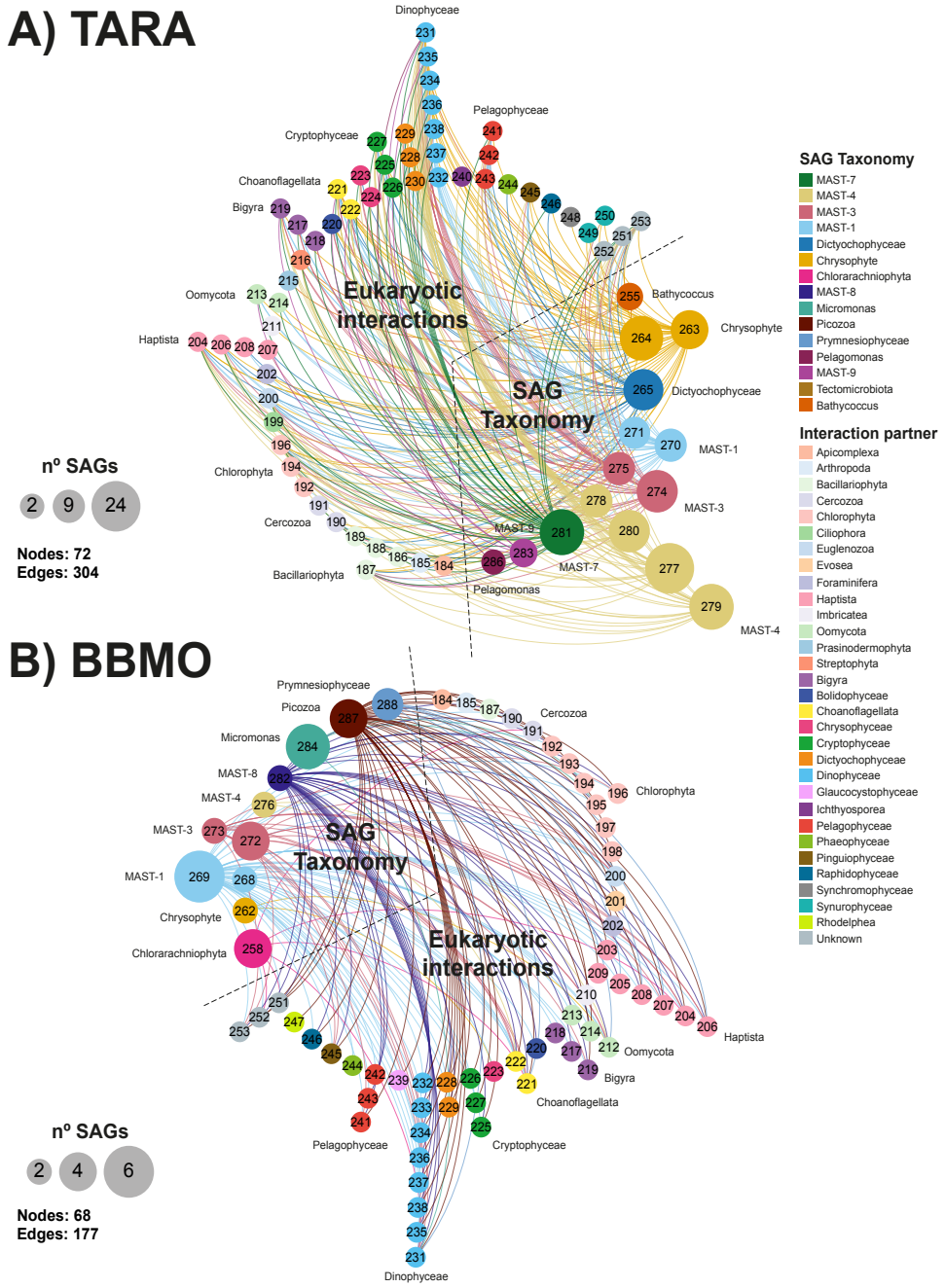

Figure 4.3. Networks of eukaryote – eukaryote interactions from deep sequenced SAGs for A) Tara Oceans and **B)** BBMO. Eukaryotes are connected to other eukaryotes if they appear in at least 2 SAGs. Edge color coincides with the taxonomy of the main taxonomic assignment of the SAG. SAG nodes are separated from the prokaryotic nodes by dashed lines and their sizes represent the number of SAGs in the dataset. Nodes are grouped to different axes by taxonomical class, and ordered along the axis based on the number of potential interactions, i.e., the most connected nodes within a class are further away from the center of the circle. The node labels (numbers within the nodes) link to the complete taxonomy of the microbes, shown in **Annex D Table 11**. The number of nodes and edges is provided.

A total of 505 and 411 potential eukaryote-eukaryote interactions were predicted for TARA and BBMO SAGs, from which 304 and 177 were considered strong potential interactions. In TARA, potential eukaryote-eukaryote interactions were equally distributed among all heterotrophic SAGs, with MAST-4 (species A, B, C and E together) being the one with the greatest number of them (27%), followed by Chrysophyte-G (12.8%), MAST-7 (11.2%), and MAST-11 (10.5%). In turn, MAST-9 (2.3%), *Bathycoccus prasinos* (2%), and *Pelagomonas calceolata* (0.9%) showed the lowest number of potential interactions (**Table 4.2**). Common potential interactions found between TARA eukaryotic SAGs involved Haptista, Bigyra, Dinophyceae, and Cryptophyceae (**Figure 4.3A, Annex D Table 9**). In BBMO, Picozoa, MAST-1D, and MAST-8B showed the largest number of potential interactions (25.4%, 23.7%, and 22.0%, respectively) (**Table 4.2**), with Haptista, Dinophyceae, Cryptophyceae, Dictyophyceae and Oomycota being the most numerous potential interactions overall (**Figure 4.3B**, **Annex D Table 10**).

4.4. DICUSSION

Microbial interactions play a crucial role in marine ecosystems, underpinning food webs and the microbial loop(17). Despite their importance, most microbial interactions in the ocean remain unknown (103,106). We provide one of the biggest collections of eukaryotic SAGs to date (3,015 SAGs) and determined potential physical interactions for an array of taxa, several of which appear not to have been reported before.

The LoCoS data from GoM and BBMO presented here has already been used to detect viral infection in protists (113). In the same study, a bacterial signal was recovered and inferred to be higher in BBMO than in GoM SAGs, which agrees with our results showing a larger number of potential interactions for winter aplastidic cells in BBMO alone (287 interactions) compared to the whole dataset of GoM, both plastidic and aplastidic cells (179 potential interactions). Previous work from marine expeditions, using association data, reported a substantial number of potential interactions in the Mediterranean Sea compared to other oceanic basins (110,337). It was suggested in a study from the *Tara oceans* expedition that this could be reflecting a bias in the number of samples (110). Yet, our LoCoS results indicated a higher proportion of eukaryote – prokaryote interactions in the sampled location in the Mediterranean Sea (BBMO) than in the North Atlantic (GoM), both datasets with

comparable numbers of SAGs (BBMO – 1,132 cells; GoM – 912 cells) and sequencing depth. Our results agree with a potentially larger number of microbial interactions in the Mediterranean Sea compared to other basins. The Mediterranean Sea is a diversity hotspot, and recent evidence point to endemic microbes (280), therefore it would not be surprising if it turns out that it contains more microbial interactions than other basins. These results agree with the idea that the number of interactions changes across ocean regions (110,337).

A decade ago, Martínez-García and colleagues (111) sequenced 315 protistan SAGs from GoM and observed a high number of potential predatory interactions between eukaryotic heterotrophs (aplastidic) and mixotrophs (plastidic) with Bacteroidota, Alphaproteobacteria and, especially, Gammaproteobacteria. Our analyses of 912 cells from GoM agree with those findings. However, we found new strong and common potential interactions between protists with Verrucomicrobiota and Planctomycetota bacteria (**Figure 4.1**). Whether Verrucomicrobiota and Planctomycetota signal within protistan SAGs is a product of predation or symbiosis is still unclear, as they have been found associated with higher eukaryotes (338,339), including marine animals such as the sea cucumber (340), the giant tiger prawn (341) or in sponges (342), but not other microorganisms. In particular, species of Verrucomicrobiota carry genes for the non-flagellar III secretion system (338), which is a protein known to mediate the interaction between eukaryotes and prokaryotes (343,344).

Among the potential interactions that we found in LoCoS SAGs are those between phototrophic eukaryotes with Flavobacteriales and Pelagibacterales. While cells of the order Flavobacteriales are often found associated with eukaryotic phytoplankton (345), either via grazing (mixotrophy) or attached (346), associations with the free-living *Pelagibacterales* seem so far uncommon among eukaryotic phytoplankton despite showing a global distribution and relatively high abundances. General patterns of potential interactions were similar in GoM and BBMO LoCoS, but specific strong potential interactions were different. For instance, GoM SAGs included Chrysophyceae, Cercozoa, and Choanoflagellata cells potentially interacting with Flavobacteriales or Pelagibacterales, most likely in a predator-prey relationship. Chrysophyceae species have been extensively documented as they can digest bacteria, such as Actinobacteria SAR324 and Bacteroidetes (Flavobacteriales) (111), Firmicutes

(347), Cyanobacteria (348), and several Gammaproteobacteria (349–351). Cercozoa have been found in a predator-prey relationship with unclassified bacteria (352), but also in an endosymbiotic association with Alphaproteobacteria (353). In some marine association networks (108), Choanoflagellata appears associated with Pelagibacterales and Actinobacteria, which agrees with our results.

BBMO LoCoS SAGs not only showed distinct potential interactions compared to GoM but also among BBMO SAGs obtained at different times of the year; *Telonema* and *Incisomonas* (MAST-3) showed strong potential interactions exclusively in summer (SH171117 and SHp170809); while Collodictyon, Imbcricatea, Oomycota, and Bacillariophyta SAGs showed strong potential interactions in winter (WH180222). Additionally, we found a larger number of potential interactions within SAGs collected in winter compared to those collected in summer. This agrees with association networks based on amplicon sequencing data from 10 years of monthly data from BBMO (109), which pointed to more potential interactions in cold than in warm waters, with season-specific key organisms. In the same study, core associations between protists and bacteria accounted for 31% (433) of the total predicted interactions (1,411), highlighting those between eukaryotic Bacillariophyta, Mamiellophyceae, and Pelagophyceae, with Alphaproteobacteria, Gammaproteobacteria, Bacteroidia, Verrucomicrobiota, and Acidobacteriota. Genes belonging to these bacterial taxonomical groups were found within LoCoS SAGs collected at BBMO, including potential interactions between Mamiellophyceae and Pelagophyceae with Alphaproteobacteria, Gammaproteobacteria, and Bacteroidota (**Figure 4.1A**); or Bacillariophyta with Alphaproteobacteria, Gammaproteobacteria, Verrucomicrobiota and SAR324 (**Figure 4.1C**). The above core associations determined with correlation analyses were observed in winter samples, corresponding to the same season in which our BBMO SAGs supporting such interactions were collected.

One limitation of Low Coverage Sequencing is that, despite allowing the sequencing of several SAGs in one run, the sequencing is done at the superficial level, missing potential interactions. As expected, we found more potential interactions in SAGs that were sequenced more deeply. For example, 287 potential eukaryote-prokaryote interactions were observed in 371 LoCoS BBMO SAGs (WH180222), compared to the 1,203 potential interactions retrieved from 81 deeply sequenced SAGs also from BBMO. General potential interaction patterns observed in LoCoS SAGs were

maintained when a deeper sequencing was applied, yet the deeper sequencing expanded the list of potential eukaryote - prokaryote interactions, including those with archaeal organisms that were barely present in LoCoS SAGs. Overall, deeply sequenced TARA and BBMO SAGs provided supporting evidence for potential interactions inferred in other studies using different techniques. For example, in global ocean (Chapter 1) (61) and temporal (109) association networks both positive and negative associations were predicted between MAST-4 and other bacteria, including Alphaproteobacteria (SAR11), Gammaproteobacteria (SAR86), Flavobacteriales, and Verrucomicrobiota. Signals for all these organisms were found within MAST-4 SAGs, along with *Synechococcus*, already demonstrated to be a prey (131). In the same line, potential predatory interactions for other unculturable heterotrophic flagellates (MAST-1, MAST-7, MAST-8, and Picozoa) and the same bacteria (SAR11, SAR86, Flavobacteriales, and *Synechococcus*) can be inferred from our SCG data.

Some association network studies pointed to *Syndiniales,* which is an obligate parasitic clade within the Dinoflagellates (354), as one of the main eukaryotic protagonists in marine interactions (109,354). Moreover, a high number of associations between MAST species and Syndiniales have been reported in the past (61,110). Here, we found signals of Syndiniales within MAST-1, MAST-7, and MAST-8 BBMO high sequenced SAGs, but also within Picozoa and Chrysophyceae, suggesting widespread parasitism in these groups, which could affect the complexity of food web dynamics (354). Within the group of interactions that were confirmed in previous studies and were also found in our SCG data, are the symbiotic relationship between Prymnesiophyceae and Foraminifera (355), *Micromonas* and Dinoflagella (356), and the predatory relationship of Chrysophyceae with Chlorophyta and Haptista (357,358). In sum, our results point to widespread eukaryote – eukaryote interactions in the ocean.

Our results support SCG as a powerful tool to either corroborate or detect physical interactions involving protists with high-throughput sequencing techniques. Nevertheless, this method has its limitations that need to be considered. For instance, SAGs often display high coverage of some genomic areas and low or no coverage of other areas (359), thus missing potential interactions. This is an issue in regular MDA-based SAG sequencing but accentuated in Low Coverage Sequencing. Yet, Low Coverage Sequencing is a cost-effective approach to screening thousands of cells and can provide useful information to select cells for additional sequencing. Our results

show that different sequencing runs using the same amplified genomes can yield different products, as some observed interactions were exclusive of one replicate.

Another limitation of our methodology is the taxonomic assignation of functional genes, in particular those of eukaryotic origin. Despite not relying only on 18S or 16S rRNA genes to infer potential interactions, which is a single gene that may not be recovered during the amplification step, the approach we have used here requires of an accurate gene prediction tool and a good reference database that includes a decent representation of the taxa being studied. As most protistan genomes are still unknown or poorly characterized, biases during these crucial steps are bound to happen. For example, taxonomy assignation for deeply sequenced SAGs of uncultured eukaryotic organisms (*e.g.,* MASTs, Chrysophyceae) was particularly difficult, as sequences from such organisms in the reference database are not correctly annotated or are missing. In those cases, we used already available 18S rRNA genes for taxonomical identification. Another example of a potentially incorrect taxonomical assignation is the brown algae *Nemacystus* (Phaeophyceae). The genus was assigned to a few SAGs from the Mediterranean Sea despite being a multicellular organism that should have been excluded during cell sorting, as only cells from the picoplankton fraction were isolated. Despite its limitations, with this new approach of using functional genes to investigate potential interactions, we were able to skip the restrictions of assigning taxonomy based exclusively on 18S and 16S rRNA genes, which are usually missing in single-amplified genomes (due to MDA biases) and has been a remarkable problem in past studies with SCG data.

To conclude, using functional genes of 3,015 protistan SAGs from different marine locations allowed us to corroborate predicted interactions from other studies and detect novel interactions (related to predation, parasitism, and symbiosis) among uncultured microorganisms. The provided and extensive collection of eukaryotic SAGs and derived microbial interaction hypotheses serve as a reference to future marine microbial interaction studies. Applying our approach to other marine microorganisms may help to better comprehend the marine interactome and the effects that global change could have on it.

GENERAL DISCUSSION

There is a myriad of microorganisms on Earth contributing to global biogeochemical cycles. In the surface ocean, the smallest microbes (picoplankton) are responsible for an important fraction of the total atmospheric carbon and nitrogen fixation, supporting ca. 50% of the global primary productivity (9). The ocean picoplankton encompasses both prokaryotes (bacteria and archaea) and tiny unicellular eukaryotes. Both groups are very different in terms of cellular structure, feeding, diversity, and reproduction, but are interconnected through biogeochemical and food web networks (17,86). However, the underlying ecological processes determining the biogeography, population dynamics, interactions, and evolution of marine microorganisms are still a mystery for the most part. Comprehending such mechanisms is essential, as changes in the ocean microbial composition could impact the global ecosystem (1).

In this thesis, we aimed at closing the existing knowledge gap on the above topics through the application of High-Throughput Sequencing (HTS) techniques and genomic approaches using global ocean data collected during the *Tara Oceans* and *Malaspina-2010* expeditions, the Gulf of Maine, and two Northwestern Mediterranean coastal sites (BBMO and SOLA stations).

Biogeography and evolution of marine protists

Marine unicellular eukaryotic predators are crucial for the functioning of the ocean ecosystem. Traditionally, these predators represented a single functional group, the heterotrophic flagellates (HFs). However, group members are evolutionary very diverse (52,207). In Chapters 1 and 2, we investigated species belonging to one abundant and widespread group of uncultured marine predators: MAST-4. Originally, MAST-4 was defined as a group of closely related organisms based on SSU 18S rRNA genes (53). However, whether these organisms represented ecotypes of the same species or different species altogether was unknown. In Chapter 1, the substantial genomic divergence observed between the four reconstructed genomes of MAST-4 (A, B, C, and E) using single-cell genomic data (*Tara Oceans* expedition (41)) suggested that these organisms are different species.

Next, we further analyzed the co-occurrence and distribution patterns of MAST-4 using amplicon sequence data (ASV) from the global surface ocean (*Malaspina-2010* expedition (36)). We found contrasting biogeographical patterns between MAST-4

species, pointing to temperature as the main driver shaping the biogeography of MAST-4, agreeing with ASV data from previous works (135). Nevertheless, from whole-genome analyses we observed differences in the repertoire and gene expression of enzymes involved in MAST-4s' degradation machinery, the glycoside hydrolases (GHs). Based on this, we suggested further niche diversification associated with prey digestion: MAST-4 species featuring similar GH composition co-excluded each other (A and C), while species with a different set of GHs appeared to be able to co-exist (B and C). We proposed an evolutionary scenario where species E remained adapted to cold waters, while the Last Common Ancestor of Species A, B, and C adapted to tropical waters. Then, species A adapted to subtropical waters to avoid competition with species C, while B remained in the tropics by changing its GH repertoire.

Genomic strategies similar to the ones presented in Chapter 1 were applied to other uncultured MAST species to obtain their genomes and assess functional diversity and composition (58,102). In these studies, hypotheses for the specialization in terms of cell motility and phagocytosis capabilities related to prey digestion were proposed for several uncultured MAST lineages (such as MAST-1, -4, -7, -8, -9, and -11). Overall, the influx of genomic data from HTS techniques provides an essential framework to study uncultured species that were not obtainable with 18S rRNA surveys alone. Here, we demonstrate that the advances in HTS and bioinformatic tools allow for a better understanding of the genetic content, evolution, and role of uncultured HFs in the ocean.

Population genomics of marine protists across the global ocean

With the increasing number of available protist genomes, new ecological questions can be answered. One of these questions involves the population dynamics of protists in the oceans. Studying the processes shaping population structure is fundamental to understanding the effects of Global Change (77). However, defining populations and investigating population structure with ASV surveys is an almost impossible task, as 16S and 18S rRNA genes often do not hold enough resolution (77).

In previous works, interspecies diversity was found in a few MAST-4 using the Internal Transcribed Spacer (ITS) region of the 18S rRNA (134,135). However, whether this intraspecies diversity represented different ecotypes or populations was not clear. After analyzing the interspecies divergence of the MAST-4 group in Chapter 1,

we aimed to compute the intraspecies divergence in the surface global ocean using the whole genomes recovered from SCG data as references. Since population genomic studies for marine protists are uncommon, there is a gap in available software tackling population genomics for non-model eukaryotic microbes. Still, we bypassed this limitation by integrating diverse methods designed for prokaryotes (POGENOM (89)), model eukaryotic organisms (SnpEff (237)), and plants and animals (FST thresholds (78,79)).

In Chapter 2, we investigated the population genomic patterns of MAST-4 using surface global ocean metagenomic read samples from the *Tara Oceans* expedition (43). We observed strong population differentiation in MAST-4A and C, and weak population differentiation in MAST-4B and E in the global ocean. We defined abundant genomic populations and sub-populations within ocean basins (**Figure 2.2**), particularly in the Mediterranean Sea (MAST-4A and C, **Figure 2.2A and C**), the Indian Ocean (MAST-4C, **Figure 2.2C**), and the Southern Ocean (MAST-4E, **Figure 2.2D**). Furthermore, we found positive selection of MAST-4 genes in specific populations, pointing to niche adaptation in these regions.

On the one hand, the intraspecies divergence of MAST-4B, C, and E was structured by temperature, similar to the interspecies divergence studied in Chapter 1. On the other hand, the intraspecies population structure of MAST-4A in temperate waters was mainly driven by salinity. This suggests that temperature is a key environmental factor driving the evolutionary diversification within the MAST-4 lineage, as each species is adapted to distinct temperature ranges. However, genomic populations within each species appear to be differently adapted to temperature and salinity. Thus, we could theorize that after adapting to temperate zones, species A diversified into different populations due to a diversity of salinity gradients between the Mediterranean Sea and other sub-tropical areas. Endemic populations in the Mediterranean Sea have been found in other marine microbes, such as the bacterium SAR116 (280), which might hint to similar processes occurring within MAST-4.

To further improve our knowledge about the biogeography and population dynamics of MAST-4 and other HFs, samples from other oceanic regions and depths should be included in future investigations. For example, MAST species (including MAST-4) are also found in the bathypelagic and mesopelagic ocean layers

(123,220,231). Using metagenomic read samples from the *Malaspina-2010* expedition could expand our knowledge about MAST-4's biogeography, population dynamics, and adaptation to different depths. Also, SCG from the other two species of MAST-4 not contemplated in this thesis (species D and F) could be added to improve the evolutionary scenario proposed here.

Due to the lack of population genomic studies using whole-genome analyses targeting marine protists, the approaches used in Chapters 1 and 2 can be extrapolated to future investigations of HFs to, for example, reconstruct genomes of uncultured protists, and using the same set of thresholds to define genomic populations (FST > 0.15) (79).

Patterns of population differentiation of marine prokaryotes on a spatiotemporal scale

In Chapter 2 we investigated the population structure of marine microbes in the surface global ocean, where different ranges of environmental conditions are observed between ocean patches. However, the marine environment is a very dynamic medium with seasonal fluctuations (281). Because population differentiation is driven by changes in environmental conditions, investigating the temporal variability is crucial to better understand what are the ecological processes shaping populations, as current genomic features are a product of past evolutionary events (72,360).

Therefore, in Chapter 3, we studied the similitudes and differences of population structure across the surface global ocean (*Tara Oceans* expedition (200)) and two long time-series of monthly data, BBMO (12 years) (100) and SOLA (7 years) (284). For this, we reconstructed and refined 495 prokaryotic MAGs from 7 years of BBMO monthly data. In the surface global ocean, genomic populations of prokaryotic MAGs were strongly differentiated and structured by temperature and salinity. In the same line, similar population structure patterns related to temperature, salinity, and light availability were observed in other marine prokaryotic organisms, such as SAR11, SAR86, *Prochlorococcus,* and *Synechococcus* (88,89,277), and marine protists, such as MAST-4 (Chapter 2).

Although genomic differentiation was high across the surface global ocean within our MAG collection, it is not uncommon to find low-differentiated genomic populations between distant locations under similar environmental conditions. Hugerth *et al.* (361) observed prokaryotic microbes in the Baltic Proper that are genetically

differentiated from closely related microbes, while being highly similar to microbes from North American waters under similar salinity conditions. We observed similar trends in some Mediterranean MAGs, such as SAR11 (**Figure 3.4B**), a ubiquitous microorganism with low-differentiated populations across the surface global ocean, but with a clear underlying population structure related to temperature and salinity (the Mediterranean Sea versus other subtropical waters).

In comparison, population differentiation over 12 and 7 years was genome-specific and could either be strong or weak. Population structure was highly influenced by seasonal environmental changes, *i.e.,* populations were defined based on warm and cold waters. Such seasonal trends were expected as seasonal abundance patterns were observed before in both BBMO and SOLA stations (109,283,284,303). To our knowledge, this thesis represents the first attempt at describing population structure patterns from two different but close (~130 km) long time-series in the Mediterranean Sea. The primary reason being that time-series often use distinct methodological sampling procedures that complicates the integration and comparison of data between them. Contrastingly, although not identical, both time-series were constructed under similar sampling conditions that favored their data integrations, *i.e.,* both are coastal stations and dispose of samples collected during the same periods (January 2009 to December 2015).

Moreover, future studies could be expanded in different directions: a) analyzing more high-quality MAGs from BBMO under the same spatiotemporal context (*e.g.,* including *Prochlorococcus* and *Synechococcus* genomes), b) using MAGs reconstructed from SOLA metagenomes, and c) including more temporal series in the study from different oceanic regions, both within and outside the Mediterranean Sea (97,99). This would allow us to assess if temporal differentiation is shaped by the same processes in different microorganisms and oceanic basins.

Genomic differentiation patterns were shared by the two stations (BBMO and SOLA), suggesting that some populations are common in both locations as a result of spatial proximity (~130 km). However, our analyses of positively selected genes pointed to differences in the adaptation processes between common populations, probably due to the distinct geographical, environmental, and climatological context of each station,

i.e., the influx of freshwater from nearby rivers and sporadic winter storms in SOLA allow for its temperature to be slightly colder throughout the year (284).

The dataset presented in Chapter 3 can become a reference to help future population genomic research to investigate adaptation at both spatial and temporal scales. For example, other genomes from the same MAG collection (BBMO) or other datasets (SOLA, *Tara Oceans*, etc.) could be investigated in both temporal datasets and the global ocean. Moreover, the ecological processes structuring population differentiation at different depths could be assessed using HTS data from the Hawaii Ocean Time-series (HOT) (97) or across the global ocean with *Malaspina-2010* (36).

The protist interactome of the ocean

In Chapters 1, 2, and 3 we tested whether biogeographical distribution patterns of species and populations across space and time were a product of environmental adaptation to abiotic factors, such as temperature and salinity. However, microbial communities consist of many microorganisms that are constantly interacting with each other through the food and biogeochemical networks (127). Deciphering the entire microbial interactome is essential to understand the ocean ecosystem thoroughly, as they guarantee its functioning. For instance, ecological interactions have crucial roles in carbon channeling, control of microalgae blooms by parasites, and phytoplankton associated bacteria influencing the growth and health of their host (106,202,315,354,362).

Efforts to collect all known information regarding the protist interactome have translated into the first Protist Interaction Database (PIDA), which encompasses all the published interactions involving protists until November 2017 (106). Regardless, the interactome of the ocean remains one of the biggest knowledge gaps in today marine microbial ecology. In recent years, association networks based on the correlation of abundances between OTUs gained popularity to investigate microbial interactions. However, the inferred associations represent interaction hypotheses and require further experimental evidence. Single-cell genomics (SCG) has become a powerful tool to find and confirm potential interactions in protists (with bacteria and viruses) and macro-organisms (111,113,363).

In Chapter 4, we investigated potential interactions in more than 3,000 eukaryotic SAGs isolated from the Gulf of Maine (113), the Blanes Bay Microbial Observatory, and a few *Tara Oceans* stations from the Mediterranean Sea and the Indian Ocean. Two types of sequencing depth were used in this SCG data: superficial and low cost-efficient low coverage sequencing (LoCoS) (120), and traditional deep-coverage sequencing (60). Most marine single-cell genomic studies rely on the identification of SSU 16S and 18S rRNA genes cooccurring within a cell, which can be difficult due to the uneven coverage (359) that limits the recovery of rRNA genes within SAGs. For instance, > 95% of the SSU rRNA genes present in LoCoS BBMO SAGs were 16S fragments, which did not allow for a good taxonomical identification of the eukaryotic cells. Moreover, only ca. 50% of these fragments matched to reference databases. For example, from the summer heterotrophic LoCoS BBMO dataset (SH171117), only 22 SAGs (~ 6% of the SAGs) showed both 16S and 18S rRNA gene fragments cooccurring, from which only 5 displayed a good-quality taxonomical representation.

To circumvent this limitation, we developed a new approach based on the prediction and taxonomical annotation of functional genes within a cell. Likewise, we are able to potentially use any amplified gene to infer physical interactions without relying upon specific sequences that might not be recovered. We detected potential interactions within cells from both LoCoS and deep sequencing techniques. We found potential interactions that were already predicted by association networks (protists interacting with Flavobacteriales, Alpha- and Gammaproteobacteria) (109,337,364), proven by cultures (106), or inferred by other SCG studies (MAST-4 with SAR11 and *Synechococcus* (111)), while other potential interactions were unreported before, especially those related to uncultured protists (such as Picozoa, MAST-8, MAST-1 or MAST-11).

Our approach is not restricted to seawater microbial data and can be applied to other fields, such as freshwater, soil, and sediment studies, or even to terrestrial microbes and macro-organisms, including plants and animals. However, potential interactions retrieved from SCG data are restricted to physical associations (attachment, endosymbiosis, parasitism, and predation). One major challenge in interactomics research is the lack of techniques, aside from cultures, to assess and confirm non-physical interactions, which might be as important as physical interactions for the

functioning of the ecosystem and represent a big portion of the marine interactome (103,365).

Overall, our approach was more successful at identifying possible interactions than previous attempts using only 18S and 16S rRNA, which were either missing or wrongly taxonomically assigned. Nevertheless, this approach is highly dependent on a good gene prediction and a correct taxonomical assignation, which can be challenging for marine eukaryotic organisms as their representation in reference databases is lacking. The current state of the field offers a large room for improvement, from developing new methodological approaches to infer associations to enhancing the current tools for gene prediction and taxonomical assignation, including the information stored in reference databases.

Advantages and challenges of HTS technologies

High throughput sequencing (HTS) techniques have become essential data sources in the study of marine microbial communities by allowing the isolation and genome reconstruction of uncultured or unknown microorganisms. Over the last two decades, global initiatives have surveyed the marine ecosystem at different scales using distinct sequencing techniques, such as the *Malaspina-2010* and *Tara Oceans* expeditions or the BBMO and SOLA time-series, which constitute the data sources of this thesis. However, these initiatives use different sampling technologies, sequencing techniques, and sequencing depths, complicating the integration of all the genomic data under the same study. Data normalization is a crucial preliminary step in analyzing genomic datasets that aims at removing global variation to make readings across different experiments comparable (366). For example, three out of seven years of SOLA metagenomic samples were sequenced with different technologies, resulting in less sequencing depths and smaller read lengths (100 vs. 150 bp), for which normalization of abundance were crucial for further population genomic analyses (Chapter 2).

Despite all the efforts in sampling and sequencing microbial diversity, the number of studies focusing on eukaryotic diversity is limited. In the *Science* family of journals, there are currently 742 studies (dated in September 2022) related to Prokaryotic, Diversity, and DNA, while only 141 to Protists, Diversity, and DNA. In this thesis, three out of four chapters are dedicated to marine eukaryotic protists. For

this reason, during the preparation of this thesis, we faced and tried to overcome some of challenges that commonly appear when working with eukaryotic sequencing data as a consequence to the low number of research dedicated to eukaryotic microbes.

The first challenge is sequencing the actual data. In SCG approaches, where eukaryotic cells are selected beforehand, the typical genome recovery from a eukaryotic SAG is about 20% (81). In part, this is a consequence of the natural complexity of eukaryotic cells, where the nucleus protects the compact and linear chromosomes. Thus, cell lysis techniques have to break different cellular structures to free the DNA, which might result in a lesser DNA amplification than expected. In addition, eukaryotic cells also have organelles (*e.g.,* chloroplast, mitochondria) with their own circular DNA, that might be more accessible and easier amplified than nuclear DNA (103). At the end, this leads to sequence biases and incomplete genomes.

A solution to increase the genome completeness during SCG is to co-assemble together multiple SAGs. This approach was applied in Chapter 1 and was effective at increasing the total genome recovery of protist species (81). Since each SAG recovers ca. 20% of the whole genome, in a co-assembly strategy, each cell contributes with new unique genomic information to complete the whole genome. Yet, the relationship between the number of SAGs co-assembled and the amount of total completeness is not linear, meaning that there is a limit to the amount of genomic information that one can recover (81), most likely due to some regions of the genome being easier amplified than others during single-cell sequencing.

High-quality genomes are required to obtain unbiased results that can answer ecological questions. Consequently, genome decontamination, *i.e.,* removing foreign or unwanted sequences from a genome, is crucial in many genomic surveys. Although some forms of contamination are interesting to assess certain topics, *e.g.,* interactions (Chapter 4), others are a product of technological drawbacks, *e.g.,* cross-contamination between multiplexed libraries (367).

For eukaryotic genomes obtained from genomic approaches, the second challenge is the lack of standardized protocols and available software for quality checking and removing contamination. While there are solid bioinformatic tools for prokaryotic genome decontamination (*e.g.,* Anvi'o (368), CheckM (125)) that can be

used to remove bacterial DNA within eukaryotic genomes (369), these normally do not resolve contamination from other eukaryotic sources, as they focus in prokaryotic genetic structures (*i.e.,* no introns, no splicing) or rely upon taxonomical assignation and gene markers detection from reference databases. Taxonomic representation of marine eukaryotes in reference databases is another pivotal issue, especially for uncultured marine species, as they often are incorrectly assigned or completely missing. Therefore, taxonomic-based cleaning approaches in eukaryotic genomes usually underperform unless a custom and curated reference database is built.

In the past five years, machine-learning based bioinformatics tools have emerged allowing to recruit eukaryotic sequences from metagenomic data (370). Such methods can be used to certain extent to decontaminate eukaryotic genomes based on tetra- and penta-nucleotide frequencies (*e.g.,* EukRep (161), Tiara (325)). In Chapter 1, we developed a decontamination pipeline using some of the machine-learning tools (ESOM (1) and EukRep) (156,161) to obtain clean reference genomes from SCG data that allowed us to accurately assess the ecological questions asked (*i.e.,* genomic differentiation, genetic and functional content).

During this thesis period (2017 – 2022), the number of advances and improvements in the field of non-model eukaryotic marine microorganisms has increased substantially. Sequencing and integrating more genomes from uncultured marine microorganisms will improve the representation and quality of eukaryotes in reference databases, and therefore, new and upgraded tools will emerge that will allow for improved genome assemblies, decontamination pipelines, gene prediction, and functional annotation of eukaryotic microbes.

Final remarks

The main objective of this thesis has been to investigate how marine microbial communities are structured and what are the processes shaping them in the ocean on a large spatiotemporal scale. Thus, we aimed to describe the biogeography, population structure, and ecological interactions of selected marine microbes. We fulfilled the four specific objectives established at the beginning of this thesis: we described biogeographical patterns and interspecies genomic diversity among protists on the surface open ocean (objective 1); we characterized intraspecies genomic differentiation

of eukaryotic (objective 2) and prokaryotic (objective 3) microorganisms in either the global ocean, 12 and 7 years of temporal data in the Mediterranean Sea, or both; and reported ecological interactions occurring within marine protists (objective 4).

BIBLIOGRAPHY

1. Han W. Oceans and Climate. In: International Encyclopedia of Geography [Internet]. John Wiley & Sons, Ltd; 2017 [cited 2022 Sep 6]. p. 1–10. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118786352.wbieg0666
2. Brini E, Fennell CJ, Fernandez-Serra M, Hribar-Lee B, Lukšič M, Dill KA. How Water's Properties Are Encoded in Its Molecular Structure and Energies. Chem Rev. 2017 Oct 11;117(19):12385–414.
3. Buonocore E, Grande U, Franzese PP, Russo GF. Trends and Evolution in the Concept of Marine Ecosystem Services: An Overview. Water [Internet]. 2021 [cited 2022 Sep 6];13(15). Available from: https://www.sciencegate.app/document/10.3390/w13152060
4. Hegerl GC, Black E, Allan RP, Ingram WJ, Polson D, Trenberth KE, et al. Challenges in Quantifying Changes in the Global Water Cycle. Bulletin of the American Meteorological Society. 2015 Jul 1;96(7):1097–115.
5. Lønborg C, Carreira C, Jickells T, Álvarez-Salgado XA. Impacts of Global Change on Ocean Dissolved Organic Carbon (DOC) Cycling. Frontiers in Marine Science [Internet]. 2020 [cited 2022 Sep 6];7. Available from: https://www.frontiersin.org/articles/10.3389/fmars.2020.00466
6. Bindoff N, Willebrand J, Artale V, Cazenave A, Gregory J, Gulev S, et al. Observations: oceanic climate and sea level. In: Climate change 2007: The physical Science Basis [Internet]. 2007 [cited 2022 Sep 6]. p. 385–432. Available from: https://hal.archives-ouvertes.fr/hal-00287145
7. Moreno J, Møller AP. Extreme climatic events in relation to global change and their impact on life histories. Current Zoology. 2011 Jun 1;57(3):375–89.
8. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Ocean plankton. Structure and function of the global ocean microbiome. Science (New York, NY). 2015 May 22;348(6237):1261359.
9. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science [Internet]. 1998 [cited 2017 Jun 19];281(5374). Available from: http://science.sciencemag.org/content/281/5374/237.long
10. Falkowski PG, Barber RT, Smetacek V. Biogeochemical Controls and Feedbacks on Ocean Primary Production. Science (New York, NY). 1998 Jul 10;281(5374):200–7.
11. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. Science. 2008 May 23;320(5879):1034–9.
12. Litchman E, de Tezanos Pinto P, Edwards KF, Klausmeier CA, Kremer CT, Thomas MK. Global biogeochemical impacts of phytoplankton: A trait-based perspective. Journal of Ecology. 2015 Nov 1;103(6):1384–96.
13. Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L, et al. Regime Shifts, Resilience, and Biodiversity in Ecosystem Management. Annual Review of Ecology, Evolution, and Systematics. 2004;35(1):557–81.
14. Borja A. Grand challenges in marine ecosystems ecology. Frontiers in Marine Science [Internet]. 2014 [cited 2022 Jul 24];1. Available from: https://www.frontiersin.org/articles/10.3389/fmars.2014.00001
15. Begon M, Townsend CR, Harper JL. Ecology: From Individuals to Ecosystems [Internet]. 4th Edition. Wiley-Blackwell; 2005 [cited 2022 Sep 6]. 750 p. Available from: https://www.wiley.com/en-us/Ecology%3A+From+Individuals+to+Ecosystems%2C+4th+Edition-p-9781405111171
16. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol. 2006 Feb;4(2):102–12.
17. DeLong EF, Karl DM. Genomic perspectives in microbial oceanography. Nature. 2005 Sep 15;437(7057):336–42.
18. Herndl GJ, Reinthaler T. Microbial control of the dark end of the biological pump. Nature Geosci. 2013 Sep;6(9):718–24.
19. Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM, et al. Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. PLOS ONE. 2011 Sep 8;6(9):e24570.
20. Ghiglione JF, Galand PE, Pommier T, Pedrós-Alió C, Maas EW, Bakker K, et al. Pole-to-pole biogeography of surface and deep marine bacterial communities. Proceedings of the National Academy of Sciences. 2012 Oct 23;109(43):17633–8.
21. Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, et al. Defining seasonal marine microbial community dynamics. ISME J. 2012 Feb;6(2):298–308.
22. Fuhrman JA. Microbial community structure and its functional implications. Nature. 2009 May;459(7244):193–9.
23. Seymour JR, Doblin MA, Jeffries TC, Brown MV, Newton K, Ralph PJ, et al. Contrasting microbial assemblages in adjacent water masses associated with the East Australian Current. Environ Microbiol Rep. 2012 Oct;4(5):548–55.
24. Needham DM, Chow CET, Cram JA, Sachdeva R, Parada A, Fuhrman JA. Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. ISME J. 2013 Jul;7(7):1274–85.
25. Beier S, Andersson AF, Galand PE, Hochart C, Logue JB, McMahon K, et al. The environment drives microbial trait variability in aquatic habitats. Molecular Ecology. 2020;29(23):4605–17.
26. Shade A, Carey CC, Kara E, Bertilsson S, McMahon KD, Smith MC. Can the black box be cracked? The augmentation of microbial ecology by high-resolution, automated sensing technologies. ISME J. 2009 Aug;3(8):881–8.
27. Baumann P, Baumann L, Woolkalis MJ, Bang SS. Evolutionary relationships in vibrio and Photobacterium: a basis for a natural classification. Annu Rev Microbiol. 1983;37:369–98.

28. Lewin RA. A classification of flexibacteria. J Gen Microbiol. 1969 Oct;58(2):189–206.
29. Haeckel E. Report on the Radiolaria collected by H.M.S. Challenger during the years 1873-76. [Edinburgh, London, Eng: Printed for H.M. Stationery Off. by Neill and Co.]; 1887. 3 v. :
30. Marteinsson VÞ, Groben R, Reynisson E, Vannier P. Biogeography of Marine Microorganisms. In: Stal LJ, Cretoiu MS, editors. The Marine Microbiome: An Untapped Source of Biodiversity and Biotechnological Potential [Internet]. Cham: Springer International Publishing; 2016 [cited 2022 Sep 6]. p. 187–207. Available from: https://doi.org/10.1007/978-3-319-33000-6_6
31. Tseng CH, Tang SL. Marine Microbial Metagenomics: From Individual to the Environment. International Journal of Molecular Sciences. 2014 May;15(5):8878–92.
32. Biller SJ, Berube PM, Dooley K, Williams M, Satinsky BM, Hackl T, et al. Marine microbial metagenomes sampled across space and time. Sci Data. 2018 Sep 4;5(1):180176.
33. Villar E, Vannier T, Vernette C, Lescot M, Cuenca M, Alexandre A, et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. Nucleic Acids Research. 2018 Jul 2;46(W1):W289–95.
34. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, et al. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLOS Biology. 2007 Mar 13;5(3):e77.
35. Amaral-Zettler L, Artigas LF, Baross J, Bharathi P.A. L, Boetius A, Chandramohan D, et al. A Global Census of Marine Microbes. In: Life in the World's Oceans [Internet]. John Wiley & Sons, Ltd; 2010 [cited 2022 Sep 6]. p. 221–45. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781444325508.ch12
36. Duarte CM. Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. Limnology and Oceanography Bulletin. 2015;
37. Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, et al. Open science resources for the discovery and analysis of Tara Oceans data. Sci Data. 2015 May 26;2(1):150023.
38. Morales L, Dachs J, González-Gaya B, Hernán G, Abalos M, Abad E. Background concentrations of polychlorinated dibenzo-p-dioxins, dibenzofurans, and biphenyls in the global oceanic atmosphere. Environ Sci Technol. 2014 Sep 2;48(17):10198–207.
39. Cózar A, Echevarría F, González-Gordillo JI, Irigoien X, Úbeda B, Hernández-León S, et al. Plastic debris in the open ocean. Proceedings of the National Academy of Sciences. 2014 Jul 15;111(28):10239–44.
40. BENZONI F, ARRIGONI R, BERUMEN ML., TAVIANI M, BONGAERTS P, FRADE PR, et al. Morphological and genetic divergence between Mediterranean and Caribbean populations of Madracis pharensis (Heller 1868) (Scleractinia, Pocilloporidae): too much for one species? Zootaxa. 2018 Sep 6;4471(3):473.
41. de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. Science. 2015 May 22;348(6237):1261605.
42. Alberti A, Poulain J, Engelen S, Labadie K, Romac S, Ferrera I, et al. Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. Scientific Data. 2017;
43. Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, et al. A global ocean atlas of eukaryotic genes. Nature Communications. 2018;
44. Logares R, Deutschmann IM, Junger PC, Giner CR, Krabberød AK, Schmidt TSB, et al. Disentangling the mechanisms shaping the surface ocean microbiota. Microbiome. 2020 Apr 20;8(1):55.
45. Giovannoni SJ. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. Annual Review of Marine Science. 2017;
46. Giovannoni SJ, Stingl U. Molecular diversity and ecology of microbial plankton. Nature. 2005 Sep 15;437(7057):343–8.
47. Erguder TH, Boon N, Wittebolle L, Marzorati M, Verstraete W. Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. FEMS Microbiology Reviews. 2009 Sep 1;33(5):855–69.
48. Díez B, Pedrós-Alió C, Massana R. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Appl Environ Microbiol. 2001 Jul;67(7):2932–41.
49. López-García P, Vereshchaka A, Moreira D. Eukaryotic diversity associated with carbonates and fluid–seawater interface in Lost City hydrothermal field. Environmental Microbiology. 2007;9(2):546–54.
50. Countway PD, Gast RJ, Savai P, Caron DA. Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. J Eukaryot Microbiol. 2005 Apr;52(2):95–106.
51. Massana R, Guillou L, Díez B, Pedrós-Alió C. Unveiling the organisms behind novel eukaryotic ribosomal DNA sequences from the ocean. Applied and Environmental Microbiology. 2002 Sep;68(9):4554–8.
52. Massana R, Terrado R, Forn I, Lovejoy C, Pedros-Alio C. Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. Environmental Microbiology. 2006 Sep;8(9):1515–22.
53. Massana R, del Campo J, Sieracki ME, Audic S, Logares R. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. The ISME journal. 2014 Apr 7;8(4):854–66.
54. Logares R, Haverkamp THA, Kumar S, Lanzén A, Nederbragt AJ, Quince C, et al. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. Journal of microbiological methods. 2012 Oct;91(1):106–13.
55. Green S, Studholme DJ, Laue BE, Dorati F, Lovell H, Arnold D, et al. Comparative Genome Analysis Provides Insights into the Evolution and Adaptation of Pseudomonas syringae pv. aesculi on Aesculus hippocastanum. PLOS ONE. 2010 Apr 19;5(4):e10224.
56. Kanter I, Kalisky T. Single cell transcriptomics: methods and applications. Frontiers in oncology. 2015;5:53.
57. Mukherjee A, Reddy MS. Metatranscriptomics: an approach for retrieving novel eukaryotic genes from

polluted and related environments. 3 Biotech. 2020 Feb 1;10(2):1–19.
58. Labarre A, López-Escardó D, Latorre F, Leonard G, Bucchini F, Obiol A, et al. Comparative genomics reveals new functional insights in uncultured MAST species. ISME Journal [Internet]. 2021 [cited 2021 Feb 9]; Available from: https://pubmed.ncbi.nlm.nih.gov/33452482/
59. Stepanauskas R. Single cell genomics: an individual look at microbes. Current opinion in microbiology. 2012 Oct;15(5):613–20.
60. Heywood JL, Sieracki ME, Bellows W, Poulton NJ, Stepanauskas R. Capturing diversity of marine heterotrophic protists: one cell at a time. The ISME journal. 2011 Apr;5(4):674–84.
61. Latorre F, Deutschmann IM, Labarre A, Obiol A, Krabberød AK, Pelletier E, et al. Niche adaptation promoted the evolutionary diversification of tiny ocean predators. Proceedings of the National Academy of Sciences. 2021 Jun 22;118(25):e2020955118.
62. Baas Becking LGM. Geobiologie of inleiding tot de milieukunde. Den Haag: W.P. Van Stockum & Zoon; 1934. 263 p.
63. Finlay BJ. Global dispersal of free-living microbial eukaryote species. Science (New York, NY). 2002 May 10;296(5570):1061–3.
64. Finlay BJ, Fenchel T. Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist. 2004 Jun;155(2):237–44.
65. Brown MV, Lauro FM, DeMaere MZ, Muir L, Wilkins D, Thomas T, et al. Global biogeography of SAR11 marine bacteria. Mol Syst Biol. 2012 Jul 17;8:595.
66. Saez AG, Probert I, Geisen M, Quinn P, Young JR, Medlin LK. Pseudo-cryptic speciation in coccolithophores. Proc Natl Acad Sci U S A. 2003 Jun 10;100(12):7163–8.
67. Bass D, Richards TA, Matthai L, Marsh V, Cavalier-Smith T. DNA evidence for global dispersal and probable endemicity of protozoa. BMC Evol Biol. 2007 Sep 13;7:162.
68. Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. Nat Rev Microbiol. 2015 Mar;13(3):133–46.
69. Gattuso JP, Gentili B, Duarte CM, Kleypas JA, Middelburg JJ, Antoine D. Light availability in the coastal ocean: impact on the distribution of benthic photosynthetic organisms and their contribution to primary production. Biogeosciences. 2006 Nov 6;3(4):489–513.
70. Sugie K, Fujiwara A, Nishino S, Kameyama S, Harada N. Impacts of Temperature, CO2, and Salinity on Phytoplankton Community Composition in the Western Arctic Ocean. Frontiers in Marine Science [Internet]. 2020 [cited 2022 Sep 6];6. Available from: https://www.frontiersin.org/articles/10.3389/fmars.2019.00821
71. Qin QL, Wang ZB, Cha QQ, Liu SS, Ren XB, Fu HH, et al. Biogeography of culturable marine bacteria from both poles reveals that 'everything is not everywhere' at the genomic level. Environmental Microbiology. 2022;24(1):98–109.
72. Koskella B, Vos M. Adaptation in Natural Microbial Populations. Annual Review of Ecology, Evolution, and Systematics. 2015;46(1):503–22.
73. Walworth NG, Zakem EJ, Dunne JP, Collins S, Levine NM. Microbial evolutionary strategies in a dynamic ocean. Proceedings of the National Academy of Sciences. 2020 Mar 17;117(11):5943–8.
74. Lynch M, Gabriel W, Wood AM. Adaptive and demographic responses of plankton populations to environmental change. Limnology and Oceanography. 1991;36(7):1301–12.
75. Rengefors K, Kremp A, Reusch TBH, Wood AM. Genetic diversity and evolution in eukaryotic phytoplankton: revelations from population genetic studies. Journal of Plankton Research. 2017 Mar 1;39(2):165–79.
76. Waples RS, Gaggiotti O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol. 2006 May;15(6):1419–39.
77. VanInsberghe D, Arevalo P, Chien D, Polz MF. How can microbial population genomics inform community ecology? Philosophical Transactions of the Royal Society B: Biological Sciences. 2020 May 11;375(1798):20190253.
78. Hartl DL, Clark AG. Principles of Population Genetics. Vol. 116. Sunderland: Sinauer Associates; 1997.
79. Frankham R, Ballou JD, Briscoe DA. Introduction to Conservation Genetics [Internet]. Cambridge: Cambridge University Press; 2002 [cited 2022 Sep 12]. Available from: https://www.cambridge.org/core/books/introduction-to-conservation-genetics/F1F8EDB8B86A1790A406064296878B23
80. Acinas SG, Sánchez P, Salazar G, Cornejo-Castillo FM, Sebastián M, Logares R, et al. Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. Commun Biol. 2021 May 21;4(1):1–15.
81. Mangot JF, Logares R, Sánchez P, Latorre F, Seeleuthner Y, Mondy S, et al. Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. Scientific Reports. 2017;7.
82. Haro-Moreno JM, López-Pérez M, Rodriguez-Valera F. Enhanced Recovery of Microbial Genes and Genomes From a Marine Water Column Using Long-Read Metagenomics. Front Microbiol. 2021 Aug 27;12:708782.
83. Pérez-Cobas AE, Gomez-Valero L, Buchrieser C. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. Microb Genom. 2020 Jul 24;6(8):mgen000409.
84. Logares R. Population genetics: the next stop for microbial ecologists? Open Life Sciences. 2011 Dec 1;6(6):887–92.
85. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004 Apr 2;304(5667):66–74.
86. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving carbon export in the oligotrophic ocean. Nature. 2016 Apr 10;532(7600):465–70.

87. Boeuf D, Edwards BR, Eppley JM, Hu SK, Poff KE, Romano AE, et al. Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. Proceedings of the National Academy of Sciences. 2019 Jun 11;116(24):11824–32.
88. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, et al. Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. eLife. 2019 Sep 1;8.
89. Sjöqvist C, Delgado LF, Alneberg J, Andersson AF. Ecologically coherent population structure of uncultivated bacterioplankton. ISME Journal. 2021 May 5;1–16.
90. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. Genome Biol. 2013 May 29;14(5):R51.
91. Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, et al. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Frontiers in Genetics [Internet]. 2015 [cited 2022 Sep 7];6. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2015.00235
92. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 2012 Oct 4;490(7418):49–54.
93. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012 Apr 4;484(7392):55–61.
94. Casteleyn G, Leliaert F, Backeljau T, Debeer AE, Kotaki Y, Rhodes L, et al. Limits to gene flow in a cosmopolitan marine planktonic diatom. Proceedings of the National Academy of Sciences. 2010 Jul 20;107(29):12952–7.
95. Whittaker KA, Rynearson TA. Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. Proceedings of the National Academy of Sciences. 2017 Mar 7;114(10):2651–6.
96. Leconte J, Timsit Y, Delmont TO, Lescot M, Piganeau G, Wincker P, et al. Equatorial to Polar genomic variability of the microalgae Bathycoccus prasinos [Internet]. bioRxiv; 2021 [cited 2022 Jul 13]. p. 2021.07.13.452163. Available from: https://www.biorxiv.org/content/10.1101/2021.07.13.452163v3
97. Karl DM, Lukas R. The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. Deep Sea Research Part II: Topical Studies in Oceanography. 1996 Jan 1;43(2):129–56.
98. Steinberg DK, Carlson CA, Bates NR, Johnson RJ, Michaels AF, Knap AH. Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. Deep Sea Research Part II: Topical Studies in Oceanography. 2001 Jan 1;48(8):1405–47.
99. d'Alcalà MR, Conversano F, Corato F, Licandro P, Mangoni O, Marino D, et al. Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): an attempt to discern recurrences and trends. Scientia Marina. 2004 Apr 30;68(S1):65–83.
100. Gasol JM, Cardelús C, Morán XAG, Balagué V, Forn I, Marrasé C, et al. Seasonal patterns in phytoplankton photosynthetic parameters and primary production at a coastal NW Mediterranean site. Scientia Marina. 2016 Sep 30;80(S1):63–77.
101. Charles F, Lantoine F, Brugel S, Chrétiennot-Dinet MJ, Quiroga I, Rivière B. Seasonal survey of the phytoplankton biomass, composition and production in a littoral NW Mediterranean site, with special emphasis on the picoplanktonic contribution. Estuarine, Coastal and Shelf Science. 2005 Oct 1;65(1):199–212.
102. Seeleuthner Y, Mondy S, Lombard V, Carradec Q, Pelletier E, Wessner M, et al. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. Nature Communications. 2018 Dec 1;9(1):1–10.
103. Krabberød A, Bjorbækmo MFM, Schalchian-Tabrizi K, Logares R. Exploring the oceanic microeukaryotic interactome with metaomics approaches. AQUATIC MICROBIAL ECOLOGY Aquat Microb Ecol. 2017;79:1–12.
104. Deutschmann IM, Lima-Mendez G, Krabberød AK, Raes J, Vallina SM, Faust K, et al. Disentangling environmental effects in microbial association networks. Microbiome. 2021 Nov 26;9(1):232.
105. Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F. The ecological role of water-column microbes in the Sea. Mar Ecol Prog Ser. 1983 Jan 20;10:257–63.
106. Bjorbækmo MFM, Evenstad A, Røsæg LL, Krabberød AK, Logares R. The planktonic protist interactome: where do we stand after a century of research? ISME J. 2020 Feb;14(2):544–59.
107. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods. 2008 Mar;5(3):235–7.
108. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J. 2011 Sep;5(9):1414–25.
109. Krabberød AK, Deutschmann IM, Bjorbækmo MFM, Balagué V, Giner CR, Ferrera I, et al. Long-term patterns of an interconnected core marine microbiota. Environmental Microbiome. 2022 May 7;17(1):22.
110. Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, et al. Determinants of community structure in the global plankton interactome. Science. 2015 May 22;348(6237):1262073–1262073.
111. Martinez-Garcia M, Brazel D, Poulton NJ, Swan BK, Gomez ML, Masland D, et al. Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. The ISME journal. 2012 Mar;6(3):703–7.
112. Castillo YM, Mangot JF, Benites LF, Logares R, Kuronishi M, Ogata H, et al. Assessing the viral content of uncultured picoeukaryotes in the global-ocean by single cell genomics. Molecular Ecology. 2019;28(18):4272–89.
113. Brown JM, Labonté JM, Brown J, Record NR, Poulton NJ, Sieracki ME, et al. Single Cell Genomics Reveals Viruses Consumed by Marine Protists. Frontiers in Microbiology. 2020 Sep 24;11:2317.
114. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, et al. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. The ISME journal. 2015 Nov;9(11):2386–99.
115. Dong X, Kleiner M, Sharp CE, Thorson E, Li C, Liu D, et al. Fast and Simple Analysis of MiSeq Amplicon

Sequencing Data with MetaAmp. Frontiers in Microbiology [Internet]. 2017 [cited 2022 Sep 12];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.01461
116. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 2016;
117. Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, et al. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science. 2011 Sep 2;333(6047):1296–300.
118. Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PSG, Reitenga KG, et al. Capturing Single Cell Genomes of Active Polysaccharide Degraders: An Unexpected Contribution of Verrucomicrobia. PLOS ONE. 2012 Apr 20;7(4):e35314.
119. Martinez-Garcia M, Swan BK, Poulton NJ, Gomez ML, Masland D, Sieracki ME, et al. High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. ISME J. 2012 Jan;6(1):113–23.
120. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, et al. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nature Communications. 2017;8(1):84.
121. Tolonen AC, Xavier RJ. Dissecting the human microbiome with single-cell genomics. Genome Med. 2017 Jun 14;9:56.
122. Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H, et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environmental Microbiology. 2014 Sep;16(9):2659–71.
123. Obiol A, Giner CR, Sánchez P, Duarte CM, Acinas SG, Massana R. A metagenomic assessment of microbial eukaryotic diversity in the global ocean. Molecular ecology resources [Internet]. 2020 May [cited 2020 Jun 26];20(3). Available from: http://www.ncbi.nlm.nih.gov/pubmed/32065492
124. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nature Biotechnology. 2017 Sep 12;35(9):833–44.
125. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research. 2015 Jul 1;25(7):1043–55.
126. Li WKW. Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: Measurements from flow cytometric sorting. Limnology and Oceanography. 1994;
127. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. Science. 2015 Feb 13;347(6223):1257594–1257594.
128. Pernthaler J. Predation on prokaryotes in the water column and its ecological implications. Nature Reviews Microbiology. 2005 Jul 10;3(7):537–46.
129. Massana R, Unrein F, Rodríguez-Martínez R, Forn I, Lefort T, Pinhassi J, et al. Grazing rates and functional diversity of uncultured heterotrophic flagellates. ISME Journal. 2009 May;3(5):588–95.
130. Derelle R, López-García P, Timpano H, Moreira D. A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (=Heterokonts). Molecular Biology and Evolution. 2016 Nov;33(11):2890–8.
131. Lin YC, Campbell T, Chung CC, Gong GC, Chiang KP, Worden AZ. Distribution patterns and phylogeny of marine stramenopiles in the north pacific ocean. Applied and environmental microbiology. 2012 May 1;78(9):3387–99.
132. Piwosz K, Wiktor JM, Niemi A, Tatarek A, Michel C. Mesoscale distribution and functional diversity of picoeukaryotes in the first-year sea ice of the Canadian Arctic. ISME Journal. 2013;
133. Piwosz K, Pernthaler J. Seasonal population dynamics and trophic role of planktonic nanoflagellates in coastal surface waters of the Southern Baltic Sea. Environmental Microbiology. 2010;
134. Rodríguez-Martínez R, Rocap G, Logares R, Romac S, Massana R. Low evolutionary diversification in a widespread and abundant uncultured protist (MAST-4). Molecular biology and evolution. 2012 May;29(5):1393–406.
135. Rodríguez-Martínez R, Rocap G, Salazar G, Massana R. Biogeography of the uncultured marine picoeukaryote MAST-4: temperature-driven distribution patterns. The ISME Journal. 2013 Apr 18;7(8):1531–43.
136. Gasol JM. A framework for the assessment of top-down vs bottom-up control of heterotrophic nanoflagellate abundance. Marine Ecology Progress Series. 1994;
137. Lee WJ, Patterson DJ. Abundance and biomass of heterotrophic flagellates, and factors controlling their abundance and distribution in sediments of Botany Bay. Microbial Ecology. 2002;
138. Meira BR de, Lansac-Tôha FM, Segovia BT, Oliveira FR de, Buosi PRB, Jati S, et al. Abundance and size structure of planktonic protist communities in a Neotropical floodplain: effects of top-down and bottom-up controls. Acta Limnologica Brasiliensia. 2017;
139. Schulze H, Kolter T, Sandhoff K. Principles of lysosomal membrane degradation: Cellular topology and biochemistry of lysosomal lipid degradation. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2009 Apr 1;1793(4):674–83.
140. Manchenko GP. Handbook of detection of enzymes on electrophoretic gels, second edition. Handbook of Detection of Enzymes on Electrophoretic Gels, Second Edition. 2002.
141. Sieracki ME, Poulton NJ, Jaillon O, Wincker P, de Vargas C, Rubinat-Ripoll L, et al. Single cell genomics yields a wide diversity of small planktonic protists across major ocean ecosystems. Scientific Reports. 2019 Dec 15;9(1):6025.
142. Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, et al. Multiple marker parallel tag

environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology. 2010;
143. Parada AE, Needham DM, Fuhrman JA. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology. 2016;
144. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology. 2007;
145. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 2013 Jan 1;41(D1):D590–6.
146. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research. 2013;
147. Albanese D, Riccadonna S, Donati C, Franceschi P. A practical tool for maximal information coefficient analysis. GigaScience. 2018;
148. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME Journal. 2016;
149. Deutschmann IM. EnDED - - Environmentally-Driven Edge Detection Program [Internet]. Zenodo; 2019. Available from: http://doi.org/10.5281/zenodo.3271730
150. Salazar G. EcolUtils: Utilities for community ecology analysis [Internet]. 2019. Available from: https://github.com/GuillemSalazar/EcolUtils
151. Stepanauskas R, Sieracki ME. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proceedings of the National Academy of Sciences of the United States of America. 2007 May 22;104(21):9052–7.
152. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology. 2012 May;19(5):455–77.
153. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics (Oxford, England). 2013 Apr 15;29(8):1072–5.
154. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution. 2018;
155. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. 2016 Mar 27 [cited 2020 Jan 26]; Available from: https://peerj.com/preprints/1900/
156. Ultsch A, Morchen F. ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. 2009 Jan 6 [cited 2016 May 15]; Available from: https://www.researchgate.net/publication/246090732_ESOM-Maps_tools_for_clustering_visualization_and_classification_with_Emergent_SOM
157. Bushnell B, Rood J, Singer E. BBMerge – Accurate paired shotgun read merging via overlap. PLoS ONE. 2017;
158. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics (Oxford, England). 2011 Feb 15;27(4):578–9.
159. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015 . http://www.repeatmasker.org. 2013.
160. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic acids research. 2016;
161. West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. Genome-reconstruction for eukaryotes from complex natural microbial communities. Genome Research. 2018;
162. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct 1;31(19):3210–2.
163. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research. 2005 Jul 1;33(Web Server):W465–7.
164. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics (Oxford, England). 2008 Mar 1;24(5):637–44.
165. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research. 2012;
166. Eddy SR. BIOINFORMATICS REVIEW Profile hidden Markov models. Bioinformatics Review. 1998;
167. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000 Jan 1;28(1):27–30.
168. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research. 2015 Oct 17;44(D1):D457-62.
169. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research. 2016;
170. Hackl T, Martin R, Barenhoff K, Duponchel S, Heider D, Fischer MG. Four high-quality draft genome assemblies of the marine heterotrophic nanoflagellate Cafeteria roenbergensis. Scientific Data. 2020 Dec 21;7(1):29.
171. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution. 2013;

172. Nylander JAA. catfasta2phyml [Internet]. Available from: https://github.com/nylander/catfasta2phyml
173. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;
174. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;
175. Suzuki R, Shimodaira H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;
176. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009 Jul 15;25(14):1754–60.
177. Koncevičius K. matrixTests: Fast Statistical Hypothesis Tests on Rows and Columns of Matrices [Internet]. 2020. Available from: https://github.com/KKPMW/matrixTests
178. Suyama M, Torrents D, Bork P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research. 2006;
179. Kosakovsky Pond SL, Frost SDW, Muse SV. HyPhy: Hypothesis testing using phylogenies. Bioinformatics. 2005;
180. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Molecular Biology and Evolution. 2015 May 1;32(5):1342–53.
181. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genetics. 2012 Jul;8(7):1002764.
182. Moreira D, López-García P. The rise and fall of Picobiliphytes: How assumed autotrophs turned out to be heterotrophs. BioEssays : news and reviews in molecular, cellular and developmental biology. 2014;36(5):468.
183. Cuvelier ML, Ortiz A, Kim E, Moehlig H, Richardson DE, Heidelberg JF, et al. Widespread distribution of a unique marine protistan lineage. Environmental Microbiology. 2008 Jun;10(6):1621–34.
184. Mangot JF, Logares R, Sánchez P, Latorre F, Seeleuthner Y, Mondy S, et al. Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. Scientific Reports. 2017;7.
185. Massana R, del Campo J, Sieracki ME, Audic S, Logares R. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. The ISME journal. 2014 Apr 7;8(4):854–66.
186. Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, et al. A global ocean atlas of eukaryotic genes. Nature Communications. 2018;
187. Berlemont R, Martiny AC. Glycoside Hydrolases across Environmental Microbial Communities. PLoS computational biology. 2016;12(12):e1005300.
188. Seeleuthner Y, Mondy S, Lombard V, Carradec Q, Pelletier E, Wessner M, et al. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. Nature Communications. 2018 Dec 1;9(1):1–10.
189. Madigan MT, Martinko JM, Stahl DA, Clark DP. Brock Biology of Microorganisms 13th Edition. 2009.
190. Rabinovich ML, Melnick MS, Bolobova AV. The Structure and Mechanism of Action of Cellulolytic Enzymes. Biochemistry (Moscow). 2002;67(8):850–71.
191. Naumoff DG. GH97 is a new family of glycoside hydrolases, which is related to the α-galactosidase superfamily. BMC Genomics. 2005;
192. Naumoff DG. GH101 family of glycoside hydrolases: Subfamily structure and evolutionary connections with other families. In: Journal of Bioinformatics and Computational Biology. 2010.
193. J. A, Ohno S. Evolution by Gene Duplication. Population (French Edition). 1971;
194. Walsh JB. How often do duplicated genes evolve new functions? Genetics. 1995 Jan;139(1):421–8.
195. Wagner A. The fate of duplicated genes: Loss or new function? BioEssays. 1998 Dec 12;20(10):785–8.
196. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends in Ecology & Evolution. 2000;15(12):496.
197. Mangot JF, Forn I, Obiol A, Massana R. Constant abundances of ubiquitous uncultured protists in the open sea assessed by automated microscopy. Environmental microbiology. 2018;20(10):3876–89.
198. Vellend M. The Theory of Ecological Communities (MPB-57). The Theory of Ecological Communities (MPB-57). 2016.
199. Lindström ES, Langenheder S. Local and regional factors influencing bacterial community assembly. Environmental Microbiology Reports. 2012 Feb 1;4(1):1–9.
200. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh HJ, Cuenca M, et al. Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. Cell. 2019;
201. Ibarbalz FM, Henry N, Brandão MC, Martini S, Busseni G, Byrne H, et al. Global Trends in Marine Plankton Diversity across Kingdoms of Life. Cell. 2019;
202. Guillou L, Alves-de-Souza C, Siano Dr R, González H. The ecological significance of small, eukaryotic parasites in marine ecosystems. Microbiology Today. 2010;
203. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, et al. The genome of the diatom Thalassiosira Pseudonana: Ecology, evolution, and metabolism. Science. 2004;
204. Vetukuri RR, Tripathy S, Malar C M, Panda A, Kushwaha SK, Chawade A, et al. Draft Genome Sequence for the Tree Pathogen Phytophthora plurivora. Genome biology and evolution. 2018;10(9):2432–42.
205. Hou Y, Lin S. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. PloS one. 2009 Sep 14;4(9):e6978.
206. Massana R, Guillou L, Terrado R, Forn I, Pedrós-Alió C. Growth of uncultured heterotrophic flagellates in unamended seawater incubations. Aquatic Microbial Ecology. 2006 Nov 24;45(2):171–80.

207. Arndt H, Dietrich D, Auer B, Cleven E josef J, Gräfenhan T, Weitere M, et al. Functional diversity of heterotrophic flagellates in aquatic ecosystems. The Flagellates: Unity, Diversity and Evolution. 2000;
208. Berlemont R, Martiny AC. Glycoside Hydrolases across Environmental Microbial Communities. PLoS computational biology. 2016;12(12):e1005300.
209. Kameshwar AKS, Qin W. Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. Mycology. 2018;9(2):93.
210. Wolfenden R, Lu X, Young G. Spontaneous Hydrolysis of Glycosides. Journal of the American Chemical Society. 1998 Jul;120(27):6814–5.
211. Turakainen H, Aho S, Korhola M. MEL gene polymorphism in the genus Saccharomyces. Applied and environmental microbiology. 1993 Aug 1;59(8):2622–30.
212. Martiny AC, Treseder K, Pusch G. Phylogenetic conservatism of functional traits in microorganisms. ISME Journal. 2013 Apr;7(4):830–8.
213. Groussin M, Gouy M. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. Molecular Biology and Evolution. 2011 Sep 1;28(9):2661–74.
214. Hu SK, Herrera EL, Smith AR, Pachiadaki MG, Edgcomb VP, Sylva SP, et al. Protistan grazing impacts microbial communities and carbon cycling at deep-sea hydrothermal vents. Proceedings of the National Academy of Sciences. 2021 Jul 20;118(29):e2102674118.
215. Jürgens K, Massana R. Protistan Grazing on Marine Bacterioplankton. In: Microbial Ecology of the Oceans [Internet]. John Wiley & Sons, Ltd; 2008 [cited 2022 Mar 24]. p. 383–441. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470281840.ch11
216. Fenchel T. The Ecology of Heterotrophic Microflagellates. In: Marshall KC, editor. Advances in Microbial Ecology [Internet]. Boston, MA: Springer US; 1986 [cited 2022 Jun 13]. p. 57–97. (Advances in Microbial Ecology). Available from: https://doi.org/10.1007/978-1-4757-0611-6_2
217. Jeuck A, Arndt H. A Short Guide to Common Heterotrophic Flagellates of Freshwater Habitats Based on the Morphology of Living Organisms. Protist. 2013 Nov 1;164(6):842–60.
218. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, et al. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. Journal of Eukaryotic Microbiology. 2019;66(1):4–119.
219. Schön ME, Zlatogursky VV, Singh RP, Poirier C, Wilken S, Mathur V, et al. Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. Nat Commun. 2021 Nov 17;12(1):6651.
220. Obiol A, Muhovic I, Massana R. Oceanic heterotrophic flagellates are dominated by a few widespread taxa. Limnology and Oceanography. 2021;66(12):4240–53.
221. Crawford DL, Oleksiak MF. Ecological population genomics in the marine environment. Briefings in Functional Genomics. 2016 Sep 1;15(5):342–51.
222. Zhao L, Qu F, Song N, Han Z, Gao T, Zhang Z. Population genomics provides insights into the population structure and temperature-driven adaptation of Collichthys lucidus. BMC Genomics. 2021 Oct 8;22(1):729.
223. Zhang BD, Li YL, Xue DX, Liu JX. Population Genomics Reveals Shallow Genetic Structure in a Connected and Ecologically Important Fish From the Northwestern Pacific Ocean. Frontiers in Marine Science [Internet]. 2020 [cited 2022 Jun 13];7. Available from: https://www.frontiersin.org/article/10.3389/fmars.2020.00374
224. Xu S, Yanagimoto T, Song N, Cai S, Gao T, Zhang X. Population genomics reveals possible genetic evidence for parallel evolution of Sebastiscus marmoratus in the northwestern Pacific Ocean. Open Biology. 2019;9(9):190028.
225. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, et al. Population Genomics of Early Events in the Ecological Differentiation of Bacteria. Science. 2012 Apr 6;336(6077):48–51.
226. Lin YC, Chin CP, Yang JW, Chiang KP, Hsieh C hao, Gong GC, et al. How Communities of Marine Stramenopiles Varied with Environmental and Biological Variables in the Subtropical Northwestern Pacific Ocean. Microb Ecol. 2022 May;83(4):916–28.
227. Monier A, Terrado R, Thaler M, Comeau A, Medrinal E, Lovejoy C. Upper Arctic Ocean water masses harbor distinct communities of heterotrophic flagellates. Biogeosciences. 2013 Jun 27;10(6):4273–86.
228. Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A, et al. Phylogenetic and ecological analysis of novel marine stramenopiles. Applied and Environmental Microbiology. 2004 Jun;70(6):3528–34.
229. Simon M, Jardillier L, Deschamps P, Moreira D, Restoux G, Bertolino P, et al. Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. Environ Microbiol. 2015 Oct;17(10):3610–27.
230. Gómez F, Moreira D, Benzerara K, López-García P. Solenicola setigera is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. Environ Microbiol. 2011 Jan;13(1):193–202.
231. Giner CR, Forn I, Romac S, Logares R, de Vargas C, Massana R. Environmental Sequencing Provides Reasonable Estimates of the Relative Abundance of Specific Picoeukaryotes. Appl Environ Microbiol. 2016 Aug 1;82(15):4757–66.
232. Gutleben J, Chaib De Mares M, van Elsas JD, Smidt H, Overmann J, Sipkema D. The multi-omics promise in context: from sequence to microbial isolate. Critical Reviews in Microbiology. 2018 Mar 4;44(2):212–29.
233. Huang L, Zhang H, Wu P, Entwistle S, Li X, Yohe T, et al. DbCAN-seq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Research. 2018;
234. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research. 2009 Jan 1;37(Database):D233–8.
235. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format

and SAMtools. Bioinformatics (Oxford, England). 2009 Aug 15;25(16):2078–9.
236. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012 Jul 17 [cited 2017 Jun 20]; Available from: http://arxiv.org/abs/1207.3907
237. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012 Jun;6(2):80–92.
238. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution. 1986 Sep 1;3(5):418–26.
239. Morelli MJ, Wright CF, Knowles NJ, Juleff N, Paton DJ, King DP, et al. Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. Veterinary Research. 2013 Mar 1;44(1):1–15.
240. Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, et al. Patterns of Gene Flow Define Species of Thermophilic Archaea. PLoS Biol. 2012 Feb 21;10(2):e1001265.
241. Reno ML, Held NL, Fields CJ, Burke PV, Whitaker RJ. Biogeography of the Sulfolobus islandicus pan-genome. Proceedings of the National Academy of Sciences. 2009 May 26;106(21):8605–10.
242. Torrado H, Carreras C, Raventos N, Macpherson E, Pascual M. Individual-based population genomics reveal different drivers of adaptation in sympatric fish. Sci Rep. 2020 Jul 29;10(1):12683.
243. Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. Science. 2006 Mar 24;311(5768):1737–40.
244. Kryazhimskiy S, Plotkin JB. The Population Genetics of dN/dS. PLoS Genet. 2008 Dec 12;4(12):e1000304.
245. Ting L, Williams TJ, Cowley MJ, Lauro FM, Guilhaus M, Raftery MJ, et al. Cold adaptation in the marine bacterium, Sphingopyxis alaskensis, assessed using quantitative proteomics. Environmental Microbiology. 2010;12(10):2658–76.
246. Zhang Z, Miteva MA, Wang L, Alexov E. Analyzing Effects of Naturally Occurring Missense Mutations. Comput Math Methods Med. 2012;2012:805827.
247. Scacheri CA, Scacheri PC. Mutations in the non-coding genome. Curr Opin Pediatr. 2015 Dec;27(6):659–64.
248. Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. Nature. 2022 Jun 8;1–7.
249. Sharp N. Mutations matter even if proteins stay the same. Nature [Internet]. 2022 Jun 8 [cited 2022 Jun 13]; Available from: https://www.nature.com/articles/d41586-022-01091-6
250. Smith WM, Pham TH, Lei L, Dou J, Soomro AH, Beatson SA, et al. Heat Resistance and Salt Hypersensitivity in Lactococcus lactis Due to Spontaneous Mutation of llmg_1816 (gdpP) Induced by High-Temperature Growth. Appl Environ Microbiol. 2012 Nov;78(21):7753–9.
251. Almeida-Dalmet S, Litchfield CD, Gillevet P, Baxter BK. Differential Gene Expression in Response to Salinity and Temperature in a Haloarcula Strain from Great Salt Lake, Utah. Genes. 2018 Jan;9(1):52.
252. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. Proc Natl Acad Sci U S A. 1998 Jun 9;95(12):6578–83.
253. Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proceedings of the National Academy of Sciences. 2018 Jun 19;115(25):6506–11.
254. Moran MA. The global ocean microbiome. Science. 2015 Dec 11;350(6266):aac8455.
255. Liu J, Meng Z, Liu X, Zhang XH. Microbial assembly, interaction, functioning, activity and diversification: a review derived from community compositional data. Mar Life Sci Technol. 2019 Nov 1;1(1):112–28.
256. Overmann J, Lepleux C. Marine Bacteria and Archaea: Diversity, Adaptations, and Culturability. In: Stal LJ, Cretoiu MS, editors. The Marine Microbiome: An Untapped Source of Biodiversity and Biotechnological Potential [Internet]. Cham: Springer International Publishing; 2016 [cited 2022 Jul 5]. p. 21–55. Available from: https://doi.org/10.1007/978-3-319-33000-6_2
257. Salazar G, Sunagawa S. Marine microbial diversity. Current Biology. 2017 Jun 5;27(11):R489–94.
258. Oleksiak M, Rajora O. Population Genomics: Marine Organisms. 2020.
259. Raes EJ, Bodrossy L, van de Kamp J, Bissett A, Ostrowski M, Brown MV, et al. Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. Proc Natl Acad Sci U S A. 2018 Aug 28;115(35):E8266–75.
260. Hernando-Morales V, Ameneiro J, Teira E. Water mass mixing shapes bacterial biogeography in a highly hydrodynamic region of the Southern Ocean. Environ Microbiol. 2017 Mar;19(3):1017–29.
261. Zorz J, Willis C, Comeau AM, Langille MGI, Johnson CL, Li WKW, et al. Drivers of Regional Bacterial Community Structure and Diversity in the Northwest Atlantic Ocean. Front Microbiol. 2019 Feb 21;10:281.
262. Maturana-Martínez C, Iriarte JL, Ha SY, Lee B, Ahn IY, Vernet M, et al. Biogeography of Southern Ocean Active Prokaryotic Communities Over a Large Spatial Scale. Frontiers in Microbiology [Internet]. 2022 [cited 2022 Jul 5];13. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2022.862812
263. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol. 2012 May 14;10(7):497–506.
264. Seo JH, Kang I, Yang SJ, Cho JC. Characterization of spatial distribution of the bacterial community in the South Sea of Korea. PLoS One. 2017 Mar 17;12(3):e0174159.
265. Doblin MA, van Sebille E. Drift in ocean currents impacts intergenerational microbial exposure to temperature. Proc Natl Acad Sci U S A. 2016 May 17;113(20):5700–5.
266. O'Donnell DR, Hamman CR, Johnson EC, Kremer CT, Klausmeier CA, Litchman E. Rapid thermal adaptation in a marine diatom reveals constraints and trade-offs. Global Change Biology. 2018;24(10):4554–65.

267. Schaum CE, Rost B, Collins S. Environmental stability affects phenotypic evolution in a globally distributed marine picoplankton. ISME J. 2016 Jan;10(1):75–84.
268. Hellweger FL, van Sebille E, Fredrick ND. Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. Science. 2014 Sep 12;345(6202):1346–9.
269. Joint I, Mühling M, Querellou J. Culturing marine bacteria – an essential prerequisite for biodiscovery. Microb Biotechnol. 2010 Sep;3(5):564–75.
270. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021 Apr;39(4):499–509.
271. Yoshitake K, Kimura G, Sakami T, Watanabe T, Taniuchi Y, Kakehi S, et al. Development of a time-series shotgun metagenomics database for monitoring microbial communities at the Pacific coast of Japan. Sci Rep. 2021 Jun 9;11(1):12222.
272. Paoli L, Ruscheweyh HJ, Forneris CC, Hubrich F, Kautsar S, Bhushan A, et al. Biosynthetic potential of the global ocean microbiome. Nature. 2022 Jul;607(7917):111–8.
273. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017 Aug;35(8):725–31.
274. Yang C, Chowdhury D, Zhang Z, Cheung WK, Lu A, Bian Z, et al. A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. Computational and Structural Biotechnology Journal. 2021 Jan 1;19:6301–14.
275. Coe A, Biller SJ, Thomas E, Boulias K, Bliem C, Arellano A, et al. Coping with darkness: The adaptive response of marine picocyanobacteria to repeated light energy deprivation. Limnology and Oceanography. 2021;66(9):3300–12.
276. Yan W, Feng X, Lin TH, Huang X, Xie L, Wei S, et al. Diverse Subclade Differentiation Attributed to the Ubiquity of Prochlorococcus High-Light-Adapted Clade II. mBio. 2022 Mar 14;13(2):e03027-21.
277. Kent AG, Baer SE, Mouginot C, Huang JS, Larkin AA, Lomas MW, et al. Parallel phylogeography of Prochlorococcus and Synechococcus. ISME J. 2019 Feb;13(2):430–41.
278. Kashtan N, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R, Chisholm SW. Fundamental differences in diversity and genomic population structure between Atlantic and Pacific Prochlorococcus. ISME J. 2017 Sep;11(9):1997–2011.
279. López-Pérez M, Haro-Moreno JM, Coutinho FH, Martinez-Garcia M, Rodriguez-Valera F. The Evolutionary Success of the Marine Bacterium SAR11 Analyzed through a Metagenomic Perspective. mSystems. 2020 Oct 6;5(5):e00605-20.
280. Roda-Garcia JJ, Haro-Moreno JM, Huschet LA, Rodriguez-Valera F, López-Pérez M. Phylogenomics of SAR116 Clade Reveals Two Subclades with Different Evolutionary Trajectories and an Important Role in the Ocean Sulfur Cycle. mSystems. 2021 Oct 26;6(5):e0094421.
281. Lambert S, Lozano JC, Bouget FY, Galand PE. Seasonal marine microorganisms change neighbours under contrasting environmental conditions. Environmental Microbiology. 2021;23(5):2592–604.
282. Pereira O, Hochart C, Boeuf D, Auguet JC, Debroas D, Galand PE. Seasonality of archaeal proteorhodopsin and associated Marine Group IIb ecotypes (Ca. Poseidoniales) in the North Western Mediterranean Sea. ISME J. 2021 May;15(5):1302–16.
283. Galand PE, Pereira O, Hochart C, Auguet JC, Debroas D. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. ISME J. 2018 Oct;12(10):2470–8.
284. Lambert S, Tragin M, Lozano JC, Ghiglione JF, Vaulot D, Bouget FY, et al. Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. ISME J. 2019 Feb;13(2):388–401.
285. Karsenti E, Acinas SG, Bork P, Bowler C, Vargas CD, Raes J, et al. A Holistic Approach to Marine Eco-Systems Biology. PLOS Biology. 2011 Oct 18;9(10):e1001177.
286. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10–2.
287. Benoit G, Peterlongo P, Mariadassou M, Drezen E, Schbath S, Lavenier D, et al. Multiple comparative metagenomics using multiset k-mer counting. PeerJ Comput Sci. 2016 Nov 14;2:e94.
288. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing; 2018. Available from: https://www.R-project.org
289. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674–6.
290. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015 Aug 27;3:e1165.
291. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014 Nov;11(11):1144–6.
292. Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome. 2014 Jan 1;2(1):26.
293. Uritskiy GV, Diruggiero J, Taylor J. MetaWRAP - A flexible pipeline for genome-resolved metagenomic data analysis 08 Information and Computing Sciences 0803 Computer Software 08 Information and Computing Sciences 0806 Information Systems. Microbiome. 2018 Sep 15;6(1):1–13.
294. Tully BJ, Sachdeva R, Graham ED, Heidelberg JF. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. 2017;
295. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the

genome taxonomy database. Bioinformatics. 2020 Mar 1;36(6):1925–7.
296. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017 Dec;11(12):2864–8.
297. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.
298. Boyd JA. Development of meta omic tools to explore microbial carbon cycling. [Internet]. University of Queensland; 2020 [cited 2022 Aug 23]. Available from: https://espace.library.uq.edu.au/view/UQ:1754b2c
299. Ruf T. The Lomb-Scargle Periodogram in Biological Rhythm Research: Analysis of Incomplete and Unequally Spaced Time-Series. Biological Rhythm Research. 1999 Apr 1;30(2):178–201.
300. Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J Classif. 2014 Oct 1;31(3):274–95.
301. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-7 [Internet]. 2020. Available from: https://CRAN.R-project.org/package=vegan
302. Auladell A, Sánchez P, Sánchez O, Gasol JM, Ferrera I. Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. ISME J. 2019 Aug;13(8):1975–87.
303. Giner CR, Balagué V, Krabberød AK, Ferrera I, Reñé A, Garcés E, et al. Quantifying long-term recurrence in planktonic microbial eukaryotes. Molecular Ecology. 2019;28(5):923–35.
304. Auladell A, Barberán A, Logares R, Garcés E, Gasol JM, Ferrera I. Seasonal niche differentiation among closely related marine bacteria. ISME J. 2022 Jan;16(1):178–89.
305. Hewson I, Steele JA, Capone DG, Fuhrman JA. Temporal and spatial scales of variation in bacterioplankton assemblages of oligotrophic surface waters. Marine Ecology Progress Series. 2006 Apr 13;311:67–77.
306. Tolar BB, King GM, Hollibaugh JT. An Analysis of Thaumarchaeota Populations from the Northern Gulf of Mexico. Front Microbiol. 2013 Apr 9;4:72.
307. Boeuf D, Eppley JM, Mende DR, Malmstrom RR, Woyke T, DeLong EF. Metapangenomics reveals depth-dependent shifts in metabolic potential for the ubiquitous marine bacterial SAR324 lineage. Microbiome. 2021 Aug 13;9(1):172.
308. Gómez-Pereira PR, Fuchs BM, Alonso C, Oliver MJ, van Beusekom JEE, Amann R. Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. ISME J. 2010 Apr;4(4):472–87.
309. López-Pérez M, Haro-Moreno JM, Iranzo J, Rodriguez-Valera F. Genomes of the "Candidatus Actinomarinales" Order: Highly Streamlined Marine Epipelagic Actinobacteria. mSystems. 2020 Dec 15;5(6):e01041-20.
310. Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M, et al. Genomic and functional adaptation in surface ocean planktonic prokaryotes. Nature. 2010 Nov;468(7320):60–6.
311. Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure. 1995 Sep 15;3(9):853–9.
312. Sanders C, Turkarslan S, Lee DW, Daldal F. Cytochrome c biogenesis: the Ccm system. Trends Microbiol. 2010 Jun;18(6):266–74.
313. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences. 2016 May 24;113(21):5970–5.
314. Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol. 2012 Jul 16;10(8):538–50.
315. Dunne JA, Lafferty KD, Dobson AP, Hechinger RF, Kuris AM, Martinez ND, et al. Parasites Affect Food Web Structure Primarily through Increased Diversity and Complexity. PLOS Biology. 2013 Jun 11;11(6):e1001579.
316. Raes J, Bork P. Molecular eco-systems biology: towards an understanding of community function. Nat Rev Microbiol. 2008 Sep;6(9):693–9.
317. Legrand C, Rengefors K, Fistarol GO, Granéli E. Allelopathy in phytoplankton - biochemical, ecological and evolutionary aspects. Phycologia. 2003 Jul 1;42(4):406–19.
318. Sañudo-Wilhelmy SA, Gómez-Consarnau L, Suffridge C, Webb EA. The Role of B Vitamins in Marine Biogeochemistry. Annual Review of Marine Science. 2014;6(1):339–67.
319. Fredrickson AG, Stephanopoulos G. Microbial competition. Science. 1981 Aug 28;213(4511):972–9.
320. Chambouvet A, Morin P, Marie D, Guillou L. Control of toxic marine dinoflagellate blooms by serial parasitic killers. Science. 2008 Nov 21;322(5905):1254–7.
321. Berney C, Romac S, Mahé F, Santini S, Siano R, Bass D. Vampires in the oceans: predatory cercozoan amoebae in marine habitats. ISME J. 2013 Dec;7(12):2387–99.
322. Jolley ET, Jones AK. interaction between Navicula muralis Grunow and an associated species of Flavobacterium. British phycological journal [Internet]. 1977 [cited 2022 Aug 24]; Available from: https://scholar.google.com/scholar_lookup?title=interaction+between+Navicula+muralis+Grunow+and+an+associated+species+of+Flavobacterium&author=Jolley%2C+E.T.&publication_year=1977
323. Miller TR, Belas R. Dimethylsulfoniopropionate metabolism by Pfiesteria-associated Roseobacter spp. Appl Environ Microbiol. 2004 Jun;70(6):3383–91.
324. Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, et al. Single-Cell Genomics Reveals Organismal Interactions in Uncultivated Marine Protists. Science. 2011 May 6;332(6030):714–7.
325. Karlicki M, Antonowicz S, Karnkowska A. Tiara: Deep learning-based classification system for eukaryotic sequences. Bioinformatics. 2021 Sep 27;btab672.
326. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010 Mar 8;11(1):1–11.
327. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized

bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology. 2018 Nov 1;36(10):996.
328. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research. 2022 Jan 7;50(D1):D785–94.
329. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017 Nov;35(11):1026–8.
330. Levy Karin E, Mirdita M, Söding J. MetaEuk-sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. Microbiome. 2020 Apr 3;8(1):1–15.
331. Steinegger M, Mirdita M, Söding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. Nat Methods. 2019 Jul;16(7):603–6.
332. Mirdita M, Von Den Driesch L, Galiez C, Martin MJ, Soding J, Steinegger M. Uniclust databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Research. 2017 Jan 1;45(D1):D170–6.
333. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biology. 2014;12(6):e1001889.
334. Johnson LK, Alexander H, Brown CT. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. Gigascience. 2019 Apr 1;8(4):giy158.
335. Richter DJ, Berney C, Strassert JFH, Poh YP, Herman EK, Muñoz-Gómez SA, et al. EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes [Internet]. bioRxiv; 2022 [cited 2022 Sep 9]. p. 2020.06.30.180687. Available from: https://www.biorxiv.org/content/10.1101/2020.06.30.180687v4
336. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. Proceedings of the International AAAI Conference on Web and Social Media. 2009 Mar 19;3(1):361–2.
337. Deutschmann IM, Delage E, Giner CR, Sebastian M, Poulain J, Aristegui J, et al. Disentangling marine microbial networks across space [Internet]. bioRxiv; 2022 [cited 2022 Sep 5]. p. 2021.07.12.451729. Available from: https://www.biorxiv.org/content/10.1101/2021.07.12.451729v2
338. Wagner M, Horn M. The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr Opin Biotechnol. 2006 Jun;17(3):241–9.
339. Orellana LH, Francis TB, Ferraro M, Hehemann JH, Fuchs BM, Amann RI. Verrucomicrobiota are specialist consumers of sulfated methyl pentoses during diatom blooms. ISME J. 2022 Mar;16(3):630–41.
340. Sakai T, Ishizuka K, Kato I. Isolation and characterization of a fucoidan-degrading marine bacterium. Mar Biotechnol (NY). 2003 Oct;5(5):409–16.
341. Fuerst JA, Gwilliam HG, Lindsay M, Lichanska A, Belcher C, Vickers JE, et al. Isolation and molecular identification of planctomycete bacteria from postlarvae of the giant tiger prawn, Penaeus monodon. Appl Environ Microbiol. 1997 Jan;63(1):254–62.
342. Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, et al. Molecular evidence for a uniform microbial community in sponges from different oceans. Appl Environ Microbiol. 2002 Sep;68(9):4431–40.
343. Abby SS, Rocha EPC. The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems. PLOS Genetics. 2012 Sep 27;8(9):e1002983.
344. Pallen MJ, Beatson SA, Bailey CM. Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perspective. FEMS Microbiol Rev. 2005 Apr;29(2):201–29.
345. Jasti S, Sieracki ME, Poulton NJ, Giewat MW, Rooney-Varga JN. Phylogenetic diversity and specificity of bacteria closely associated with Alexandrium spp. and other phytoplankton. Appl Environ Microbiol. 2005 Jul;71(7):3483–94.
346. Sapp M, Schwaderer AS, Wiltshire KH, Hoppe HG, Gerdts G, Wichels A. Species-specific bacterial communities in the phycosphere of microalgae? Microb Ecol. 2007 May;53(4):683–99.
347. Aaronson SY 1974. The Biology and Ultrastructure of Phagotrophy in Ochromonas danica (Chrysophyceae: Chrysomdnadida). Microbiology. 83(1):21–9.
348. Van Donk E, Cerbin S, Wilken S, Helmsing NR, Ptacnik R, Verschoor AM. The effect of a mixotrophic chrysophyte on toxic and colony-forming cyanobacteria. Freshwater Biology. 2009;54(9):1843–55.
349. Chrzanowski TH, Šimek K. Prey-size selection by freshwater flagellated protozoa. Limnology and Oceanography. 1990;35(7):1429–36.
350. Boenigk J, Matz C, Jürgens K, Arndt H. The Influence of Preculture Conditions and Food Quality on the Ingestion and Digestion Process of Three Species of Heterotrophic Nanoflagellates. Microb Ecol. 2001 Aug;42(2):168–76.
351. Grossmann L, Bock C, Schweikert M, Boenigk J. Small but Manifold - Hidden Diversity in “Spumella-like Flagellates.” J Eukaryot Microbiol. 2016 Jul;63(4):419–39.
352. del Campo J, Not F, Forn I, Sieracki ME, Massana R. Taming the smallest predators of the oceans. ISME J. 2013 Feb;7(2):351–8.
353. Hess S, Suthaus A, Melkonian M. “Candidatus Finniella” (Rickettsiales, Alphaproteobacteria), Novel Endosymbionts of Viridiraptorid Amoeboflagellates (Cercozoa, Rhizaria). Appl Environ Microbiol. 2016 Jan 15;82(2):659–70.
354. Suter EA, Pachiadaki M, Taylor GT, Edgcomb VP. Eukaryotic Parasites Are Integral to a Productive Microbial Food Web in Oxygen-Depleted Waters. Frontiers in Microbiology [Internet]. 2022 [cited 2022 Aug 30];12. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.764605
355. Decelle J, Probert I, Bittner L, Desdevises Y, Colin S, de Vargas C, et al. An original mode of symbiosis in

open ocean plankton. Proc Natl Acad Sci U S A. 2012 Oct 30;109(44):18000–5.
356. Nishitani G, Nagai S, Hayakawa S, Kosaka Y, Sakurada K, Kamiyama T, et al. Multiple plastids collected by the dinoflagellate Dinophysis mitra through kleptoplastidy. Appl Environ Microbiol. 2012 Feb;78(3):813–21.
357. Goldman JC, Caron DA. Experimental studies on an omnivorous microflagellate: implications for grazing and nutrient regeneration in the marine microbial food chain. Deep Sea Research Part A Oceanographic Research Papers. 1985 Aug 1;32(8):899–915.
358. Boraas ME, Estep KW, Johnson PW, Sieburth JMcN. Phagotrophic Phototrophs: The Ecological Significance of Mixotrophy1,2. The Journal of Protozoology. 1988;35(2):249–52.
359. Sidore AM, Lan F, Lim SW, Abate AR. Enhanced sequencing coverage with digital droplet multiple displacement amplification. Nucleic Acids Res. 2016 Apr 20;44(7):e66.
360. Brooks AN, Turkarslan S, Beer KD, Lo FY, Baliga NS. Adaptation of cells to new environments. Wiley Interdiscip Rev Syst Biol Med. 2011 Sep;3(5):544–61.
361. Hugerth LW, Larsson J, Alneberg J, Lindh MV, Legrand C, Pinhassi J, et al. Metagenome-assembled genomes uncover a global brackish microbiome. Genome Biology. 2015 Dec 14;16(1):279.
362. Buck KR, Bentham WN. A novel symbiosis between a cyanobacterium, Synechococcus sp., an aplastidic protist, Solenicola setigera, and a diatom, Leptocylindrus mediterraneus, in the open ocean. Marine Biology. 1998 Oct 1;132(3):349–55.
363. Ghazanfar S, Lin Y, Su X, Lin DM, Patrick E, Han ZG, et al. Investigating higher-order interactions in single-cell data with scHOT. Nat Methods. 2020 Aug;17(8):799–806.
364. Chaffron S, Delage E, Budinich M, Vintache D, Henry N, Nef C, et al. Environmental vulnerability of the global ocean epipelagic plankton community interactome. Science Advances. 2021;7(35):eabg1921.
365. Menge BA. Indirect Effects in Marine Rocky Intertidal Interaction Webs: Patterns and Importance. Ecological Monographs. 1995;65(1):21–74.
366. Ghandi M, Beer MA. Group Normalization for Genomic Data. PLOS ONE. 2012 Aug 13;7(8):e38695.
367. Larsson AJM, Stanley G, Sinha R, Weissman IL, Sandberg R. Computational correction of index switching in multiplexed sequencing libraries. Nat Methods. 2018 May;15(5):305–7.
368. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nature Microbiology. 2021 Jan 1;6(1):3–6.
369. Delmont TO, Eren AM. Identifying contamination with advanced visualization and analysis practices: metagenomic approaches for eukaryotic genome assemblies. PeerJ. 2016 Jan 1;4:e1839.
370. Fierst JL, Murdock DA. Decontaminating eukaryotic genome assemblies with machine learning. BMC Bioinformatics. 2017 Dec 1;18(1):533.

ANNEX A – SUPPLEMENTARY MATERIAL FOR CHAPTER 1

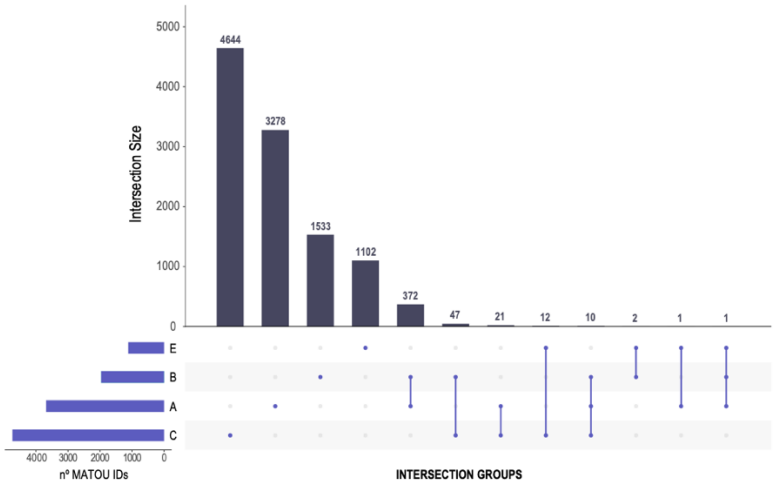

Figure 1. Number of Unigenes (i.e., representative genes after clustering genes at 95% identity) from the MATOU database found in MAST-4 and the number of genes shared by the four species. Note that the different groups are ordered by group size and that the biggest groups are those including only one MAST-4 species, followed by the groups constituted by the combination of two or more species.

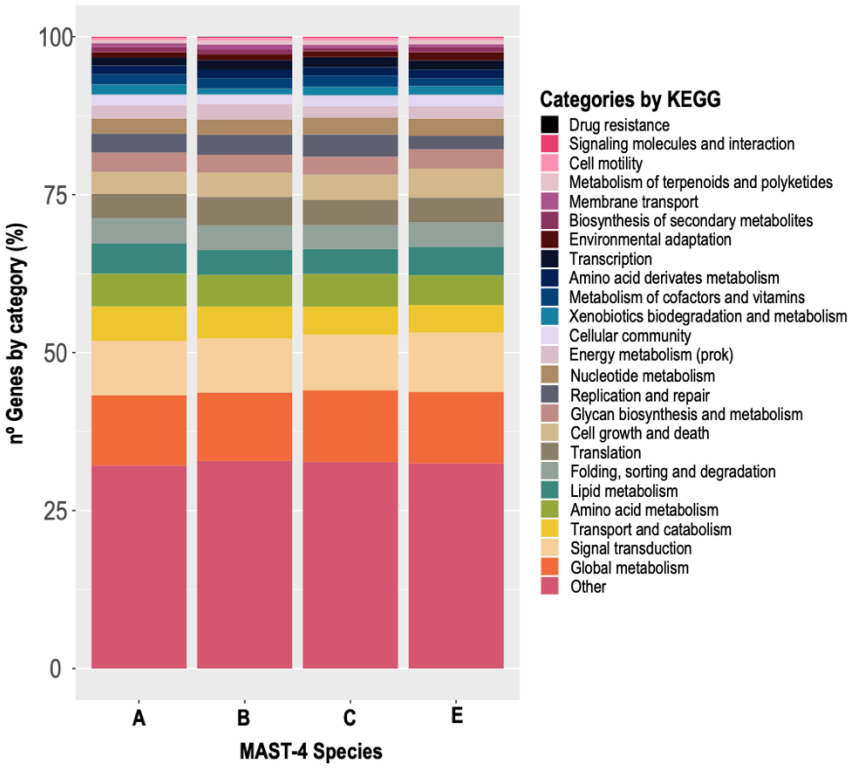

Figure 2. Functional profile of MAST-4 genes according to KEGG. KEGG annotations are indicated as percentage of genes falling into functional categories. The category "Other" is an artificial grouping including all the annotations belonging to human related pathways such as 'Alzheimer' or 'Influenza A'.

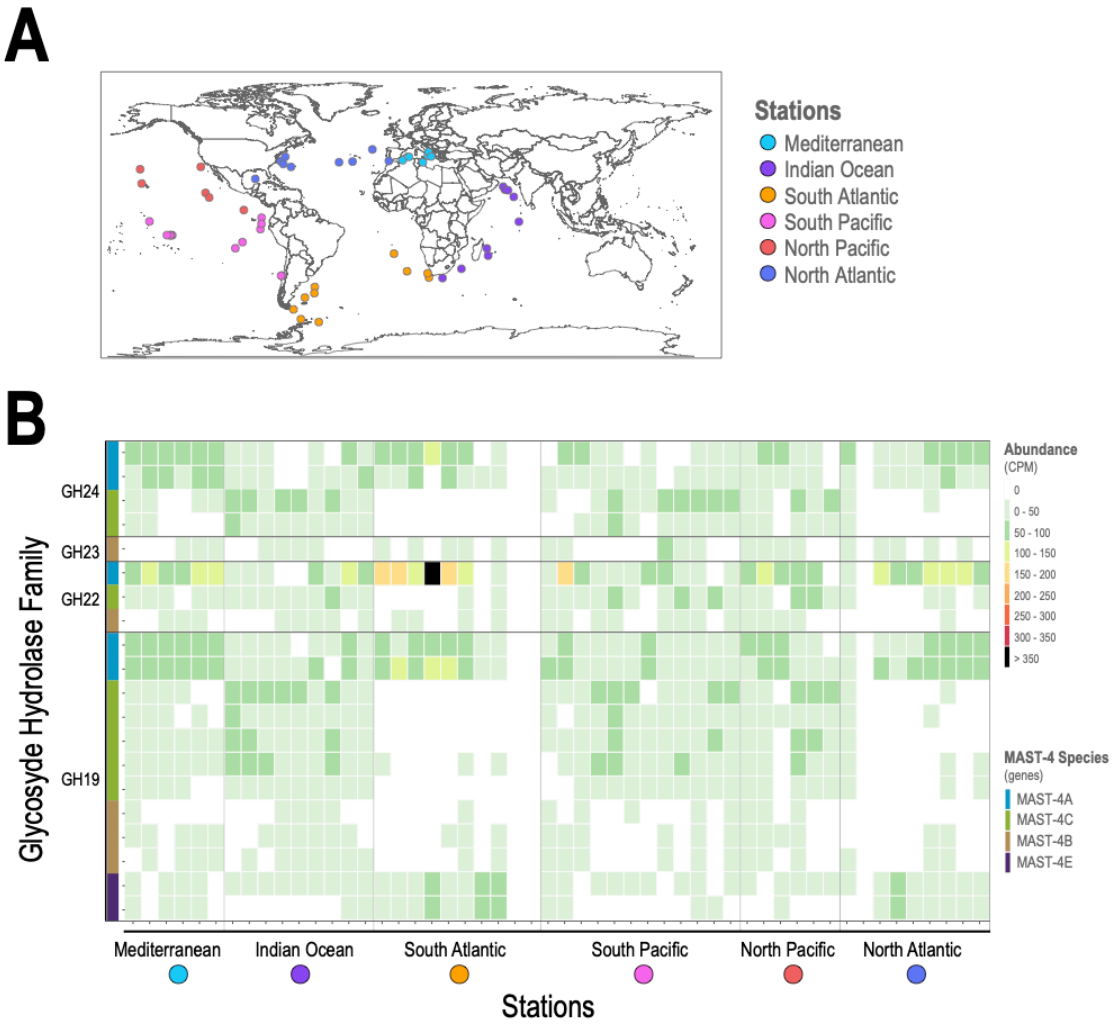

Figure 3. Abundance of GH genes in MAST-4A/B/C/E. Panel A) Geographic location of metagenomic samples of Tara Oceans. **Panel B)** Heatmap of the Glycoside Hydrolase family abundances in MAST-4 (see their expression in **Figure 1.5C**). Samples are in the x-axis grouped by the ocean region and ordered following the expedition's trajectory. Genes in the y-axis are organized by family and each species is indicated with a color. GH22, GH23 and GH24 are families of lysozymes and GH19 is a family of chitinases that can also act as lysozymes in some organisms.

Table 1. Summary of each SAG's environmental data from the *TARA Oceans* expedition. Legend: Sample Depths: D - DCM, S - SUR; Platform: GS - National Sequencing Center of Genoscope; OR - Oregon Health & Science University.

SAG ID	Species	Station	Location	Sample Depth	Sequencing Date (dd/mm/year)	Platform	Sequencing Depth (Gb)	BUSCO completeness (%)	Assembly Size (Mbp)	N50	GC content (%)	ENA ID
AA538_M19	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	6.9	29.8	8.95	12,558	33.04	SAMEA3692426
AA538_N22	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	8.4	23.4	7.54	10,215	32.63	SAMEA3692427
AA538_F10	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	6	29.8	9.55	10,961	32.5	SAMEA3692431
AA538_G04	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.7	27.4	8.38	9,809	32.53	SAMEA3692428
AA538_G20	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.6	27.1	9.17	13,842	32.72	SAMEA3692429

AA538_K07	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4	4.7	1.64	5,239	36.15	SAMEA3692430
AA538_E21	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	5.7	49.9	20.31	12,807	32.69	SAMEA4557820
AA538_E15	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	6.4	32	9.44	11,361	32.9	SAMEA3663802
AA538_C11	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (OR)	2.7	40.6	15.66	10,281	32.6	SAMEA3663804
AB537_A17	MAST-4A	41	Indian Ocean	D	1/1/13	HiSeq2000 (OR)	4	25.1	7.19	11,907	32.66	SAMEA3663803
AA538_J18	MAST-4A	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	3.2	21.7	6.98	7,576	32.39	SAMEA4557821
AA538_E19	MAST-4A	23	Adriatic Sea	D	10/3/14	HiSeq2000 (OR)	2.4	26.1	8.66	8,957	32.49	SAMEA4557817
AA538_G20_2	MAST-4A	23	Adriatic Sea	D	10/3/14	HiSeq2000 (OR)	6.8	30.7	10.24	10,324	32.81	SAMEA4557819
AB537_K04	MAST-4A	41	Indian Ocean	D	10/3/14	HiSeq2000 (OR)	3.5	16.9	4.45	18,740	32.53	SAMEA4557818
AA539_A11	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	10.5	33.7	15.04	8,199	33.66	SAMEA7773355
AA539_C15	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	8.8	45.5	17.9	9,238	33.49	SAMEA7773356
AA539_D06	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	3.6	17.8	5.07	7,682	32.77	SAMEA7773357
AA539_E05	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	3.5	18.5	5.76	10,763	32.93	SAMEA7773358
AA539_I04	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	8.5	11.2	3.74	5,764	35.92	SAMEA7773359
AA539_L09	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	8	9.9	3.59	5,184	34.93	SAMEA7773360
AA539_N11	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	7.3	17.4	6.05	6,678	34.19	SAMEA7773361
AA539_O23	MAST-4A	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	7.1	11.6	3.75	6,130	35.48	SAMEA7773362
AB242_E18	MAST-4A	51	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	7.1	35	11.8	12,694	33.64	SAMEA7773363
AB240_N06	MAST-4B	41	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	6.2	42.6	12.81	8,850	34.56	SAMEA7773364
AB240_P13	MAST-4B	41	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	8	25.5	9.66	6,742	34.8	SAMEA7773365
AB535_I05	MAST-4B	46	Adriatic Sea	S	5/2/16	HiSeq4000 (GS)	8.7	29.1	9.97	7,711	34.6	SAMEA7773366
AB206_K13	MAST-4B	47	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	8	37.4	11.75	8,792	34.52	SAMEA7773367
AB208_E03	MAST-4B	48	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	7.1	21.2	6.86	8,131	35.07	SAMEA7773368
AB209_C10	MAST-4B	48	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	8.6	10.3	2.72	5,268	36.27	SAMEA7773369
AB209_D14	MAST-4B	48	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	9.7	14.5	4.93	6,590	36.23	SAMEA7773370
AB209_G07	MAST-4B	48	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	2.7	22.1	7.81	6,881	34.87	SAMEA7773371
AB242_M03	MAST-4B	51	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	8.4	30.1	12.55	7,827	34.97	SAMEA7773372
AB536_E17	MAST-4C	41	Indian Ocean	D	1/1/13	HiSeq2000 (GS)	6.1	33.7	8.33	18,733	40.15	SAMEA3692437
AB536_F22	MAST-4C	41	Indian Ocean	D	1/1/13	HiSeq2000 (GS)	5.5	33.3	9.67	17,256	40.59	SAMEA3692438

AB536_J08	MAST-4C	41	Indian Ocean	D	1/1/13	HiSeq2000 (GS)	5.1	24.8	7.33	15,429	40.52	SAMEA3692439
AB536_M21	MAST-4C	41	Indian Ocean	D	1/1/13	HiSeq2000 (GS)	4.7	17.1	4.69	14,717	40.33	SAMEA3692440
AB197_D11	MAST-4C	39	Adriatic Sea	S	5/2/16	HiSeq4000 (GS)	10.1	27.7	8.18	10,535	40.37	SAMEA7773373
AB197_D19	MAST-4C	39	Adriatic Sea	S	5/2/16	HiSeq4000 (GS)	7.2	27.7	7.98	10,443	40.42	SAMEA7773374
AB240_A08	MAST-4C	41	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	4.4	43.5	13.48	15,688	40.55	SAMEA7773375
AB240_K06	MAST-4C	41	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	6.6	30	9.12	12,577	40.47	SAMEA7773376
AB241_N04	MAST-4C	41	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	6.6	42.9	12.62	14,165	40.38	SAMEA7773377
AB537_D14	MAST-4C	41	Indian Ocean	D	5/2/16	HiSeq4000 (GS)	6.4	22.1	6.06	11,641	40.59	SAMEA7773378
AB537_G02	MAST-4C	41	Indian Ocean	D	5/2/16	HiSeq4000 (GS)	8.9	54.8	20.09	14,366	40.75	SAMEA7773379
AB537_J09	MAST-4C	41	Indian Ocean	D	5/2/16	HiSeq4000 (GS)	6.5	41.3	13.27	16,943	40.71	SAMEA7773380
AB537_K09	MAST-4C	41	Indian Ocean	D	5/2/16	HiSeq4000 (GS)	8.3	39.3	12.16	14,100	40.47	SAMEA7773381
AB537_L03	MAST-4C	41	Indian Ocean	D	5/2/16	HiSeq4000 (GS)	6.1	43.9	15.42	17,397	40.68	SAMEA7773382
AB538_D10	MAST-4C	41	Indian Ocean	D	5/2/16	HiSeq4000 (GS)	5.8	7.6	2.34	6,077	40.59	SAMEA7773383
AB538_O04	MAST-4C	41	Indian Ocean	D	5/2/16	HiSeq4000 (GS)	7	8.3	3.1	6,079	40.25	SAMEA7773384
AB535_A16	MAST-4C	46	Maldives	S	5/2/16	HiSeq4000 (GS)	5.2	31.1	8.95	9,468	40.57	SAMEA7773385
AB535_N14	MAST-4C	46	Maldives	S	5/2/16	HiSeq4000 (GS)	6	13.8	3.46	8,047	39.97	SAMEA7773386
AB535_P02	MAST-4C	46	Maldives	S	5/2/16	HiSeq4000 (GS)	5.9	38.6	14.26	11,347	40.78	SAMEA7773387
AB242_E07	MAST-4C	31	Indian Ocean	S	5/2/16	HiSeq4000 (GS)	8.1	26.1	7.73	8,564	40.6	SAMEA7773388
AA538_A02	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.5	23.4	7.27	11,624	44.28	SAMEA3692436
AA538_A03	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.5	21.4	7.84	12,985	44.17	SAMEA3692416
AA538_C05	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.6	24	6.95	12,026	44.58	SAMEA3692417
AA538_F08	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4	25.4	6.55	10,536	44.54	SAMEA3692418
AA538_J09	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.7	23.8	7.27	11,885	44.39	SAMEA3692419
AA538_A11	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	6.8	28.8	9.68	11,474	44.49	SAMEA3692432
AA538_L23	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.4	10.6	2.97	7,547	43.74	SAMEA3692433
AA538_M11	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.2	9.9	2.32	8,900	42.88	SAMEA3692434
AA538_N16	MAST-4E	23	Adriatic Sea	D	1/1/13	HiSeq2000 (GS)	4.9	19.2	4.69	8,128	43.79	SAMEA3692435
AA539_C21	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	5.6	8.3	2.43	6,091	43.08	SAMEA7773389
AA539_D08	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	7.6	9.9	2.44	5,844	43.23	SAMEA7773390

AA539_F21	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	6.4	25.1	7.41	8,254	43.83	SAMEA7773391
AA539_I19	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	5.3	8.5	2.73	6,687	43.98	SAMEA7773392
AA539_L11	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	4.6	19.4	6.88	9,398	43.9	SAMEA7773393
AA539_M19	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	4.1	29.7	12.59	8,319	43.83	SAMEA7773394
AA539_N05	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	4.7	20.1	7.61	9,287	44.37	SAMEA7773395
AA539_O18	MAST-4E	23	Adriatic Sea	D	5/2/16	HiSeq4000 (GS)	6.4	12.9	3.43	5,661	42.55	SAMEA7773396

Table 2. Basic assembly statistics from QUAST, BUSCO and AUGUSTUS for all the co-assemblies before and after the cleaning pipeline. Legend: Norm and Non-norm indicate whether or not the raw reads were normalized using BBNORM prior to the co-assembly.

	MAST-4A Norm.	**MAST-4A Non-norm.**	**MAST-4B Norm.**	**MAST-4B Non-norm.**	**MAST-4C Norm.**	**MAST-4C Non-norm.**	**MAST-4E Norm.**	**MAST-4E Non-norm.**
Total # SAGs	23	23	9	9	20	20	17	17
Before Cleaning								
Assembly size (Mb)	63.1	72.6	38.7	46.1	62	71.1	40	45.2
# contigs	11,382	15,253	5,548	8,634	8,948	13,225	4,077	5,815
GC content	33.8	33.77	35.75	35.5	41.28	40.59	44.81	44
N50	10,024	8,336	14,446	10,442	13,793	10,223	21,539	17,495
Completeness % (BUSCO)	90.1	92.1	79.6	83.8	91.4	91.8	82.5	84.8
After Cleaning								
Assembly size (Mb)	47.4	48	29	30.5	47.8	47.4	30.7	33.3
# contigs	4,787	5,198	2,282	2,792	3,953	4,531	1,739	2,163
GC content	33.13	33.13	34.67	34.73	41.17	41.14	45.68	45.77
N50	12,683	11,696	17,930	14,576	17,109	13,909	26,613	23,298
Completeness % (BUSCO)	80.5	81.2	66.7	71.6	83.5	80	70.7	73.3
# predicted genes	15,508	15,484	10,019	10,667	16,260	16,070	9,042	9,593

Table 3. BUSCO v3 proteins used to generate the multi-gene phylogeny of MAST-4 and from which contig they were retrieved in the co-assemblies.

BUSCO ID	**Contig MAST-4A**	**Contig MAST-4B**	**Contig MAST-4C**	**Contig MAST-4E**
EOG09370082	scaffold890_size14996	scaffold1032_size10703	scaffold198_size34289	scaffold1124_size10288
EOG0937017X	scaffold456_size20825	scaffold150_size31631	scaffold274_size30937	scaffold151_size41591
EOG093701IQ	scaffold476_size20367	scaffold993_size11015	scaffold499_size28111	scaffold580_size19217
EOG093701S0	scaffold1158_size12787	scaffold1162_size9409	scaffold459_size24190	scaffold548_size19878
EOG093701SQ	scaffold2211_size8311	scaffold63_size43378	scaffold2853_size6204	scaffold952_size12336
EOG093704Q0	scaffold3309_size5706	scaffold2298_size4325	scaffold316_size29145	scaffold37_size71445
EOG0937050C	scaffold49_size41508	scaffold589_size16261	scaffold2757_size6444	scaffold491_size21471
EOG093705DM	scaffold4113_size4453	scaffold689_size14745	scaffold3549_size4712	scaffold2_size138925

EOG093705E5	scaffold3949_size4657	scaffold2652_size3487	scaffold1037_size15433	scaffold333_size27322
EOG093705EY	scaffold3913_size4702	scaffold969_size11251	scaffold1685_size10654	scaffold1631_size6177
EOG093705VV	scaffold615_size18160	scaffold486_size18156	scaffold3811_size4289	scaffold39_size70496
EOG093705YA	scaffold1356_size11562	scaffold2208_size4599	scaffold739_size18929	scaffold176_size39241
EOG093705YE	scaffold87_size36286	scaffold1478_size7481	scaffold424_size25136	scaffold266_size31642
EOG0937068H	scaffold154_size30382	scaffold364_size20750	scaffold2434_size7397	scaffold177_size39107
EOG093706PM	scaffold2896_size6541	scaffold1069_size10252	scaffold277_size30817	scaffold1099_size10570
EOG093707RF	scaffold328_size23591	scaffold578_size16496	scaffold157_size37680	scaffold762_size15240
EOG093707VN	scaffold4041_size5577	scaffold1030_size10715	scaffold3166_size5452	scaffold678_size16752
EOG093708IM	scaffold46_size41920	scaffold180_size28687	scaffold432_size24981	scaffold387_size24815
EOG093708OP	scaffold211_size27426	scaffold225_size26309	scaffold1164_size14129	scaffold617_size18391
EOG093708TP	scaffold2422_size7664	scaffold940_size11512	scaffold1453_size11944	scaffold1197_size9497
EOG0937091Y	scaffold2911_size6515	scaffold370_size20677	scaffold3125_size5566	scaffold1085_size10791
EOG09370AJP	scaffold1823_size9440	scaffold3_size96191	scaffold1076_size14982	scaffold192_size37380
EOG09370AS9	scaffold262_size25729	scaffold208_size27147	scaffold290_size30169	scaffold3_size134068
EOG09370AV1	scaffold4504_size3960	scaffold1292_size8589	scaffold183_size35467	scaffold2129_size3990
EOG09370AWB	scaffold3008_size6325	scaffold494_size18011	scaffold441_size24761	scaffold1_size156486
EOG09370B1L	scaffold865_size15194	scaffold1638_size6686	scaffold148_size38367	scaffold266_size31642
EOG09370B7D	scaffold2703_size6990	scaffold1222_size9038	scaffold256_size31546	scaffold51_size63121
EOG09370CAV	scaffold2960_size6441	scaffold808_size12924	scaffold134_size39517	scaffold1204_size9457
EOG09370CIV	scaffold453_size20879	scaffold234_size25659	scaffold3_size83363	scaffold341_size27079
EOG09370CZT	scaffold2313_size11097	scaffold4_size93649	scaffold35_size57743	scaffold411_size24286

Table 4. Samples of metagenomes and metatranscriptomes from the TARA Oceans expeditions mapped against MAST-4 genes.

Tara Oceans ID	BioSamples ID	Type	Stations	Depth	Filter	Total number of reads	Accession Number (Sample)	Accession Number (run)	Related Biomolecular Data (url)	Sample Code
TARA_A200000123	SAMEA2591060	MetaT	TARA_7	SRF	0.8 - 5	258,519,466	ERS477934	ERR550396,ERR550403	http://www.ebi.ac.uk/ena/data/view/ERS477934	7SUR2GGMM14
TARA_A200000123	SAMEA2591060	MetaG	TARA_7	SRF	0.8 - 5	549,473,594	ERS477934	ERR315802,ERR315821	http://www.ebi.ac.uk/ena/data/view/ERS477934	7SUR1GGMM11
TARA_A100000551	SAMEA2591093	MetaT	TARA_23	SRF	0.8 - 5	418,232,176	ERS477988	ERR550521	http://www.ebi.ac.uk/ena/data/view/ERS477988	23SUR3GGMM14
TARA_A100000552	SAMEA2591095	MetaG	TARA_23	SRF	0.8 - 5	553,865,076	ERS477990	ERR538173,ERR318582	http://www.ebi.ac.uk/ena/data/view/ERS477990	23SUR1GGMM11
TARA_X000000323	SAMEA2619396	MetaT	TARA_4	SRF	0.8 - 5	420,350,248	ERS487919	ERR1719198	http://www.ebi.ac.uk/ena/data/view/ERS487919	4SUR1GGMM14
TARA_X000000323	SAMEA2619396	MetaG	TARA_4	SRF	0.8 - 5	362,874,162	ERS487919	ERR868369	http://www.ebi.ac.uk/ena/data/view/ERS487919	4SUR1GGMM11
TARA_X000000954	SAMEA2619534	MetaT	TARA_9	SRF	0.8 - 5	342,547,022	ERS488122	ERR1711995,ERR1711932	http://www.ebi.ac.uk/ena/data/view/ERS488122	9SUR1GGMM14
TARA_X000000954	SAMEA2619534	MetaG	TARA_9	SRF	0.8 - 5	361,242,930	ERS488122	ERR868407	http://www.ebi.ac.uk/ena/data/view/ERS488122	9SUR1GGMM11

TARA_A100000595	SAMEA2619675	MetaT	TARA_18	SRF	0.8 - 5	343,719,868	ERS488338	ERR1712185	http://www.ebi.ac.uk/ena/data/view/ERS488338	18SUR1GGMM14
TARA_A100000595	SAMEA2619675	MetaG	TARA_18	SRF	0.8 - 5	337,957,720	ERS488338	ERR868393	http://www.ebi.ac.uk/ena/data/view/ERS488338	18SUR1GGMM11
TARA_A100000534	SAMEA2619745	MetaT	TARA_22	SRF	0.8 - 5	508,959,870	ERS488446	ERR1719224,ERR1719453	http://www.ebi.ac.uk/ena/data/view/ERS488446	22SUR1GGMM14
TARA_A100000534	SAMEA2619745	MetaG	TARA_22	SRF	0.8 - 5	410,839,244	ERS488446	ERR868403	http://www.ebi.ac.uk/ena/data/view/ERS488446	22SUR1GGMM11
TARA_A100000393	SAMEA2619777	MetaT	TARA_25	SRF	0.8 - 5	361,912,434	ERS488497	ERR1712022,ERR1711998	http://www.ebi.ac.uk/ena/data/view/ERS488497	25SUR1GGMM14
TARA_A100000393	SAMEA2619777	MetaG	TARA_25	SRF	0.8 - 5	370,809,278	ERS488497	ERR868356	http://www.ebi.ac.uk/ena/data/view/ERS488497	25SUR1GGMM11
TARA_N000000316	SAMEA2619943	MetaT	TARA_36	SRF	0.8 - 5	468,881,322	ERS488730	ERR1719152,ERR1719488	http://www.ebi.ac.uk/ena/data/view/ERS488730	36SUR1GGMM14
TARA_N000000316	SAMEA2619943	MetaG	TARA_36	SRF	0.8 - 5	332,258,346	ERS488730	ERR868428,ERR868406	http://www.ebi.ac.uk/ena/data/view/ERS488730	36SUR1GGMM11
TARA_N000000029	SAMEA2620010	MetaT	TARA_38	SRF	0.8 - 5	306,420,568	ERS488809	ERR1711871,ERR1711968	http://www.ebi.ac.uk/ena/data/view/ERS488809	38SUR1GGMM14
TARA_N000000029	SAMEA2620010	MetaG	TARA_38	SRF	0.8 - 5	375,209,692	ERS488809	ERR868498,ERR868503	http://www.ebi.ac.uk/ena/data/view/ERS488809	38SUR0GGMM11
TARA_N000000006	SAMEA2620071	MetaT	TARA_39	SRF	0.8 - 5	483,268,850	ERS488885	ERR1719502,ERR1719267	http://www.ebi.ac.uk/ena/data/view/ERS488885	39SUR1GGMM14
TARA_N000000006	SAMEA2620071	MetaG	TARA_39	SRF	0.8 - 5	330,030,576	ERS488885	ERR868397,ERR868408	http://www.ebi.ac.uk/ena/data/view/ERS488885	39SUR1GGMM11
TARA_N000000071	SAMEA2620204	MetaT	TARA_41	SRF	0.8 - 5	328,987,220	ERS489053	ERR1712023,ERR1712005	http://www.ebi.ac.uk/ena/data/view/ERS489053	41SUR1GGMM14
TARA_N000000071	SAMEA2620204	MetaG	TARA_41	SRF	0.8 - 5	395,921,700	ERS489053	ERR868486	http://www.ebi.ac.uk/ena/data/view/ERS489053	41SUR1GGMM11
TARA_N000000267	SAMEA2620378	MetaT	TARA_46	SRF	0.8 - 5	515,526,368	ERS489279	ERR1719218,ERR1719498	http://www.ebi.ac.uk/ena/data/view/ERS489279	46SUR1GGMM14
TARA_N000000267	SAMEA2620378	MetaG	TARA_46	SRF	0.8 - 5	343,262,538	ERS489279	ERR868478	http://www.ebi.ac.uk/ena/data/view/ERS489279	46SUR1GGMM11
TARA_N000000231	SAMEA2620500	MetaT	TARA_51	SRF	0.8 - 5	400,600,874	ERS489448	ERR1719408,ERR1719306	http://www.ebi.ac.uk/ena/data/view/ERS489448	51SUR3GGMM14
TARA_N000000214	SAMEA2620503	MetaG	TARA_51	SRF	0.8 - 5	352,304,690	ERS489451	ERR868391	http://www.ebi.ac.uk/ena/data/view/ERS489451	51SUR1GGMM11
TARA_N000000598	SAMEA2620556	MetaT	TARA_52	SRF	0.8 - 5	370,897,412	ERS489543	ERR1711907,ERR1712199	http://www.ebi.ac.uk/ena/data/view/ERS489543	52SUR1GGMM15
TARA_N000000598	SAMEA2620556	MetaG	TARA_52	SRF	0.8 - 5	378,414,006	ERS489543	ERR599280	http://www.ebi.ac.uk/ena/data/view/ERS489543	52SUR0GGMM11
TARA_N000000522	SAMEA2620802	MetaT	TARA_64	SRF	0.8 - 5	307,884,038	ERS489933	ERR1711943,ERR1712100	http://www.ebi.ac.uk/ena/data/view/ERS489933	64SUR1GGMM14
TARA_N000000522	SAMEA2620802	MetaG	TARA_64	SRF	0.8 - 5	305,166,002	ERS489933	ERR599266	http://www.ebi.ac.uk/ena/data/view/ERS489933	64SUR1GGMM11
TARA_N000000933	SAMEA2620865	MetaT	TARA_65	SRF	0.8 - 5	465,743,312	ERS490039	ERR1719390,ERR1719261	http://www.ebi.ac.uk/ena/data/view/ERS490039	65SUR2GGMM14
TARA_N000000959	SAMEA2620870	MetaG	TARA_65	SRF	0.8 - 5	229,850,178	ERS490044	ERR599210,ERR1740328	http://www.ebi.ac.uk/ena/data/view/ERS490044	65SUR1GGMM12
TARA_N000000805	SAMEA2620939	MetaT	TARA_66	SRF	0.8 - 5	314,391,250	ERS490134	ERR1712216,ERR1712205	http://www.ebi.ac.uk/ena/data/view/ERS490134	66SUR1GGMM14
TARA_N000000805	SAMEA2620939	MetaG	TARA_66	SRF	0.8 - 5	249,842,444	ERS490134	ERR599261	http://www.ebi.ac.uk/ena/data/view/ERS490134	66SUR1GGMM11
TARA_N000000756	SAMEA2620988	MetaT	TARA_67	SRF	0.8 - 5	294,190,952	ERS490201	ERR1711906,ERR1711966	http://www.ebi.ac.uk/ena/data/view/ERS490201	67SUR5GGMM14
TARA_N000000756	SAMEA2620988	MetaG	TARA_67	SRF	0.8 - 5	296,830,858	ERS490201	ERR599302	http://www.ebi.ac.uk/ena/data/view/ERS490201	67SUR1GGMM11
TARA_N000000722	SAMEA2621029	MetaT	TARA_68	SRF	0.8 - 5	397,710,764	ERS490281	ERR1719484,ERR1719344	http://www.ebi.ac.uk/ena/data/view/ERS490281	68SUR1GGMM14
TARA_N000000722	SAMEA2621029	MetaG	TARA_68	SRF	0.8 - 5	321,213,290	ERS490281	ERR599257	http://www.ebi.ac.uk/ena/data/view/ERS490281	68SUR1GGMM11

TARA_N00000 0678	SAMEA26210 82	MetaT	TARA_70	SRF	0.8 - 5	340,732,858	ERS490343	ERR1712094,ERR1711924	http://www.ebi.ac.uk/ena/data/view/ERS490343	70SUR1GGM M14
TARA_N00000 0678	SAMEA26210 82	MetaG	TARA_70	SRF	0.8 - 5	301,556,840	ERS490343	ERR599305	http://www.ebi.ac.uk/ena/data/view/ERS490343	70SUR1GGM M11
TARA_N00000 1491	SAMEA26213 15	MetaT	TARA_80	SRF	0.8 - 5	389,357,552	ERS490751	ERR1711967,ERR1712070	http://www.ebi.ac.uk/ena/data/view/ERS490751	80SUR1GGM M14
TARA_N00000 1491	SAMEA26213 15	MetaG	TARA_80	SRF	0.8 - 5	397,406,080	ERS490751	ERR868387	http://www.ebi.ac.uk/ena/data/view/ERS490751	80SUR1GGM M11
TARA_N00000 1436	SAMEA26213 62	MetaT	TARA_81	SRF	0.8 - 5	467,261,852	ERS490817	ERR1740127,ERR1740138	http://www.ebi.ac.uk/ena/data/view/ERS490817	81SUR1GGM M14
TARA_N00000 1436	SAMEA26213 62	MetaG	TARA_81	SRF	0.8 - 5	362,982,552	ERS490817	ERR868372	http://www.ebi.ac.uk/ena/data/view/ERS490817	81SUR1GGM M11
TARA_N00000 1386	SAMEA26214 12	MetaT	TARA_82	SRF	0.8 - 5	399,345,252	ERS490896	ERR1719149,ERR1719424	http://www.ebi.ac.uk/ena/data/view/ERS490896	82SUR1GGM M14
TARA_N00000 1386	SAMEA26214 12	MetaG	TARA_82	SRF	0.8 - 5	413,831,578	ERS490896	ERR599298	http://www.ebi.ac.uk/ena/data/view/ERS490896	82SUR0GGM M11
TARA_N00000 1374	SAMEA26214 70	MetaT	TARA_83	SRF	0.8 - 5	392,454,868	ERS490977	ERR1740125,ERR1740126	http://www.ebi.ac.uk/ena/data/view/ERS490977	83SUR1GGM M14
TARA_N00000 1374	SAMEA26214 70	MetaG	TARA_83	SRF	0.8 - 5	335,710,694	ERS490977	ERR868388	http://www.ebi.ac.uk/ena/data/view/ERS490977	83SUR1GGM M11
TARA_N00000 1438	SAMEA26214 98	MetaT	TARA_84	SRF	0.8 - 5	369,399,294	ERS491012	ERR1711865,ERR1712028	http://www.ebi.ac.uk/ena/data/view/ERS491012	84SUR1GGM M14
TARA_N00000 1438	SAMEA26214 98	MetaG	TARA_84	SRF	0.8 - 5	301,358,538	ERS491012	ERR599254	http://www.ebi.ac.uk/ena/data/view/ERS491012	84SUR0GGM M11
TARA_N00000 1028	SAMEA26215 22	MetaG	TARA_85	SRF	0.8 - 5	285,775,170	ERS491057	ERR599335	http://www.ebi.ac.uk/ena/data/view/ERS491057	85SUR0GGM M11
TARA_N00000 1029	SAMEA26215 23	MetaT	TARA_85	SRF	0.8 - 5	470,137,682	ERS491058	ERR1740133,ERR1740130	http://www.ebi.ac.uk/ena/data/view/ERS491058	85SUR2GGM M14
TARA_N00000 1299	SAMEA26217 72	MetaT	TARA_92	SRF	0.8 - 5	477,018,734	ERS491398	ERR1719503,ERR1719329	http://www.ebi.ac.uk/ena/data/view/ERS491398	92SUR1GGM M14
TARA_N00000 1299	SAMEA26217 72	MetaG	TARA_92	SRF	0.8 - 5	464,764,696	ERS491398	ERR868413	http://www.ebi.ac.uk/ena/data/view/ERS491398	92SUR1GGM M11
TARA_N00000 1296	SAMEA26217 91	MetaT	TARA_93	SRF	0.8 - 5	371,748,314	ERS491433	ERR1712103,ERR1711965	http://www.ebi.ac.uk/ena/data/view/ERS491433	93SUR2GGM M14
TARA_N00000 1296	SAMEA26217 91	MetaG	TARA_93	SRF	0.8 - 5	432,700,952	ERS491433	ERR868416	http://www.ebi.ac.uk/ena/data/view/ERS491433	93SUR1GGM M11
TARA_N00000 1608	SAMEA26221 06	MetaT	TARA_100	SRF	0.8 - 5	515,771,452	ERS491845	ERR1719161,ERR1719388	http://www.ebi.ac.uk/ena/data/view/ERS491845	100SUR1GGM M14
TARA_N00000 1608	SAMEA26221 06	MetaG	TARA_100	SRF	0.8 - 5	399,989,404	ERS491845	ERR868493	http://www.ebi.ac.uk/ena/data/view/ERS491845	100SUR1GGM M11
TARA_N00000 1650	SAMEA26221 84	MetaT	TARA_102	SRF	0.8 - 5	467,577,532	ERS491949	ERR1719172,ERR1719375	http://www.ebi.ac.uk/ena/data/view/ERS491949	102SUR1GGM M14
TARA_N00000 1650	SAMEA26221 84	MetaG	TARA_102	SRF	0.8 - 5	325,889,108	ERS491949	ERR868357	http://www.ebi.ac.uk/ena/data/view/ERS491949	102SUR1GGM M11
TARA_N00000 1730	SAMEA26223 25	MetaT	TARA_109	SRF	0.8 - 5	368,072,090	ERS492154	ERR1719235	http://www.ebi.ac.uk/ena/data/view/ERS492154	109SUR1GGM M14
TARA_N00000 1730	SAMEA26223 25	MetaG	TARA_109	SRF	0.8 - 5	359,273,186	ERS492154	ERR868441,ERR868374	http://www.ebi.ac.uk/ena/data/view/ERS492154	109SUR1GGM M11
TARA_N00000 1750	SAMEA26223 91	MetaT	TARA_110	SRF	0.8 - 5	332,445,356	ERS492243	ERR1712078,ERR1711886	http://www.ebi.ac.uk/ena/data/view/ERS492243	110SUR1GGM M14
TARA_N00000 1750	SAMEA26223 91	MetaG	TARA_110	SRF	0.8 - 5	332,511,730	ERS492243	ERR868442	http://www.ebi.ac.uk/ena/data/view/ERS492243	110SUR1GGM M11
TARA_N00000 1812	SAMEA26224 63	MetaT	TARA_111	SRF	0.8 - 5	380,931,346	ERS492332	ERR1711904,ERR1711985	http://www.ebi.ac.uk/ena/data/view/ERS492332	111SUR2GGM M14
TARA_N00000 1812	SAMEA26224 63	MetaG	TARA_111	SRF	0.8 - 5	333,731,898	ERS492332	ERR868476	http://www.ebi.ac.uk/ena/data/view/ERS492332	111SUR1GGM M11
TARA_N00000 1938	SAMEA26226 61	MetaT	TARA_122	SRF	0.8 - 5	473,706,518	ERS492651	ERR1712182,ERR1712118,ERR1711869	http://www.ebi.ac.uk/ena/data/view/ERS492651	122SUR1GGM M14
TARA_N00000 1938	SAMEA26226 61	MetaG	TARA_122	SRF	0.8 - 5	603,984,892	ERS492651	ERR868475,ERR868513	http://www.ebi.ac.uk/ena/data/view/ERS492651	122SUR1GGM M11

TARA_N00001992	SAMEA2622717	MetaT	TARA_123	SRF	0.8 - 5	429,513,964	ERS492740	ERR1719256,ERR1719298,ERR1719217	http://www.ebi.ac.uk/ena/data/view/ERS492740	123SUR3GGMM14
TARA_N00001992	SAMEA2622717	MetaG	TARA_123	SRF	0.8 - 5	623,726,444	ERS492740	ERR868466,ERR868469	http://www.ebi.ac.uk/ena/data/view/ERS492740	123SUR1GGMM11
TARA_N00002037	SAMEA2622770	MetaT	TARA_124	SRF	0.8 - 5	505,566,386	ERS492825	ERR1719301,ERR1719160,ERR1719214	http://www.ebi.ac.uk/ena/data/view/ERS492825	124SUR1GGMM14
TARA_N00002037	SAMEA2622770	MetaG	TARA_124	SRF	0.8 - 5	542,739,706	ERS492825	ERR868363,ERR868489	http://www.ebi.ac.uk/ena/data/view/ERS492825	124SUR1GGMM11
TARA_N00002019	SAMEA2622826	MetaT	TARA_125	SRF	0.8 - 5	560,541,472	ERS492897	ERR1719395,ERR1719316,ERR1719207	http://www.ebi.ac.uk/ena/data/view/ERS492897	125SUR1GGMM14
TARA_N00002019	SAMEA2622826	MetaG	TARA_125	SRF	0.8 - 5	267,060,780	ERS492897	ERR868382,ERR868352	http://www.ebi.ac.uk/ena/data/view/ERS492897	125SUR1GGMM11
TARA_N00002289	SAMEA2622914	MetaT	TARA_128	SRF	0.8 - 5	446,561,198	ERS493057	ERR1719364,ERR1719311,ERR1719405	http://www.ebi.ac.uk/ena/data/view/ERS493057	128SUR1GGMM14
TARA_N00002289	SAMEA2622914	MetaG	TARA_128	SRF	0.8 - 5	374,266,272	ERS493057	ERR868462	http://www.ebi.ac.uk/ena/data/view/ERS493057	128SUR1GGMM11
TARA_N00002352	SAMEA2623018	MetaT	TARA_131	SRF	0.8 - 5	376,365,764	ERS493224	ERR1712054,ERR1711975	http://www.ebi.ac.uk/ena/data/view/ERS493224	131SUR1GGMM14
TARA_N00002352	SAMEA2623018	MetaG	TARA_131	SRF	0.8 - 5	354,877,116	ERS493224	ERR868485	http://www.ebi.ac.uk/ena/data/view/ERS493224	131SUR1GGMM11
TARA_N00002416	SAMEA2623072	MetaT	TARA_132	SRF	0.8 - 5	374,725,650	ERS493313	ERR1711988,ERR1711987	http://www.ebi.ac.uk/ena/data/view/ERS493313	132SUR1GGMM14
TARA_N00002416	SAMEA2623072	MetaG	TARA_132	SRF	0.8 - 5	407,723,756	ERS493313	ERR868480	http://www.ebi.ac.uk/ena/data/view/ERS493313	132SUR1GGMM11
TARA_N00002179	SAMEA2623206	MetaT	TARA_135	SRF	0.8 - 5	477,681,814	ERS493519	ERR1719181,ERR1719155	http://www.ebi.ac.uk/ena/data/view/ERS493519	135SUR1GGMM14
TARA_N00002179	SAMEA2623206	MetaG	TARA_135	SRF	0.8 - 5	334,088,220	ERS493519	ERR868433	http://www.ebi.ac.uk/ena/data/view/ERS493519	135SUR1GGMM11
TARA_N00002961	SAMEA2623266	MetaT	TARA_136	SRF	0.8 - 5	377,973,560	ERS493612	ERR1719171,ERR1719241	http://www.ebi.ac.uk/ena/data/view/ERS493612	136SUR1GGMM14
TARA_N00002961	SAMEA2623266	MetaG	TARA_136	SRF	0.8 - 5	373,081,762	ERS493612	ERR868378	http://www.ebi.ac.uk/ena/data/view/ERS493612	136SUR1GGMM11
TARA_N00002925	SAMEA2623284	MetaT	TARA_137	SRF	0.8 - 5	461,186,060	ERS493645	ERR1719387,ERR1719252	http://www.ebi.ac.uk/ena/data/view/ERS493645	137SUR1GGMM14
TARA_N00002925	SAMEA2623284	MetaG	TARA_137	SRF	0.8 - 5	335,779,338	ERS493645	ERR868477	http://www.ebi.ac.uk/ena/data/view/ERS493645	137SUR1GGMM11
TARA_N00003037	SAMEA2623419	MetaT	TARA_139	SRF	0.8 - 5	390,873,990	ERS493853	ERR1719185,ERR1719313	http://www.ebi.ac.uk/ena/data/view/ERS493853	139SUR1GGMM14
TARA_N00003037	SAMEA2623419	MetaG	TARA_139	SRF	0.8 - 5	423,658,582	ERS493853	ERR868420	http://www.ebi.ac.uk/ena/data/view/ERS493853	139SUR1GGMM11
TARA_N00003083	SAMEA2623479	MetaT	TARA_142	SRF	0.8 - 5	383,957,742	ERS493954	ERR1719165,ERR1719487	http://www.ebi.ac.uk/ena/data/view/ERS493954	142SUR2GGMM14
TARA_N00003083	SAMEA2623479	MetaG	TARA_142	SRF	0.8 - 5	327,695,642	ERS493954	ERR868430	http://www.ebi.ac.uk/ena/data/view/ERS493954	142SUR1GGMM11
TARA_N00003179	SAMEA2623603	MetaT	TARA_144	SRF	0.8 - 5	232,331,470	ERS494131	ERR1712089,ERR1712041	http://www.ebi.ac.uk/ena/data/view/ERS494131	144SUR1GGMM14
TARA_N00003179	SAMEA2623603	MetaG	TARA_144	SRF	0.8 - 5	489,642,732	ERS494131	ERR873964	http://www.ebi.ac.uk/ena/data/view/ERS494131	144SUR1GGMM12
TARA_N00003219	SAMEA2623641	MetaT	TARA_145	SRF	0.8 - 5	324,271,990	ERS494184	ERR1719199	http://www.ebi.ac.uk/ena/data/view/ERS494184	145SUR1GGMM14
TARA_N00003219	SAMEA2623641	MetaG	TARA_145	SRF	0.8 - 5	398,293,800	ERS494184	ERR868411	http://www.ebi.ac.uk/ena/data/view/ERS494184	145SUR1GGMM11
TARA_N00003253	SAMEA2623685	MetaT	TARA_146	SRF	0.8 - 5	340,205,542	ERS494248	ERR1719409	http://www.ebi.ac.uk/ena/data/view/ERS494248	146SUR1GGMM14
TARA_N00003253	SAMEA2623685	MetaG	TARA_146	SRF	0.8 - 5	368,598,422	ERS494248	ERR868351	http://www.ebi.ac.uk/ena/data/view/ERS494248	146SUR1GGMM11
TARA_N00002103	SAMEA2623723	MetaT	TARA_147	SRF	0.8 - 5	347,805,760	ERS494304	ERR1719514,ERR1719478	http://www.ebi.ac.uk/ena/data/view/ERS494304	147SUR1GGMM14
TARA_N00002103	SAMEA2623723	MetaG	TARA_147	SRF	0.8 - 5	359,399,988	ERS494304	ERR868366	http://www.ebi.ac.uk/ena/data/view/ERS494304	147SUR1GGMM11

TARA_N00000 2697	SAMEA26238 17	MetaT	TARA_150	SRF	0.8 - 5	254,647,306	ERS494454	ERR1719258	http://www.ebi.ac.uk/ena/data/view/ERS494454	150SUR1GGMM14
TARA_N00000 2697	SAMEA26238 17	MetaG	TARA_150	SRF	0.8 - 5	413,027,964	ERS494454	ERR868354	http://www.ebi.ac.uk/ena/data/view/ERS494454	150SUR1GGMM11
TARA_N00000 2741	SAMEA26238 61	MetaT	TARA_151	SRF	0.8 - 5	344,199,012	ERS494529	ERR1719440	http://www.ebi.ac.uk/ena/data/view/ERS494529	151SUR1GGMM14
TARA_N00000 2741	SAMEA26238 61	MetaG	TARA_151	SRF	0.8 - 5	341,449,348	ERS494529	ERR868459	http://www.ebi.ac.uk/ena/data/view/ERS494529	151SUR1GGMM11
TARA_N00000 2789	SAMEA26239 01	MetaT	TARA_152	SRF	0.8 - 5	349,688,472	ERS494594	ERR1719353	http://www.ebi.ac.uk/ena/data/view/ERS494594	152SUR1GGMM14
TARA_N00000 2789	SAMEA26239 01	MetaG	TARA_152	SRF	0.8 - 5	336,402,944	ERS494594	ERR868445	http://www.ebi.ac.uk/ena/data/view/ERS494594	152SUR1GGMM11

Table 5. OTUs used in the Network Association with MICTools. Low and Upp CI_values determine the Confidence Interval, while the Sign implies the position of the observed value based on the CI: HIGHER - Observed value is > Upp CI, LOWER - Observed value is < Low CI, NON-SIGNIFICANT - Observed value is within both boundaries.

Table too large to fit, link to the published version in: **Dataset_S05** at https://www.pnas.org/doi/10.1073/pnas.2020955118#supplementary-materials

Table 6. Summary of all hits of MAST-4 genes to different databases.

Genome	# Total Genes	# Genes mapped EggNOG	EggNOG (%)	# Genes mapped KEGG	KEGG (%)	# Genees mapped CAZY	CAZY (%)	# Genes mapped to UNIGENE	UNIGENE (%)
MAST-4A	15508	11399	73.50	4388	28.30	503	3.24	3855	24.86
MAST-4B	10019	7436	74.22	3004	29.98	303	3.02	1987	19.83
MAST-4C	16260	12031	73.99	4909	30.19	497	3.06	5448	33.51
MAST-4E	9042	6966	77.04	3108	34.37	309	3.42	1191	13.17

Table 7. Summary of all GHs gene families found in MAST-4 genomes. A value of 0 indicates that no gene was annotated as part of such GH family, while a value of 1 indicates that at least one gene was found.

GH family	MAST-4A	MAST-4B	MAST-4C	MAST-4E
GH1	1	0	0	1
GH10	1	1	1	1
GH109	1	1	1	1
GH110	1	0	1	0
GH13	1	1	1	1
GH120	1	0	0	0
GH135	1	1	1	1

GH136	1	0	1	1
GH14	1	0	1	0
GH141	1	1	1	1
GH144	1	0	0	0
GH15	1	0	0	1
GH16	1	1	1	1
GH18	1	1	1	1
GH19	1	1	1	1
GH2	1	1	1	1
GH20	1	1	1	1
GH22	1	1	1	0
GH24	1	0	1	0
GH25	1	1	1	1
GH27	1	1	1	1
GH28	1	1	1	1
GH3	1	1	1	1
GH31	1	1	1	1
GH32	1	1	1	1
GH33	1	1	1	1
GH35	1	1	1	1
GH36	1	1	1	1
GH37	1	1	1	1
GH38	1	1	1	1
GH39	1	0	1	1
GH43	1	1	1	1
GH47	1	1	1	1
GH5	1	1	1	1
GH54	1	1	1	1
GH55	1	1	1	1
GH56	1	1	1	1
GH59	1	1	1	1
GH63	1	0	1	0
GH65	1	1	1	0
GH67	1	1	1	1
GH74	1	1	1	1
GH78	1	1	1	1
GH79	1	1	1	1
GH86	1	0	0	1
GH89	1	1	1	1
GH92	1	0	1	1
GH99	1	1	1	1
GH105	0	1	0	1
GH117	0	1	0	0

GH130	0	1	1	0
GH23	0	1	0	0
GH104	0	0	1	0
GH128	0	0	1	0
GH30	0	0	1	0
GH139	0	0	0	1
GH29	0	0	0	1

Table 8. Expression (TPM) means for the 20 most expressed GH genes (Figure 1.5) and the 152 single-copy housekeeping genes (from BUSCO eukaryota_odb9) found in MAST-4 for each *Tara Oceans* station. p-values corresponding to the difference of the means for each station are indicated (Wilcoxon test).

	Mean TPM expression		
Station	**Housekeeping genes**	**GH genes**	**p-value (Wilcoxon test)**
4	12.88	166.18	0.00
7	11.38	90.25	0.00
9	16.36	15.68	0.58
18	15.69	37.38	0.73
22	19.42	241.86	0.01
23	12.52	146.55	0.01
25	7.87	121.95	0.01
36	18.52	81.55	0.22
38	7.49	73.70	0.00
39	11.25	100.30	0.00
41	14.06	46.57	0.00
46	14.16	140.61	0.03
51	8.62	110.25	0.01
52	6.34	127.90	0.00
64	26.35	70.61	0.00
65	15.61	183.07	0.01
66	14.44	79.80	0.07
67	29.71	67.39	0.89
68	15.36	175.35	0.12
70	5.11	156.73	0.28
80	15.21	125.57	0.16
81	22.48	163.58	0.19
82	31.97	71.92	0.87
83	29.55	72.62	0.48
84	0.00	0.00	NaN
85	0.00	0.00	NaN
92	31.35	47.91	0.18

93	23.78	32.60	0.38
100	8.30	102.54	0.00
102	10.85	103.62	0.00
109	16.02	77.17	0.00
110	16.22	50.03	0.02
111	6.21	66.73	0.03
122	0.69	3.46	0.39
123	18.91	39.04	0.00
124	10.18	100.21	0.00
125	16.86	65.41	0.00
128	16.37	99.18	0.00
131	8.58	73.13	0.00
132	12.46	52.98	0.10
135	15.84	109.73	0.16
136	15.50	87.85	0.00
137	16.75	112.20	0.00
139	17.74	91.41	0.01
142	11.32	174.70	0.00
144	24.96	0.00	0.53
145	25.49	153.91	0.40
146	25.34	212.97	0.54
147	17.36	188.94	0.04
150	14.72	147.48	0.07
151	14.09	130.82	0.02
152	18.29	203.48	0.45

Table 9. List of all homologous genes between all four MAST-4 species. For each homolog alignment, the number of branches and sites with positive selection is given along with the function from dbCAN (CAZymes).

Homologous genes aligned				Positive Selection		Function	
MAST-4E	**MAST-4B**	**MAST-4A**	**MAST-4C**	**# Branches**	**Selected Branches**	**# Sites**	**dbCANfamily**
g4327	g417	g3952	g13398	2	MAST4-A,B	11	NaN
g1052	g1838	g9756	g232	2	MAST4-B,E	0	NaN
g185	g9615	g12806	g9727	2	MAST4-A,B	0	NaN
g6381	g142	g10016	g2472	1	MAST4B	15	NaN
g2405	g4515	g578	g2172	1	MAST4A	10	NaN
g2270	g6403	g1544	g11439	1	Node	6	NaN
g1701	g3272	g3566	g5587	1	MAST4B	5	NaN
g1601	g5788	g11046	g2181	1	MAST4B	4	NaN
g6713	g1319	g9114	g11598	1	MAST4A	4	NaN
g2401	g7860	g1073	g4064	1	MAST4B	3	NaN
g2566	g1048	g160	g7601	1	MAST4E	3	NaN
g3555	g8666	g10237	g5514	1	MAST4B	3	CBM9
g4322	g6250	g4048	g3569	1	MAST4A	3	NaN
g484	g2045	g6677	g1432	1	MAST4B	3	CE1
g5529	g7137	g13451	g11772	1	MAST4B	3	NaN
g5725	g8986	g8759	g11385	1	MAST4A	3	GH74
g6078	g9056	g14486	g11587	1	MAST4B	3	NaN
g6707	g3019	g4436	g11078	1	MAST4B	3	NaN
g873	g5501	g4252	g9077	1	MAST4A	3	NaN
g2913	g9382	g12339	g10781	1	MAST4B	2	NaN
g3052	g2298	g3076	g1150	1	MAST4A	2	NaN

g3240	g2479	g8499	g16220	1	MAST4A	2	NaN
g3543	g3780	g8990	g7554	1	MAST4B	2	NaN
g3544	g7744	g8994	g962	1	MAST4A	2	NaN
g3838	g231	g12981	g9686	1	MAST4B	2	NaN
g5256	g2364	g11459	g2751	1	MAST4A	2	NaN
g5847	g3269	g10064	g3351	1	Node	2	NaN
g726	g365	g12182	g183	1	MAST4A	2	NaN
g8663	g736	g12614	g11715	1	MAST4A	2	NaN
g8785	g7349	g6445	g12396	1	MAST4A	2	NaN
g8787	g2954	g10268	g14289	1	MAST4C	2	NaN
g1948	g1405	g12633	g14750	1	MAST4B	1	NaN
g251	g3871	g8800	g12503	1	MAST4A	1	NaN
g302	g5146	g8469	g2048	1	MAST4C	1	NaN
g3984	g7713	g3093	g11108	1	MAST4C	1	NaN
g4396	g7435	g3026	g15210	1	MAST4E	1	NaN
g6390	g5454	g13809	g13112	1	MAST4B	1	NaN
g6446	g174	g2549	g9238	1	MAST4A	1	NaN
g8687	g8905	g11288	g6349	1	MAST4A	1	NaN
g1829	g9742	g1134	g2509	1	MAST4C	0	NaN
g223	g4790	g2954	g12289	1	MAST4A	0	NaN
g2784	g9977	g15198	g7827	1	MAST4B	0	NaN
g2947	g2227	g6861	g13193	1	MAST4C	0	NaN
g3475	g5923	g8974	g6707	1	MAST4C	0	NaN
g3831	g3377	g5608	g14765	1	MAST4A	0	NaN
g4236	g1706	g15432	g5148	1	MAST4C	0	NaN
g4868	g4676	g10502	g6405	1	MAST4B	0	NaN
g5642	g8532	g4460	g5264	1	MAST4B	0	NaN
g6052	g1047	g8448	g485	1	MAST4E	0	NaN
g7130	g3749	g6186	g14645	1	MAST4A	0	NaN
g7298	g4947	g1208	g1743	1	MAST4B	0	NaN
g7748	g7342	g4657	g4391	1	MAST4B	0	NaN
g8152	g2484	g5271	g11851	1	MAST4B	0	NaN
g8616	g571	g5232	g7668	1	MAST4A	0	NaN
g8702	g4096	g1845	g1242	1	Node	0	NaN
g8783	g5496	g3630	g2291	1	Node	0	NaN
g882	g2181	g4965	g8481	1	MAST4B	0	NaN
g99	g8178	g13719	g5400	1	MAST4C	0	NaN
g1451	g7357	g14281	g14531	1	MAST4A	2	NaN
g1656	g2999	g988	g9066	1	MAST4B	1	NaN
g8586	g4232	g3305	g7397	0		14	NaN
g3938	g2541	g4951	g6699	0		12	NaN
g3672	g8853	g1661	g4675	0		11	NaN
g3895	g4667	g1880	g511	0		10	NaN
g3903	g6754	g1992	g15185	0		9	NaN
g7127	g6693	g14702	g9732	0		9	GH30_1
g5514	g8733	g4848	g14172	0		8	NaN
g1127	g5427	g14499	g14345	0		7	NaN
g1609	g1831	g4033	g4338	0		7	NaN
g1632	g455	g13400	g10033	0		7	NaN
g1738	g8676	g1611	g7358	0		7	NaN
g2994	g2431	g6983	g13108	0		7	GH13_17
g3258	g2217	g13025	g8804	0		7	GH19
g3380	g9297	g753	g3668	0		7	NaN
g4308	g2754	g3278	g12520	0		7	NaN
g6595	g4645	g235	g14376	0		7	NaN
g787	g3993	g4841	g3853	0		7	NaN
g8050	g6215	g3136	g10306	0		7	NaN
g8770	g3741	g5978	g192	0		7	NaN
g174	g4446	g4179	g14386	0		6	GH78
g2262	g7878	g5272	g8419	0		6	NaN
g2463	g1766	g11716	g5708	0		6	NaN
g2937	g7332	g14230	g6665	0		6	NaN
g2983	g6380	g6754	g10743	0		6	NaN
g3414	g7772	g13863	g3522	0		6	GH78
g3450	g355	g55	g16100	0		6	NaN
g3722	g6557	g14947	g5710	0		6	NaN
g4178	g2672	g13742	g4273	0		6	NaN
g4592	g7394	g12822	g15280	0		6	NaN
g5979	g9346	g5060	g10155	0		6	NaN
g6019	g8874	g12905	g2142	0		6	NaN
g6354	g7133	g11164	g15689	0		6	NaN
g6616	g8566	g12520	g5598	0		6	NaN
g8215	g5951	g14573	g4788	0		6	NaN
g8608	g6620	g12787	g8462	0		6	NaN
g1223	g4847	g11431	g7443	0		5	GT4
g1930	g9222	g13389	g1433	0		5	NaN
g2080	g7911	g6476	g9124	0		5	NaN
g2670	g5614	g13935	g12610	0		5	NaN
g4433	g220	g234	g14055	0		5	NaN
g4783	g8757	g7010	g6063	0		5	NaN
g5162	g9605	g14879	g897	0		5	NaN
g5234	g7634	g8171	g5122	0		5	NaN

g5716	g5310	g12110	g6248	0	5	NaN
g5966	g9305	g1095	g599	0	5	NaN
g6070	g3537	g13817	g5475	0	5	GH89
g7014	g2713	g7592	g8074	0	5	CBM23
g7747	g1826	g4654	g4394	0	5	NaN
g8534	g8156	g4480	g8284	0	5	NaN
g8786	g2027	g4391	g13548	0	5	NaN
g880	g9315	g1859	g8483	0	5	NaN
g8875	g6657	g12154	g7496	0	5	NaN
g1150	g283	g10277	g824	0	4	NaN
g189	g3624	g10353	g14185	0	4	NaN
g20	g6605	g7565	g983	0	4	GH79
g2184	g260	g11090	g3753	0	4	NaN
g2644	g4466	g2773	g11361	0	4	NaN
g2759	g7592	g1489	g5052	0	4	NaN
g2889	g6689	g1717	g8360	0	4	NaN
g2962	g4848	g15422	g7440	0	4	NaN
g3066	g6682	g14559	g15032	0	4	NaN
g3218	g9208	g8587	g8622	0	4	NaN
g3489	g627	g7655	g1449	0	4	NaN
g3518	g1183	g1123	g1722	0	4	NaN
g3526	g1454	g14991	g6847	0	4	NaN
g4601	g695	g14543	g14827	0	4	NaN
g4697	g5789	g2945	g4498	0	4	NaN
g4816	g4354	g5258	g15038	0	4	NaN
g4971	g377	g5578	g10541	0	4	NaN
g5047	g2519	g2552	g2849	0	4	NaN
g5167	g2328	g5753	g9474	0	4	NaN
g5913	g8972	g7631	g3531	0	4	NaN
g6371	g4781	g4476	g2610	0	4	NaN
g6756	g593	g1832	g9760	0	4	NaN
g705	g905	g11437	g2505	0	4	NaN
g7364	g7608	g207	g5012	0	4	NaN
g7393	g5559	g5170	g7728	0	4	NaN
g7402	g3281	g7161	g98	0	4	NaN
g8044	g588	g509	g9024	0	4	NaN
g8186	g768	g11040	g16152	0	4	NaN
g8363	g7321	g1752	g7875	0	4	GT1
g8466	g8369	g12678	g13520	0	4	NaN
g8552	g4261	g588	g13828	0	4	NaN
g888	g5732	g11545	g12113	0	4	NaN
g8960	g8635	g11705	g5119	0	4	NaN
g8991	g5506	g12958	g3520	0	4	NaN
g944	g5877	g7767	g14548	0	4	NaN
g1023	g4816	g6752	g8170	0	3	NaN
g1162	g8653	g6329	g14065	0	3	NaN
g1217	g4755	g5958	g9837	0	3	NaN
g1239	g8082	g467	g4401	0	3	NaN
g1429	g625	g7653	g9231	0	3	NaN
g1471	g6342	g6255	g10338	0	3	NaN
g169	g6775	g9339	g1592	0	3	NaN
g1956	g6312	g10086	g15217	0	3	NaN
g1966	g2007	g1070	g3670	0	3	NaN
g2046	g6506	g5692	g10022	0	3	NaN
g2071	g9038	g15283	g8253	0	3	NaN
g2103	g1752	g5179	g14963	0	3	NaN
g2343	g7476	g12754	g5350	0	3	NaN
g2486	g3292	g3790	g1167	0	3	NaN
g2500	g8138	g5198	g12860	0	3	NaN
g2502	g319	g13564	g8530	0	3	NaN
g2826	g8904	g1546	g10348	0	3	NaN
g297	g8077	g6284	g7155	0	3	GH16
g3151	g5637	g3443	g4240	0	3	NaN
g3169	g2396	g6373	g2624	0	3	NaN
g3190	g2794	g13870	g1068	0	3	NaN
g3265	g1868	g11189	g11630	0	3	NaN
g3437	g1231	g12186	g14566	0	3	NaN
g3487	g4941	g6158	g8592	0	3	NaN
g3545	g7743	g8993	g963	0	3	NaN
g3745	g9961	g12654	g9609	0	3	NaN
g3836	g228	g12984	g9689	0	3	NaN
g3859	g808	g11048	g15844	0	3	NaN
g3860	g784	g960	g10435	0	3	GH28
g3953	g3587	g548	g7719	0	3	NaN
g4001	g4353	g964	g7792	0	3	GT4
g4208	g5644	g14003	g5675	0	3	NaN
g4240	g1778	g983	g8326	0	3	NaN
g4306	g2752	g7002	g12518	0	3	NaN
g4939	g5326	g2520	g9573	0	3	NaN
g5018	g8945	g6715	g2741	0	3	NaN
g5096	g80	g4425	g2457	0	3	NaN
g51	g926	g7799	g2303	0	3	GH30_1

g52	g2317	g1585	g9646	0	3	NaN
g5298	g196	g9669	g5243	0	3	NaN
g5378	g7915	g6472	g6356	0	3	NaN
g5516	g7759	g236	g8196	0	3	NaN
g564	g8706	g3988	g9091	0	3	NaN
g5728	g4923	g11729	g6354	0	3	GT4
g5771	g1428	g11452	g6439	0	3	NaN
g5785	g868	g14157	g13693	0	3	NaN
g593	g892	g4663	g2827	0	3	NaN
g6114	g7163	g10379	g13594	0	3	NaN
g6181	g7118	g4911	g13511	0	3	NaN
g6204	g815	g14810	g12897	0	3	NaN
g6214	g3397	g1996	g457	0	3	NaN
g6303	g8939	g7618	g6730	0	3	NaN
g6665	g8259	g4999	g14520	0	3	NaN
g671	g2843	g10143	g14302	0	3	NaN
g6729	g8176	g3658	g8448	0	3	GH99
g6732	g2836	g1299	g4233	0	3	NaN
g6858	g9287	g464	g13498	0	3	NaN
g6871	g3685	g13431	g13804	0	3	NaN
g6935	g7642	g7885	g3370	0	3	NaN
g7222	g4026	g8029	g6965	0	3	NaN
g729	g7565	g11509	g8760	0	3	NaN
g7313	g1413	g5925	g11241	0	3	GH3
g736	g9477	g1496	g11285	0	3	NaN
g7388	g3413	g10034	g7458	0	3	NaN
g7546	g106	g3793	g12768	0	3	NaN
g7630	g8988	g1199	g4073	0	3	GH28
g766	g4452	g13503	g6860	0	3	NaN
g7720	g6674	g8667	g8857	0	3	NaN
g7792	g2750	g6879	g12516	0	3	NaN
g7869	g7028	g2720	g15416	0	3	NaN
g8042	g590	g1829	g9022	0	3	NaN
g8059	g5160	g5760	g8189	0	3	NaN
g8082	g2848	g15135	g5579	0	3	NaN
g8207	g40	g13946	g9249	0	3	NaN
g8921	g4437	g5428	g15341	0	3	NaN
g9032	g1110	g12389	g11528	0	3	NaN
g935	g507	g14945	g8353	0	3	NaN
g938	g5320	g4282	g13117	0	3	GH3
g966	g470	g2165	g13735	0	3	NaN
g996	g5651	g9549	g10063	0	3	NaN
g1050	g9247	g2613	g14062	0	2	NaN
g1116	g8773	g9501	g9901	0	2	NaN
g1455	g5313	g9392	g15794	0	2	NaN
g1520	g8459	g6656	g13051	0	2	NaN
g1526	g1517	g643	g12924	0	2	NaN
g1647	g4310	g5110	g13651	0	2	NaN
g1751	g9667	g1040	g10674	0	2	NaN
g1752	g4406	g3190	g5376	0	2	NaN
g1794	g5740	g10140	g453	0	2	NaN
g1841	g9763	g15217	g8042	0	2	NaN
g2271	g6404	g1545	g11438	0	2	NaN
g2408	g8481	g8119	g11826	0	2	NaN
g2440	g3108	g7505	g7540	0	2	NaN
g2458	g1115	g4182	g1889	0	2	CBM40
g2466	g5873	g10689	g12361	0	2	NaN
g2508	g1107	g3490	g15115	0	2	NaN
g2514	g9069	g1918	g2740	0	2	NaN
g2633	g3399	g13139	g7818	0	2	NaN
g2663	g122	g3408	g12872	0	2	NaN
g2681	g9204	g8590	g8626	0	2	NaN
g2686	g587	g11603	g706	0	2	NaN
g2716	g851	g7704	g7495	0	2	NaN
g2724	g7736	g13340	g2279	0	2	NaN
g2775	g8529	g11878	g10186	0	2	NaN
g2791	g4085	g1371	g231	0	2	NaN
g3059	g4779	g10485	g11476	0	2	NaN
g3101	g6240	g3609	g15605	0	2	NaN
g3124	g7550	g4555	g8435	0	2	NaN
g3254	g6729	g10049	g2674	0	2	NaN
g3281	g5139	g3461	g8922	0	2	NaN
g3485	g9726	g8129	g12547	0	2	NaN
g3503	g2591	g1927	g5747	0	2	NaN
g3561	g9225	g13380	g15478	0	2	NaN
g3623	g2421	g3407	g11512	0	2	NaN
g376	g4832	g10479	g14705	0	2	NaN
g3769	g1249	g9332	g7040	0	2	NaN
g3797	g2708	g1900	g13021	0	2	NaN
g3815	g6266	g9361	g502	0	2	NaN
g3818	g8901	g9198	g10350	0	2	NaN
g3845	g8261	g4997	g15647	0	2	NaN

g4186	g4181	g8217	g15304	0	2	NaN
g4212	g3934	g370	g13166	0	2	NaN
g4224	g5566	g11823	g15197	0	2	NaN
g4247	g3730	g12289	g8332	0	2	NaN
g4291	g9035	g1242	g14711	0	2	NaN
g4312	g5016	g8758	g7367	0	2	NaN
g4313	g7006	g13068	g11935	0	2	NaN
g4377	g7331	g6175	g4182	0	2	NaN
g4425	g5113	g12452	g10342	0	2	NaN
g4635	g9367	g2863	g15318	0	2	NaN
g467	g4640	g7725	g10490	0	2	NaN
g4682	g1312	g11457	g3253	0	2	NaN
g4817	g8807	g14335	g16121	0	2	NaN
g485	g1624	g11782	g13566	0	2	NaN
g4942	g280	g4362	g3943	0	2	NaN
g4990	g2994	g6558	g10149	0	2	NaN
g5038	g766	g12716	g5466	0	2	NaN
g5169	g1049	g9111	g483	0	2	NaN
g519	g7217	g10876	g12369	0	2	NaN
g5192	g8165	g4338	g9932	0	2	NaN
g5397	g5331	g1637	g14290	0	2	NaN
g5636	g3620	g14194	g2188	0	2	NaN
g5676	g5828	g11915	g12310	0	2	NaN
g5703	g7949	g6921	g724	0	2	NaN
g5714	g3750	g10992	g14643	0	2	NaN
g572	g2533	g6293	g9551	0	2	NaN
g5776	g6846	g13029	g12966	0	2	NaN
g5858	g2224	g5350	g15312	0	2	NaN
g5862	g8433	g7781	g11464	0	2	AA7
g587	g3811	g3497	g1864	0	2	NaN
g5928	g7521	g7611	g3614	0	2	NaN
g5940	g5188	g7833	g8270	0	2	NaN
g6009	g3352	g11795	g10447	0	2	NaN
g6040	g6359	g10446	g13474	0	2	NaN
g6345	g2113	g11160	g7991	0	2	NaN
g6365	g5671	g8775	g12939	0	2	NaN
g6424	g826	g14314	g887	0	2	NaN
g6428	g1616	g10228	g9496	0	2	NaN
g6434	g1736	g1483	g2265	0	2	NaN
g6476	g7132	g11163	g15688	0	2	NaN
g6584	g9957	g12073	g11846	0	2	NaN
g6692	g981	g15055	g8800	0	2	NaN
g6786	g237	g4274	g15129	0	2	NaN
g6787	g238	g4273	g15128	0	2	NaN
g683	g3027	g4813	g12531	0	2	NaN
g686	g3432	g5190	g14595	0	2	NaN
g6933	g4924	g11728	g2451	0	2	NaN
g6997	g5983	g3787	g6443	0	2	NaN
g7036	g3245	g7756	g11310	0	2	NaN
g7108	g109	g3310	g9479	0	2	NaN
g7157	g5383	g9082	g11366	0	2	NaN
g724	g8804	g3432	g200	0	2	NaN
g7432	g7190	g8664	g4751	0	2	NaN
g7459	g2656	g3664	g3650	0	2	NaN
g7469	g3923	g1097	g9188	0	2	NaN
g7472	g2852	g14436	g15600	0	2	NaN
g763	g3311	g12145	g1262	0	2	NaN
g7638	g7690	g4310	g7788	0	2	NaN
g7671	g5301	g13904	g10384	0	2	NaN
g7797	g1989	g5778	g15254	0	2	NaN
g780	g7652	g11699	g5479	0	2	NaN
g7915	g425	g3351	g5911	0	2	NaN
g8006	g3298	g13258	g5506	0	2	NaN
g8066	g6350	g9928	g13358	0	2	NaN
g8093	g1925	g10141	g5307	0	2	NaN
g8097	g617	g10906	g650	0	2	NaN
g8100	g4058	g2602	g598	0	2	NaN
g8145	g787	g9235	g12165	0	2	NaN
g8148	g788	g12568	g12166	0	2	NaN
g834	g5035	g6565	g5034	0	2	NaN
g8428	g8419	g13777	g14001	0	2	NaN
g8434	g8461	g13532	g13048	0	2	NaN
g8458	g8725	g7784	g12718	0	2	NaN
g8660	g9263	g6204	g11719	0	2	NaN
g8749	g9963	g4091	g5724	0	2	NaN
g877	g6736	g14933	g11179	0	2	NaN
g8804	g856	g2980	g9742	0	2	NaN
g8905	g2842	g7724	g14303	0	2	NaN
g8954	g8984	g15009	g2591	0	2	NaN
g901	g9142	g14953	g4153	0	2	NaN
g1021	g6681	g952	g8432	0	1	NaN
g1063	g9923	g10848	g13744	0	1	NaN

g1128	g5428	g14497	g14346	0	1	NaN
g1153	g3277	g901	g8991	0	1	NaN
g1184	g6877	g5425	g12250	0	1	NaN
g1259	g6696	g11762	g6457	0	1	GH3
g1281	g557	g13417	g67	0	1	NaN
g1316	g4392	g9641	g13438	0	1	NaN
g143	g4325	g9718	g10860	0	1	NaN
g1431	g9202	g2854	g2879	0	1	NaN
g1449	g2866	g13676	g15099	0	1	NaN
g1472	g6344	g6254	g13877	0	1	NaN
g1576	g9421	g2919	g7249	0	1	NaN
g1614	g2535	g1198	g2421	0	1	NaN
g1638	g460	g2006	g10322	0	1	NaN
g1757	g9164	g4403	g14606	0	1	NaN
g186	g1130	g3021	g2009	0	1	NaN
g1886	g5420	g6432	g1842	0	1	NaN
g1954	g9329	g3400	g4681	0	1	NaN
g2018	g8994	g824	g11882	0	1	NaN
g204	g632	g4088	g8153	0	1	NaN
g2078	g609	g11002	g11468	0	1	NaN
g2339	g8352	g12371	g12674	0	1	NaN
g2462	g4075	g15464	g11593	0	1	NaN
g2472	g9929	g10851	g14103	0	1	NaN
g2547	g3938	g2324	g10904	0	1	NaN
g2562	g8123	g157	g5427	0	1	NaN
g2591	g6817	g10310	g1528	0	1	NaN
g2613	g3192	g3283	g3869	0	1	NaN
g2614	g6572	g11877	g8955	0	1	NaN
g2692	g9739	g6938	g2535	0	1	NaN
g2837	g6391	g584	g2173	0	1	NaN
g2907	g9959	g12075	g11844	0	1	NaN
g2910	g9589	g2289	g10889	0	1	NaN
g2981	g1588	g7973	g5353	0	1	NaN
g2989	g4797	g10550	g10648	0	1	NaN
g3053	g2299	g3077	g1149	0	1	NaN
g3060	g2919	g3416	g3822	0	1	NaN
g3104	g5497	g3629	g2290	0	1	NaN
g3134	g8562	g14767	g3589	0	1	NaN
g3146	g5556	g14765	g14473	0	1	NaN
g3170	g2398	g5295	g2623	0	1	NaN
g3175	g9359	g14660	g14794	0	1	NaN
g3207	g6838	g7017	g13257	0	1	NaN
g3293	g4348	g692	g2725	0	1	NaN
g3329	g5357	g4007	g2934	0	1	NaN
g336	g6884	g2127	g8822	0	1	NaN
g3408	g3451	g13407	g9879	0	1	NaN
g3453	g9450	g14409	g1194	0	1	NaN
g3473	g7588	g7273	g16093	0	1	NaN
g36	g394	g774	g7110	0	1	NaN
g3653	g216	g7873	g14059	0	1	NaN
g3697	g5997	g15051	g4919	0	1	NaN
g3881	g5253	g23	g15025	0	1	NaN
g4003	g8878	g3460	g9968	0	1	NaN
g4056	g1542	g1347	g1361	0	1	NaN
g4114	g2902	g5222	g13225	0	1	NaN
g4117	g5599	g9299	g2437	0	1	NaN
g4125	g7970	g899	g998	0	1	NaN
g4139	g361	g8104	g9970	0	1	NaN
g4155	g290	g3488	g12859	0	1	NaN
g417	g7563	g7136	g3901	0	1	NaN
g4245	g1881	g4414	g8330	0	1	NaN
g4269	g1363	g11881	g4168	0	1	NaN
g433	g9159	g17	g11460	0	1	NaN
g4356	g7622	g14632	g7761	0	1	NaN
g4472	g8589	g6698	g10693	0	1	NaN
g4501	g4852	g6080	g8317	0	1	CE10
g4517	g4107	g10404	g9941	0	1	NaN
g4575	g6820	g10314	g2953	0	1	NaN
g4794	g7946	g10199	g9826	0	1	NaN
g4926	g4398	g5448	g1082	0	1	NaN
g4951	g858	g12306	g4824	0	1	NaN
g5016	g3681	g10215	g11660	0	1	NaN
g5037	g765	g12715	g5465	0	1	NaN
g5044	g2685	g12838	g6911	0	1	NaN
g5104	g9410	g76	g7174	0	1	NaN
g5226	g599	g9567	g11876	0	1	NaN
g5230	g978	g9578	g4489	0	1	NaN
g5258	g3268	g10065	g3350	0	1	NaN
g5332	g6160	g10433	g11956	0	1	NaN
g5417	g9650	g457	g11638	0	1	NaN
g5432	g2444	g12499	g4629	0	1	NaN
g5495	g8768	g13182	g11224	0	1	NaN

g5526	g5027	g13449	g14116	0	1	NaN
g5566	g6113	g14295	g11256	0	1	NaN
g5568	g6255	g10663	g4315	0	1	NaN
g5593	g7257	g4281	g16197	0	1	NaN
g5600	g7105	g1877	g8719	0	1	NaN
g562	g5508	g3986	g9093	0	1	NaN
g5678	g9949	g821	g12728	0	1	NaN
g5968	g9897	g2229	g12867	0	1	NaN
g6024	g8235	g6359	g13629	0	1	NaN
g6144	g6714	g4907	g3038	0	1	NaN
g6152	g4482	g14755	g15810	0	1	NaN
g6162	g366	g1673	g838	0	1	NaN
g6164	g5728	g15166	g8812	0	1	NaN
g6261	g5189	g3434	g8269	0	1	NaN
g6338	g1608	g13132	g8796	0	1	NaN
g6347	g932	g6997	g16105	0	1	NaN
g640	g7507	g4669	g1039	0	1	NaN
g6401	g2019	g9628	g8263	0	1	NaN
g6425	g3998	g7966	g12314	0	1	NaN
g6444	g305	g2548	g9239	0	1	NaN
g6447	g175	g2550	g9237	0	1	NaN
g6464	g348	g12431	g2575	0	1	NaN
g6506	g1438	g8334	g5715	0	1	CE10
g652	g6959	g15102	g15575	0	1	NaN
g6541	g8462	g11101	g8210	0	1	NaN
g667	g2923	g10474	g7623	0	1	NaN
g6682	g4352	g963	g7791	0	1	CBM32
g6694	g1173	g15063	g3541	0	1	NaN
g6748	g3967	g6709	g10383	0	1	NaN
g6882	g637	g7410	g7520	0	1	NaN
g6916	g4130	g14512	g4221	0	1	NaN
g6955	g9687	g6923	g14803	0	1	NaN
g7017	g7183	g13297	g6880	0	1	NaN
g7037	g8450	g15303	g693	0	1	NaN
g7073	g1408	g580	g494	0	1	NaN
g7153	g2473	g13319	g5178	0	1	NaN
g7256	g739	g3803	g8349	0	1	NaN
g7261	g7134	g11165	g7111	0	1	NaN
g7315	g9965	g3619	g11841	0	1	NaN
g7336	g7110	g13229	g11380	0	1	NaN
g7359	g6598	g7058	g1726	0	1	NaN
g7561	g671	g13873	g16243	0	1	NaN
g7570	g6107	g3286	g2990	0	1	NaN
g7606	g4841	g14277	g10574	0	1	NaN
g7724	g1504	g8510	g9596	0	1	NaN
g7738	g414	g10840	g4884	0	1	NaN
g7746	g1825	g4653	g4395	0	1	NaN
g7800	g8713	g13529	g5884	0	1	NaN
g7884	g1520	g9586	g9567	0	1	NaN
g7977	g204	g9960	g353	0	1	NaN
g8104	g4435	g8139	g10324	0	1	NaN
g8121	g4988	g9202	g1224	0	1	NaN
g8188	g770	g11038	g16154	0	1	NaN
g8204	g9670	g2318	g8231	0	1	NaN
g8275	g6856	g15144	g15596	0	1	NaN
g8282	g8386	g2872	g14630	0	1	NaN
g8350	g9895	g14116	g658	0	1	NaN
g836	g779	g14072	g9407	0	1	NaN
g8454	g3662	g7645	g5252	0	1	NaN
g8585	g5288	g5714	g12049	0	1	NaN
g8656	g4422	g10436	g3388	0	1	NaN
g8685	g5461	g12647	g6487	0	1	NaN
g8778	g4460	g8602	g11866	0	1	NaN
g8803	g6228	g9661	g9556	0	1	NaN
g8807	g8294	g1804	g11970	0	1	NaN
g8885	g1014	g5465	g7471	0	1	NaN
g8914	g4938	g10381	g15816	0	1	GT8
g932	g6858	g9027	g14023	0	1	NaN
g1132	g7825	g6070	g6511	0	0	NaN
g1207	g3870	g8801	g7572	0	0	NaN
g1229	g5374	g9101	g6563	0	0	NaN
g1314	g9838	g9639	g13436	0	0	NaN
g1325	g4095	g864	g1418	0	0	NaN
g134	g8376	g13726	g7409	0	0	NaN
g1363	g2114	g11937	g15685	0	0	NaN
g1385	g4226	g13232	g12422	0	0	NaN
g1389	g5795	g13966	g743	0	0	NaN
g1486	g8317	g2471	g9447	0	0	NaN
g1523	g6631	g10442	g7920	0	0	NaN
g1525	g2744	g8889	g12013	0	0	NaN
g1649	g1177	g7259	g3302	0	0	NaN
g166	g6175	g5483	g10326	0	0	NaN

g1679	g9037	g8773	g14712	0	0	NaN
g176	g7023	g13928	g15421	0	0	NaN
g1825	g4833	g10478	g6891	0	0	NaN
g1855	g7182	g13296	g6881	0	0	NaN
g1888	g5421	g1897	g1843	0	0	NaN
g1904	g4583	g15187	g3619	0	0	NaN
g2003	g4237	g7402	g6207	0	0	NaN
g2016	g5366	g3551	g12014	0	0	NaN
g2113	g4639	g8613	g10487	0	0	NaN
g2199	g9218	g7947	g3021	0	0	NaN
g2247	g7131	g124	g8456	0	0	NaN
g2269	g1680	g3557	g7124	0	0	NaN
g2299	g3616	g12195	g1874	0	0	NaN
g2308	g6386	g2936	g10461	0	0	NaN
g2324	g147	g1486	g12976	0	0	NaN
g2341	g8350	g12369	g7557	0	0	NaN
g2348	g8518	g8312	g6999	0	0	NaN
g2407	g6786	g8054	g516	0	0	NaN
g2460	g4445	g10028	g13633	0	0	NaN
g2470	g1921	g337	g4553	0	0	NaN
g250	g3870	g8801	g7572	0	0	NaN
g2507	g68	g11073	g12731	0	0	NaN
g2559	g9596	g11093	g2637	0	0	NaN
g2595	g2854	g3492	g3744	0	0	NaN
g2646	g4712	g971	g2224	0	0	NaN
g2757	g7427	g6792	g563	0	0	NaN
g2789	g1601	g13128	g872	0	0	NaN
g281	g5269	g9465	g11541	0	0	NaN
g2840	g994	g9853	g10725	0	0	NaN
g285	g4165	g3268	g9001	0	0	NaN
g3171	g6236	g6374	g2626	0	0	NaN
g321	g6536	g5183	g337	0	0	NaN
g3317	g2678	g12006	g2386	0	0	NaN
g3360	g7924	g11034	g133	0	0	NaN
g3392	g9194	g104	g12724	0	0	NaN
g3426	g2578	g11901	g4164	0	0	NaN
g3455	g4913	g15121	g7640	0	0	NaN
g3498	g4005	g14416	g4798	0	0	NaN
g3571	g1659	g7953	g533	0	0	NaN
g3587	g3258	g7542	g11745	0	0	NaN
g3589	g2662	g8113	g7360	0	0	NaN
g3676	g4510	g10070	g3345	0	0	NaN
g3709	g5793	g13963	g2431	0	0	NaN
g3734	g2008	g1076	g4067	0	0	NaN
g3752	g746	g3808	g8907	0	0	NaN
g3785	g9819	g403	g4189	0	0	NaN
g3841	g8260	g4996	g15646	0	0	NaN
g3856	g6798	g190	g2561	0	0	NaN
g3971	g9484	g8432	g6569	0	0	NaN
g4050	g3211	g740	g16103	0	0	NaN
g4091	g8468	g6372	g9052	0	0	NaN
g4127	g4732	g3964	g10225	0	0	NaN
g4131	g5385	g384	g11368	0	0	NaN
g4140	g8879	g5166	g9967	0	0	NaN
g418	g7253	g10722	g2030	0	0	NaN
g4188	g9181	g1224	g7764	0	0	NaN
g4195	g8147	g10735	g3721	0	0	NaN
g4227	g3919	g10858	g4862	0	0	NaN
g4301	g6588	g5884	g7374	0	0	NaN
g4404	g2101	g7883	g1274	0	0	NaN
g4449	g5396	g2424	g537	0	0	NaN
g4471	g8584	g14829	g8538	0	0	NaN
g4508	g3561	g4378	g12734	0	0	NaN
g4878	g5907	g9379	g16044	0	0	NaN
g49	g2825	g14012	g14051	0	0	NaN
g4978	g2924	g10472	g12837	0	0	NaN
g5029	g775	g5027	g9580	0	0	NaN
g5049	g6319	g3045	g2850	0	0	NaN
g5064	g2469	g7969	g7177	0	0	NaN
g5123	g2394	g14285	g14526	0	0	NaN
g5140	g1316	g2183	g7890	0	0	NaN
g5149	g6238	g1472	g6246	0	0	NaN
g5177	g3706	g13472	g7305	0	0	NaN
g521	g4842	g6915	g13727	0	0	NaN
g5242	g9373	g9162	g16256	0	0	NaN
g5295	g877	g2511	g5242	0	0	NaN
g5333	g6630	g3894	g12256	0	0	NaN
g5528	g5026	g13450	g11771	0	0	NaN
g5533	g3958	g8210	g12343	0	0	NaN
g5543	g4089	g962	g8765	0	0	NaN
g5557	g2342	g1382	g7198	0	0	NaN
g5595	g338	g4280	g10049	0	0	NaN

g563	g5507	g3987	g9092	0	0	NaN
g5696	g363	g13372	g10612	0	0	NaN
g5702	g825	g13022	g12039	0	0	NaN
g5738	g988	g5679	g10158	0	0	NaN
g5880	g9773	g1518	g2844	0	0	NaN
g5882	g7494	g8535	g1607	0	0	NaN
g5887	g8857	g12347	g853	0	0	NaN
g5959	g4158	g317	g5771	0	0	NaN
g5978	g9345	g5059	g10154	0	0	NaN
g610	g2335	g6830	g6660	0	0	NaN
g6150	g8927	g12478	g8711	0	0	NaN
g616	g7388	g2450	g1633	0	0	NaN
g6171	g4536	g4486	g10023	0	0	NaN
g6245	g6897	g4927	g14953	0	0	NaN
g6275	g9147	g8698	g4161	0	0	NaN
g6308	g3584	g6663	g442	0	0	NaN
g6309	g3637	g4127	g7834	0	0	NaN
g6409	g9578	g13824	g10026	0	0	NaN
g6431	g2683	g11998	g6913	0	0	NaN
g6448	g176	g2551	g4890	0	0	CE10
g6488	g7751	g5807	g6402	0	0	NaN
g6514	g9674	g13857	g9498	0	0	NaN
g6650	g6004	g14975	g12909	0	0	NaN
g6687	g3492	g1732	g3748	0	0	NaN
g6709	g6738	g2783	g2185	0	0	NaN
g6761	g3150	g9559	g11396	0	0	NaN
g6770	g7413	g86	g15278	0	0	NaN
g6860	g9185	g9137	g6057	0	0	NaN
g688	g8361	g11500	g3925	0	0	NaN
g6885	g904	g11438	g2506	0	0	NaN
g689	g9780	g11770	g11523	0	0	NaN
g691	g8362	g11501	g3924	0	0	NaN
g6913	g789	g9232	g11093	0	0	NaN
g6931	g3162	g14061	g284	0	0	NaN
g6938	g2649	g8185	g13353	0	0	NaN
g6991	g6577	g99	g10639	0	0	NaN
g7	g3569	g2583	g7009	0	0	NaN
g7075	g1406	g582	g2015	0	0	NaN
g7091	g7779	g5251	g11280	0	0	NaN
g7102	g3104	g7876	g6607	0	0	NaN
g7104	g8629	g10218	g2698	0	0	NaN
g7117	g1381	g6748	g10404	0	0	NaN
g714	g8159	g8112	g8287	0	0	NaN
g720	g646	g7115	g10819	0	0	NaN
g7211	g1651	g4977	g5673	0	0	NaN
g7254	g533	g12763	g8906	0	0	NaN
g7259	g944	g11055	g8902	0	0	NaN
g7266	g9328	g3399	g9494	0	0	NaN
g7289	g8322	g5168	g9202	0	0	NaN
g7291	g7805	g9488	g2787	0	0	NaN
g73	g4433	g10325	g4643	0	0	NaN
g7325	g1797	g3615	g2652	0	0	NaN
g7421	g5164	g5352	g15314	0	0	NaN
g7426	g6353	g15351	g11652	0	0	NaN
g7498	g3924	g11722	g5020	0	0	NaN
g7505	g4186	g12591	g14248	0	0	NaN
g7542	g5941	g3335	g6303	0	0	NaN
g7554	g4526	g7627	g5912	0	0	NaN
g7575	g4768	g540	g16115	0	0	NaN
g7601	g4340	g2001	g13045	0	0	NaN
g7616	g10009	g12160	g15663	0	0	NaN
g7637	g7689	g7887	g7789	0	0	NaN
g764	g3312	g11357	g1261	0	0	NaN
g7729	g5271	g5234	g3010	0	0	NaN
g7810	g2133	g10789	g6189	0	0	NaN
g7828	g2018	g4406	g10787	0	0	NaN
g7838	g6086	g2917	g5001	0	0	NaN
g7850	g9921	g10845	g9218	0	0	NaN
g7860	g5416	g6436	g8874	0	0	NaN
g794	g9639	g12655	g11364	0	0	NaN
g8072	g6790	g5776	g6424	0	0	NaN
g8174	g9690	g12860	g4978	0	0	NaN
g8244	g2235	g6998	g6874	0	0	NaN
g8287	g4789	g2955	g7115	0	0	NaN
g8312	g9324	g14327	g12456	0	0	NaN
g8370	g3206	g10288	g9345	0	0	NaN
g8382	g6268	g8672	g13856	0	0	NaN
g8399	g7399	g6021	g5305	0	0	NaN
g8416	g5199	g8449	g7630	0	0	NaN
g8553	g9224	g13381	g2453	0	0	NaN
g8620	g8829	g4065	g4207	0	0	NaN
g8625	g6914	g12	g2664	0	0	NaN

g864	g8190	g555	g34	0	0	NaN
g8734	g7935	g3673	g289	0	0	NaN
g8750	g9748	g3260	g15377	0	0	NaN
g8827	g5974	g5432	g10755	0	0	NaN
g8865	g140	g2831	g12247	0	0	NaN
g8873	g1920	g339	g5524	0	0	NaN
g8901	g32	g13310	g6833	0	0	NaN
g8939	g7346	g4661	g2620	0	0	NaN
g8949	g9098	g12033	g10278	0	0	NaN
g8957	g540	g11557	g8579	0	0	NaN
g8971	g5797	g5373	g8278	0	0	NaN
g8972	g1948	g12606	g8275	0	0	NaN
g9012	g7029	g6822	g6745	0	0	NaN
g9030	g1475	g3217	g13525	0	0	NaN
g943	g1144	g4749	g6029	0	0	NaN

ANNEX B – SUPPLEMENTARY MATERIAL FOR CHAPTER 2

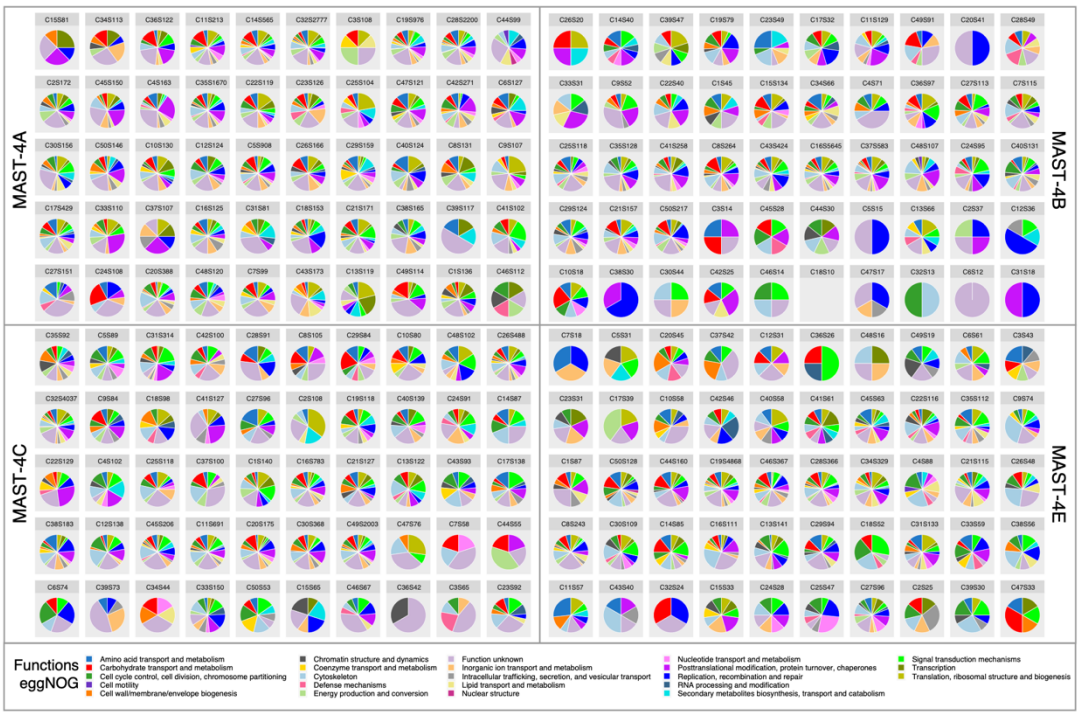

Figure 1. Functional annotation of genetic clusters for MAST-4 species using the eggNOG database. A total of 50 gene clusters were delineated based on similarities of dN/dS ratios across stations (see **Figure 2.3**). Note that gene cluster names are in the form of CXSY where X is the cluster number (1 to 50) and Y is the number of genes within the cluster. Genes without a hit in the database were not considered. A cluster without a pie chart indicates that no gene was found in the database.

Table 1. Metagenomic read samples from* Tara Oceans *expedition used in Chapter 2.

Table too large to fit, available on-line at:
https://doi.org/10.5281/zenodo.7078952

Table 2. Environmental metadata for* Tara Oceans *stations.

Station	labels	Latitude	Longitude	Temperature	Salinity	Density	Distance_coast	Chlorophyll_A	Depth	PAR	Samples	NO3	NO2	PO4	Si
TA_SUR_GGMM_4	TA4_NAO	36.563	-6.553	20.238088	36.599199	25.909909	18.5085	0.094049	9	14.834 81	4	NA	0.018 33	0.025 6	0.557 1
TA_SUR_GGMM_5	TA5_MS	36.03	-4.405	20.730517	37.051 35	26.105 1	54.8537	0.799237	9	21.106 998	5	NA	0	0.11	0.38
TA_SUR_GGMM_6	TA6_MS	36.529	-4.251	18.589717	37.212 883	26.799 7	22.3729	2.49515	9	1.2153 99	6	NA	NA	NA	NA
TA_SUR_GGMM_7	TA7_MS	37.0299	1.9575	23.8141	37.522	25.611 15	48.0551	0.065832	9	1.7943 69	7		0.044 5	0.047 5	0.793 5
TA_SUR_GGMM_9	TA9_MS	39.1633	5.916	24.4828	37.804 2	25.621	156.9683	0.0058	9	19.939 398	9	NA	0.01	0.02	0.75
TA_SUR_GGMM_11	TA11_MS	41.6663	2.7994	NA	NA	NA	NA	NA	9	NA	11	NA	NA	NA	NA
TA_SUR_GGMM_16	TA16_MS	37.398	15.454	20.858567	38.144 833	26.929 367	30.1662	0.0172	5	17.807 575	16	NA	0.02	0.04	0.67
TA_SUR_GGMM_18	TA18_MS	35.756	14.287	21.486 65	37.895 75	26.565 263	13.8783	0.0018	5	19.262 582	18	NA	0.018	0.026	0.56

TA_SUR_GGMM_20	TA20_MS	34.451	14.973	21.5215	38.4002	26.940675	149.2171	-0.0146	5	19.275779	20	NA	0.01	0.01	0.54
TA_SUR_GGMM_22	TA22_MS	39.729	17.4	17.298758	37.839767	27.623392	51.2518	0.0856	5	9.740976	22	NA	0.07	0.02	2.03
TA_SUR_GGMM_23	TA23_MS	42.176	17.729	17.211742	38.22805	27.9429	54.4621	0.0338	5	10.129158	23	NA	0.009125	0.012	1.372375
TA_SUR_GGMM_25	TA25_MS	39.333	19.421	18.319192	38.185333	27.633617	39.754	0.0366	5	4.662947	25	NA	0.00425	0.00525	1.0505
TA_SUR_GGMM_30	TA30_MS	33.929	32.789	20.4425	39.42311	28.0195	72.8349	0.01488	5	14.994186	30	0.188705	0.00125	0.00025	0.816
TA_SUR_GGMM_32	TA32_RS	23.391	37.254	25.811708	39.7709	26.698217	88.3462	0.0024	5	21.787103	32	0.279747	0.01075	0.00875	0.8775
TA_SUR_GGMM_34	TA34_RS	18.445	39.884	27.635025	38.64	25.2559	74.2982	0.066	5	2.864579	34	0	0.011	0.144	2.6405
TA_SUR_GGMM_36	TA36_IO	20.824	63.525	25.681369	36.525962	24.284594	415.5842	0.06795	5	17.265706	36	1.545287	0.0455	0.3575	1.143
TA_SUR_GGMM_38	TA38_IO	19.017	64.576	26.313595	36.618044	24.148331	584.8344	0.05004	5	23.003181	38	0.632675	0.107	0.337	1.278
TA_SUR_GGMM_39	TA39_IO	18.647	66.463	27.089113	36.310588	23.677262	464.6231	0.0258	5	26.627413	39	0	0.01	0.23	1.44
TA_SUR_GGMM_41	TA41_IO	14.582	70.011	29.151544	36.0499	22.799819	249.7827	0	5	34.540752	41	0	0.007	0.1565	1.3865
TA_SUR_GGMM_42	TA42_IO	5.992	73.919	30.12785	34.56775	21.358942	52.8424	0	5	NA	42	0	0.001	0.0765	2.171
TA_SUR_GGMM_43	TA43_IO	4.66	73.489	29.938975	34.493275	21.37115	1.773	0.0732	5	NA	43	0	0	0.07	1.82
TA_SUR_GGMM_45	TA45_IO	0.941	71.71	30.59345	35.0481	21.558975	153.5512	0.0036	5	NA	45	2.576	0.02	1.94	15.14
TA_SUR_GGMM_46	TA46_IO	-0.659	73.162	30.12475	35.1113	21.767 45	1.3428	0.048	5	34.954927	46	0.477	0	0.1	1.94
TA_SUR_GGMM_51	TA51_IO	-21.476	54.283	27.488383	34.939958	22.516108	102.9889	0	5	25.554368	51	0.244	0	0.09	1.28
TA_SUR_GGMM_52	TA52_IO	-17.023	53.508	27.956483	34.54545	22.067625	409.6259	0.0332	5	30.189334	52	0.532471	0	0.119	2.666
TA_SUR_GGMM_58	TA58_IO	-17.455	42.32	26.550545	35.1145	22.947 05	182.7374	0.02022	5	21.826547	58	NA	0.01	0	2.46
TA_SUR_GGMM_64	TA64_IO	-29.508	37.929	22.2414	35.32467	24.397 02	550.5186	0.0492	5	13.575859	64	0	0.004	0.084	1.7665
TA_SUR_GGMM_65	TA65_IO	-35.226	26.334	21.76357	35.449638	24.629613	85.2841	0.10548	5	0.86456	65	NA	NA	NA	NA
TA_SUR_GGMM_66	TA66_SAO	-34.905	18.016	15.013362	35.3233	26.2141	79.9944	0.1611	5	9.435831	66	2.426286	0.3025	0.343	2.7435
TA_SUR_GGMM_67	TA67_SAO	-32.292	17.206	13.04322	34.870735	26.284 07	81.2122	1.20489	5	0.068852	67	1.264607	0.171	1.016	13.88
TA_SUR_GGMM_68	TA68_SAO	-31.039	4.62	16.8632	35.686585	26.074 37	1113.2454	0.19332	5	24.301431	68	1.103865	0.201	0.1455	1.9865
TA_SUR_GGMM_70	TA70_SAO	-20.229	-3.413	19.770001	36.35839	25.859525	1161.2626	0.0456	5	22.209965	70	1.763081	0.0465	0.309	1.1485
TA_SUR_GGMM_72	TA72_SAO	-8.691	-18.006	25.064425	36.41678	24.393095	396.0494	0.00864	5	28.638751	72	0.416799	0.003	0.104	0.868
TA_SUR_GGMM_76	TA76_SAO	-21.029	-35.231	23.374825	37.076575	25.400 18	332.8446	-0.00072	5	31.396196	76	0.185971	0.001404	0.055883	0.814172
TA_SUR_GGMM_78	TA78_SAO	-30.158	-43.323	20.07775	36.32553	25.754115	554.4294	0.00144	5	41.250614	78	0.025333	0	0	0.749
TA_SUR_GGMM_80	TA80_SAO	-40.699	-51.952	19.934577	35.622685	25.255 28	64.5554	0.04392	5	32.686463	80	0	0	0	1.039
TA_SUR_GGMM_81	TA81_SAO	-44.497	-52.214	13.6805	34.793925	26.094 4	565.4919	0.099	5	16.7857	81	1.856	0.115	0.429	1.173
TA_SUR_GGMM_82	TA82_SAO	-47.165	-58.012	7.579417	34.05025	26.589567	835.6982	0.2258	5	NA	82	18.434187	0.146	1.302	1.945
TA_SUR_GGMM_83	TA83_SAO	-54.418	-65.023	7.12605	33.254213	26.026 55	468.833	0.1905	5	25.541 55	83	11.004	0.198	1.176	0.936
TA_SUR_GGMM_84	TA84_SO	-60.395	-60.471	1.9034	33.7198	26.9578	15.0017	0.0219	5	28.651 83	84	24.903191	0.266	1.723	16.55
TA_SUR_GGMM_85	TA85_SO	-62.176	-49.503	0.73159	34.327465	27.521 75	200.5774	0.12516	5	10.772055	85	29.733 83	0.1055	2.1055	80.55
TA_SUR_GGMM_86	TA86_SO	-64.309	-53.057	-0.494375	33.254413	26.718162	233.8466	0.0366	5	16.703 32	86	11.992	0.082	1.294	62.46
TA_SUR_GGMM_89	TA89_SPO	-57.764	-67.419	5.8264	34.0449	26.8197	29.5753	2.028	5	27.992267	89	NA	0.243	1.609	3.293
TA_SUR_GGMM_92	TA92_SPO	-33.69	-71.977	15.912325	34.3948	25.301 65	98.7436	4.9071	9	1.096275	92	NA	0.039	0.795	0.49

		7													
TA_SUR_GGMM_93	TA93_SPO	-33.762	-72.615	18.122955	34.31598	24.71804	41.3902	0.13782	5	18.154071	93	0	0.03975	0.52125	0.1445
TA_SUR_GGMM_95	TA95_SPO	-31.386	-93.986	22.4257	34.934	24.048488	744.7631	0.0255	5	36.208597	95	0.557	0	0.22	0.41
TA_SUR_GGMM_96	TA96_SPO	-29.655	-101.268	23.84533	35.76868	24.269535	593.1643	-0.00432	5	39.162501	96	0.149932	0	0.1575	0.5305
TA_SUR_GGMM_97	TA97_SPO	-28.169	-107.668	24.716415	36.049015	24.22107	418.9277	-0.00144	5	31.459258	97	0.274	0	0.12	0.57
TA_SUR_GGMM_98	TA98_SPO	-26.261	-110.992	25.18103	36.40163	24.346275	204.7463	-0.0072	5	30.95361	98	0	0.000563	0.1815	0.385875
TA_SUR_GGMM_100	TA100_SPO	-13.162	-96.283	25.349495	35.82876	23.86109	1404.6916	0.10404	5	NA	100	5.005302	0.137	0.67725	1.15025
TA_SUR_GGMM_102	TA102_SPO	-5.218	-85.27	24.958885	34.764535	23.176695	805.0024	0.16782	5	24.439054	102	13.010708	0.3285	1.02575	5.04425
TA_SUR_GGMM_106	TA106_SPO	0.037	-84.62	25.702575	34.51722	22.76179	412.6693	0.12024	5	NA	106	6.059	0.14	0.55	3.6
TA_SUR_GGMM_109	TA109_SPO	1.8	-84.545	27.497475	33.472445	21.408655	496.135	0.16428	5	NA	109	1.673601	0.0455	0.2865	1.7085
TA_SUR_GGMM_110	TA110_SPO	-1.913	-84.616	23.888136	35.012435	23.68419	512.0231	0.11724	5	25.413264	110	8.074076	0.31925	0.77275	3.8635
TA_SUR_GGMM_111	TA111_SPO	-16.932	-100.662	22.781833	35.98269	24.742485	1067.5202	0.0408	5	15.111047	111	3.149999	0.0415	0.5	1.0325
TA_SUR_GGMM_112	TA112_SPO	-23.22	-129.578	24.26287	36.43917	24.65312	653.4677	0	5	7.292033	112	0.167491	0.0075	0.14225	0.7425
TA_SUR_GGMM_113	TA113_SPO	-23.114	-134.92	23.791125	36.51495	24.851175	102.7922	0.0438	5	22.426681	113	0	0.01	0.18	0.8
TA_SUR_GGMM_122	TA122_SPO	-8.969	-139.338	26.60467	35.37032	23.122935	4.603	0.06984	5	28.45742	122	5.660372	0.11925	0.56795	2.1905
TA_SUR_GGMM_123	TA123_SPO	-8.879	-140.304	26.586685	35.3551	23.11712	2.4728	0.0924	5	NA	123	4.805128	0.14425	0.53495	2.2205
TA_SUR_GGMM_124	TA124_SPO	-8.999	-140.588	26.572419	35.384031	23.143631	36.1359	0.1494	5	24.609687	124	4.965553	0.1665	0.63025	2.53325
TA_SUR_GGMM_125	TA125_SPO	-8.89	-142.61	26.80604	35.42651	23.10143	227.3463	0.09801	5	NA	125	5.460123	0.1945	0.55725	1.80475
TA_SUR_GGMM_128	TA128_SPO	-0.469	-153.305	26.218438	35.129025	23.062688	449.3824	0.1431	5	29.350133	128	2.964091	0.2715	0.53975	2.668
TA_SUR_GGMM_129	TA129_NPO	6.732	-153.089	28.29544	34.74947	22.10997	388.3567	0.08916	5	NA	129	0	0.04	0.23	1.37
TA_SUR_GGMM_130	TA130_NPO	11.265	-152.462	27.5021	34.482	22.167425	921.0306	0.0075	5	NA	130	0	0.01	0.19	0.65
TA_SUR_GGMM_131	TA131_NPO	22.747	-158.052	26.275305	35.27298	23.153445	919.5766	0	5	34.262057	131	0	0	0.05	1.04
TA_SUR_GGMM_132	TA132_NPO	31.506	-159.014	25.164595	35.19025	23.434915	113.969	-0.00288	5	16.532129	132	0	0.002	0.0062	2.4278
TA_SUR_GGMM_135	TA135_NPO	32.983	-121.832	17.46236	33.468295	24.22939	215.5801	0.16416	5	7.813335	135	0	0.01	0.28	1.22
TA_SUR_GGMM_136	TA136_NPO	17.022	-118.914	24.612275	34.524725	23.0995	490.6032	0.0717	5	16.754029	136	0	0.01	0.31	1.8
TA_SUR_GGMM_137	TA137_NPO	14.161	-116.699	26.478605	33.84798	22.01524	453.4002	0.1386	5	22.246192	137	3.740793	0.067	0.456	2.538
TA_SUR_GGMM_138	TA138_NPO	6.215	-103.017	26.605515	33.52041	21.72899	10.2085	0.00588	5	NA	138	0.803918	0.003	0.1575	1.1855
TA_SUR_GGMM_139	TA139_NPO	6.491	-95.449	26.396617	33.205433	21.557358	665.1344	0.248	5	NA	139	2.892	0.1	0.53	3.17
TA_SUR_GGMM_142	TA142_NAO	25.602	-88.417	24.98456	36.177285	24.236425	398.5282	0.05556	5	18.336382	142	0	0.0195	0.003	1.218
TA_SUR_GGMM_143	TA143_NAO	29.885	-79.682	24.812165	36.21913	24.32047	88.8528	0.036	5	9.063101	143	1.19	0.02	0.01	1.26
TA_SUR_GGMM_144	TA144_NAO	36.369	-72.815	22.939295	36.39483	25.00969	45.2654	0.06012	5	11.529651	144	0	0.02	0	1.18
TA_SUR_GGMM_145	TA145_NAO	39.163	-70.076	13.993715	35.12152	26.282555	262.746	0.13044	5	1.097164	145	5.013141	0.093	0.33275	2.41425
TA_SUR_GGMM_146	TA146_NAO	34.731	-71.248	19.312775	36.513137	26.097587	353.0857	0.06129	5	NA	146	1.279722	0.2225	0.02	0.944
TA_SUR_GGMM_147	TA147_NAO	32.954	-66.533	20.150733	36.57475	25.923533	466.8747	0.0512	5	21.236944	147	0	0.22	0	0.8
TA_SUR_GGMM_148	TA148_NAO	31.782	-64.145	20.481195	36.596745	25.85157	172.4671	0.04404	5	21.500162	148	1.009924	0.093	0.0015	0.728

TA_SUR_GG MM_149	TA149_ NAO	34.09 8	-49.84	18.8255 4	36.412 74	26.146 925	749.564	0.08145	5	21.533 512	149	0.8080 32	0.25	0.06	0.8
TA_SUR_GG MM_150	TA150_ NAO	35.8	- 37.102	17.6271 42	36.285 292	26.349 408	1394.884	0.0782	5	16.225 385	150	0.1208 53	0.035 5	0.005 5	0.835 8
TA_SUR_GG MM_151	TA151_ NAO	36.19 4	- 29.801	17.3030 05	36.229 605	26.385 64	387.5094	0.0132	5	25.012 121	151	0.2960 16	0.019	0.012 5	0.654 5
TA_SUR_GG MM_152	TA152_ NAO	43.66 8	- 16.662	14.3167 5	35.991 67	26.886 34	614.5118	0.12096	5	9.2290 2	152	3.2092 19	0.310 5	0.159 5	1.236

Table 3. Number of variants and their effects by type for each MAST-4 species across all stations, both in counts and percentage. SNP – Single-Nucleotide Polymorphism; MNP – Multiple-Nucleotide Polymorphism; INS – Insertions; DEL – Deletions; MIXED – Mix of Multiple-Nucleotide Polymorphism and INDELs.

	MAST-4A		**MAST-4B**		**MAST-4C**		**MAST-4E**	
	Nº	**%**	**Nº**	**%**	**Nº**	**%**	**Nº**	**%**
Number of variants by type								
SNP	735,967	85.18	115,191	87.87	585,004	87.50	120,952	88.06
MNP	62,463	7.23	10,755	8.20	60,579	9.06	8,317	6.06
INS	29,676	3.43	2,066	1.58	9,898	1.48	3,148	2.29
DEL	26,765	3.10	2,545	1.94	10,843	1.62	4,277	3.11
MIXED	9,138	1.06	534	0.41	2,289	0.34	663	0.48
Number of effects by impact								
HIGH	3,778	0.14	353	0.08	1,463	0.07	439	0.10
LOW	287,944	10.89	61,316	14.00	333,785	15.20	38,496	8.62
MODERATE	273,784	10.36	32,464	7.41	173,654	7.91	32,342	7.24
MODIFIER	2,078,495	78.61	343,797	78.51	1,687,836	76.83	375,597	84.05
Number of effects by functional class								
MISSENSE	238,756	46.05	27,518	32.17	145,238	31.73	29,076	43.86
NONSENSE	765	0.15	79	0.09	316	0.07	50	0.08
SILENT	279,002	53.81	57,953	67.74	312,180	68.20	37,172	56.07
Number of effects by region								
DOWNSTREAM	932,929	35.29	164,630	37.59	817,223	37.20	167,049	37.38
EXON	562,897	21.29	93,864	21.43	507,454	23.10	70,884	15.86
INTERGENIC	264,479	10.00	33,604	7.67	141,793	6.46	59,028	13.21
INTRON	34,230	1.30	3,360	0.77	17,960	0.82	7,061	1.58
SPLICE _ACCEPTOR	307	0.01	32	0.01	136	0.01	39	0.01
SPLICE_DONOR	307	0.01	28	0.01	155	0.01	45	0.01
SPLICE_REGION	1,995	0.08	209	0.05	1,157	0.05	309	0.07
TRANSCRIPT	2	0.00	1	0.00	3	0.00	0	0.00
UPSTREAM	846,839	32.03	142,199	32.47	710,840	32.36	142,457	31.88
UTR_5_PRIME	16	0.00	3	0.00	17	0.00	2	0.00

Table 4. Summary of MAST-4 populations and sub-populations for each species. Average temperature (ºC) and salinity (psu) values are listed for each population with its standard deviation.

Species	**Population**	**Mean Temp (ºC)**	**Mean Salinity (psu)**	**Area**	**nº Stations**
MAST-4A	A1	20.93 ± 2.32	38.02 ± 2.32	Mediterranean Sea	10
	A2	20.16 ± 4.47	35.99 ± 4,47	Sub-tropical	32
	A3	21.76	35.45	South Africa	1

MAST-4B	A4	25.64 ± 1.61	34.71 ± 1.61	Tropical	7
	B1.1	25.72 ± 0.79	35.23 ± 0.79	Sub-tropical	2
	B1.2	28.68 ± 1.63	34.84 ± 1.63	Tropical	9
MAST-4C	C1	23.01 ± 2.00	37.46 ± 2.00	Mediterranean Sea	4
	C2	21.76	35.45	South Africa	1
	C3	29.99 ± 0.53	35.05 ± 0.53	Indian Ocean	5
	C4	26.01 ± 1.36	35.43 ± 1.36	Tropical	30
MAST-4E	E1.1	8.55 ± 3.50	34.04 ± 3.50	Sub-polar	4
	E1.2	17.60 ± 2.15	35.84 ± 2.15	sub-tropical	12

Table 5. Functional annotation of exclusive genes for each MAST-4 genomic population with eggNOG and CAZy databases.

Population	Gene	GH_ID	NOG_ID	eggNOG description
A1	g10053	<NA>	<NA>	
	g11688	<NA>	E111NSQ	
	g15265	<NA>	COG0846	NAD+ binding
	g1644	<NA>	<NA>	
	g1759	<NA>	<NA>	
	g2856	<NA>	<NA>	
	g3081	<NA>	<NA>	
	g6119	<NA>	E111IST	
	g6222	<NA>	E111FVP	
	g7137	<NA>	E10YZID	
	g8653	<NA>	COG0457	Function unknown
	g9050	<NA>	COG1020	D-alanine ligase activity
	g9791	<NA>	E10XQB4	
A2	g5512	<NA>	COG1227	inorganic diphosphatase activity
A3	g10170	GT74	E10ZUAF	
	g10436	<NA>	E111NN4	
	g11044	<NA>	E10YJ3R	
	g11328	<NA>	E10XRR8	
	g11641	<NA>	E10YBHR	
	g1179	<NA>	E111W2T	
	g12372	<NA>	COG2032	superoxide dismutase activity
	g12522	<NA>	<NA>	
	g12996	<NA>	<NA>	
	g13385	<NA>	COG5159	proteasome assembly
	g13571	<NA>	COG2271	monocarboxylate transporter
	g13635	<NA>	E10XP1S	
	g14016	<NA>	E10ZREK	
	g14073	<NA>	<NA>	

	g14146	<NA>	E10YUEN	
	g14181	<NA>	<NA>	
	g14576	<NA>	<NA>	
	g14955	<NA>	E10YEG6	
	g155	<NA>	COG5225	ribosomal large subunit export from nucleus
	g1727	<NA>	E111IF0	
	g3699	<NA>	E10Y6B4	
	g3762	<NA>	E111HJY	
	g3882	<NA>	COG0664	cyclic nucleotide binding
	g4295	<NA>	E111T0Q	
	g5242	<NA>	E10Y4IU	
	g529	<NA>	E10XR76	
	g5647	<NA>	E10ZD6K	
	g5866	<NA>	<NA>	
	g6027	<NA>	E10XQAQ	
	g6317	<NA>	E111JXB	
	g6961	<NA>	COG0518	GMP synthase (glutamine-hydrolyzing) activity
	g698	<NA>	E10YPTM	
	g7188	<NA>	COG5496	Function unknown
	g8038	<NA>	E10ZHVU	
	g8803	<NA>	COG0042	Catalyzes the synthesis of 5,6-dihydrouridine
	g9009	<NA>	E111NV5	
	g9178	<NA>	E10XPSE	
	g1063	<NA>	E10XP3K	
	g1268	<NA>	<NA>	
	g2117	<NA>	<NA>	
	g2254	<NA>	<NA>	
	g2413	<NA>	<NA>	
	g2950	<NA>	<NA>	
	g4109	<NA>	E111NSQ	
	g4282	<NA>	E111NSQ	
	g5161	<NA>	E111ZF3	
B1	g5464	<NA>	E1105IF	
	g5867	<NA>	E112AFJ	
	g6138	<NA>	E10ZBX3	
	g6309	<NA>	E10XQ2Y	
	g6623	<NA>	<NA>	
	g6720	<NA>	<NA>	
	g8226	<NA>	<NA>	
	g89	<NA>	COG5021	ubiquitin protein ligase activity
	g9059	<NA>	COG4642	regulation of ryanodine-sensitive calcium-release channel activity
	g9377	<NA>	COG5184	guanyl-nucleotide exchange factor activity

	g9626	<NA>	<NA>	
	g9757	<NA>	E10XQGY	
C1	g10592	<NA>	E10XV39	
	g13334	<NA>	<NA>	
	g13701	<NA>	<NA>	
	g1601	<NA>	<NA>	
C2	g10014	<NA>	E10XTFP	
	g11562	<NA>	COG0465	ATP-dependent zinc metallopeptidase
	g11995	<NA>	E11031H	
	g12250	<NA>	E10XR4D	
	g13884	<NA>	E10ZZ6Z	
	g14692	<NA>	E10XQIX	
	g15006	<NA>	<NA>	
	g15112	<NA>	E1116QF	
	g15873	<NA>	E111ZF3	
	g1675	<NA>	<NA>	
	g1909	<NA>	E10Z5DF	
	g2990	<NA>	E111FEZ	
	g3679	<NA>	COG0141	histidinol dehydrogenase activity
	g5948	<NA>	COG0227	**Translation, ribosomal structure and biogenesis**
	g6598	<NA>	E1100KJ	
	g6986	<NA>	E10Y7DR	
	g7275	<NA>	E10ZUTN	
	g7325	<NA>	E10ZFMC	
	g9378	<NA>	<NA>	
	g9778	<NA>	E10XVGS	
C3	g10873	<NA>	E10XTUD	
	g14203	<NA>	<NA>	
	g2262	<NA>	<NA>	
	g3784	<NA>	COG5126	Ca2 -binding protein (EF-Hand superfamily
	g4216	<NA>	COG4624	ron-sulfur cluster assembly
	g4450	<NA>	<NA>	
	g4968	<NA>	COG3968	glutamine synthetase
	g6180	<NA>	<NA>	
	g6236	<NA>	COG5104	mRNA splicing, via spliceosome
	g8267	<NA>	<NA>	
E1	g1166	<NA>	E10Y6RP	
	g1549	<NA>	E10XS0P	
	g1682	<NA>	<NA>	
	g177	<NA>	E111FFY	
	g2134	<NA>	E10XNQS	
	g2353	<NA>	E10Y6RP	

g289	<NA>	COG2939	PFAM Peptidase S10, serine carboxypeptidase
g3732	<NA>	E10XUCG	
g4082	GH18	COG3858	chitin binding
g4166	<NA>	COG0004	ammonium transporteR
g4459	<NA>	E10YUX1	
g468	<NA>	<NA>	
g4995	<NA>	E10XPDX	
g5070	<NA>	E111MDJ	
g5298	<NA>	E110362	
g5692	<NA>	<NA>	
g635	<NA>	E111NSQ	
g7028	<NA>	<NA>	
g7887	<NA>	<NA>	
g8216	<NA>	E10Z6C3	
g8240	<NA>	E10Y9H0	
g8322	<NA>	COG4889	Function unknown
g8983	<NA>	E10ZUNB	

ANNEX C – SUPPLEMENTARY MATERIAL FOR CHAPTER 3

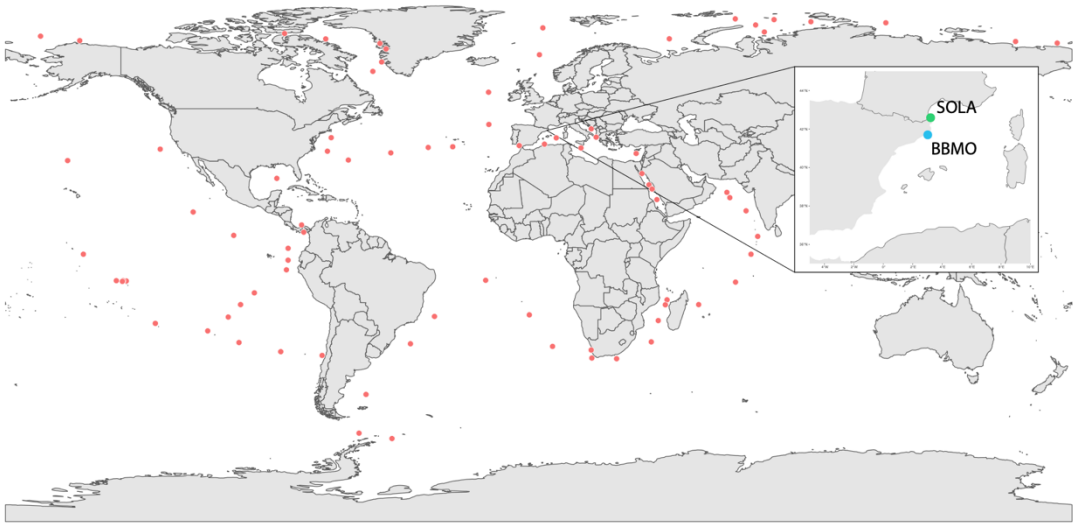

Figure 1. Locations of all Samples from BBMO (blue), SOLA (green), and TARA (red).

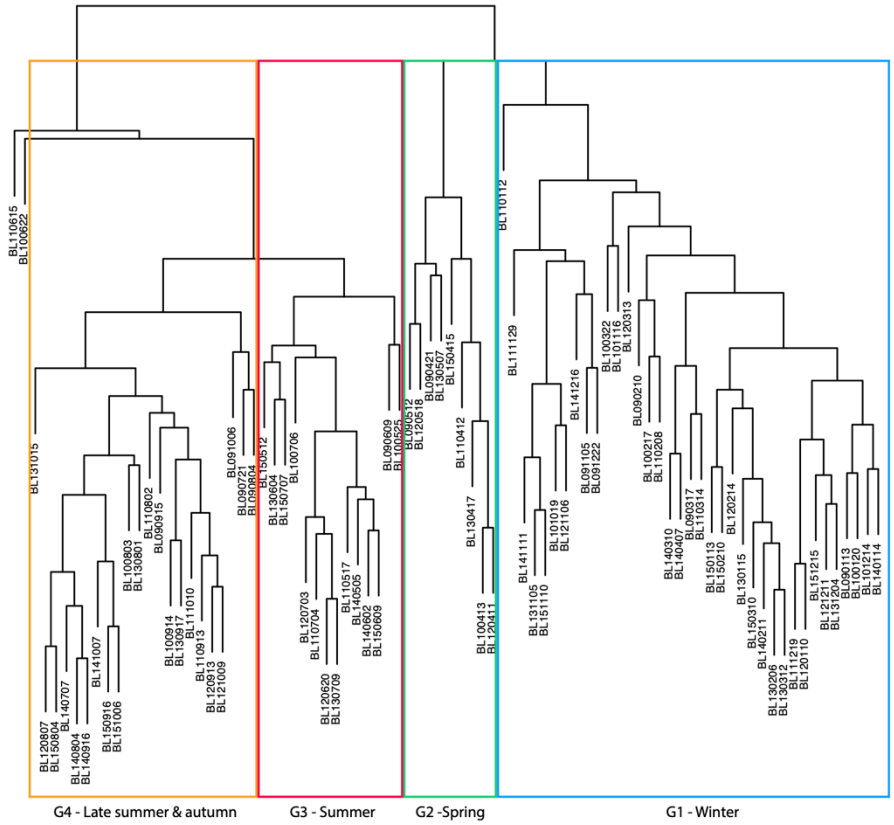

Figure 2. UPGMA clustering of Bray-Curtis dissimilarities among the 84 BBMO metagenomes calculated using SIMKA.

Table 1. Metadata for BBMO samples from January 2009 to December 2020.

Samples	Day_length	Temperature	Secchi	Salinity	Chla_total	Chla_3um	PO4	NH4	NO2	NO3	Si	year	month	day	season
BL090113	9.42	12.72	14	38.08	0.43	NA	0.086	0.411	0.186	1.501	2.499	2009	1	13	winter
BL090210	10.37	12.16	12	38.06	0.83	NA	NA	NA	NA	NA	NA	2009	2	10	winter
BL090317	11.95	13	15	38.18	NA	NA	0.049	0.327	0.033	0.287	2.066	2009	3	17	spring
BL090421	13.56	14.58	15.5	38.15	1.16	0.08	0.044	0.363	0.031	0.093	0.532	2009	4	21	spring
BL090512	14.4	17.55	12	37.76	0.54	0.27	0.048	0.406	0.024	0.098	0.735	2009	5	12	spring
BL090609	15.12	18.97	13	37.98	0.3	0.13	0.069	0.478	0.022	0.113	1.464	2009	6	9	sprng
BL090721	14.78	23.38	13	38.13	0.13	0.03	0.045	0.449	0.011	0.63	0.97	2009	7	21	summer
BL090804	14.33	24.3	20	38.17	0.34	0.18	0.029	0.099	0.028	0.346	1.042	2009	8	4	summer
BL090915	12.55	23.51	18	38.14	0.34	0.24	0.096	0.166	0.01	0.349	1.206	2009	9	15	autumn
BL091006	11.58	20.32	16	38.07	0.74	0.35	0.102	0.156	0.034	0.512	1.23	2009	10	6	autumn
BL091105	10.26	17.29	20	38.16	0.76	0.49	0.117	0.101	0.045	0.264	1.354	2009	11	5	autumn
BL091222	9.15	13.81	8	37.98	1.13	0.37	0.104	0.083	0.172	1.162	1.621	2009	12	22	winter
BL100120	9.61	12.92	16	37.93	1	0.73	0.108	0.067	0.231	2.794	1.902	2010	1	20	winter
BL100217	10.67	12.32	10	37.92	0.96	0.58	0.107	0.453	0.239	0	2.493	2010	2	17	winter
BL100322	12.19	12.39	9	37.61	1.95	0.82	0.131	0.145	0.107	1.496	1.565	2010	3	22	spring
BL100413	13.25	13.65	15	37.81	0.49	0.29	0.112	0.144	0.084	1.057	1.82	2010	4	14	spring
BL100525	14.81	15.19	8	37.63	0.93	0.37	0.091	1.151	0.14	0.43	1.561	2010	5	25	spring
BL100622	15.2	19.24	15	37.49	0.56	0.27	0.106	0.268	0.062	0.508	0.812	2010	6	22	summer
BL100706	15.1	23.05	10	37.74	0.43	0.18	0.123	0.056	0.005	0.082	0.304	2010	7	6	summer
BL100803	14.37	24.11	15	37.75	0.44	0.21	0.151	0.122	0.065	0.209	0.779	2010	8	3	summer
BL100914	12.59	24.43	22	38.03	0.31	0.1	0.118	0.201	0.025	0.127	0.617	2010	9	14	autumn
BL101019	10.99	17.61	11	37.86	0.78	0.38	0.139	0.225	0.055	0.574	1.331	2010	10	19	autumn
BL101116	9.85	15.76	8	38.17	0.53	0.35	0.076	0.144	0.15	0.421	0.874	2010	11	16	winter
BL101214	9.19	14.45	18	38.17	0.51	0.36	0.184	0.116	0.251	1.459	1.518	2010	12	14	winter
BL110112	9.39	12.89	12.5	38	0.81	0.52	0.112	0.616	0.275	1.539	2.141	2011	1	12	winter
BL110208	10.29	12.36	13.5	37.82	0.96	0.54	0.107	0.559	0.238	1.412	2.031	2011	2	8	winter
BL110314	11.81	12.25	8.5	37.82	1.05	0.54	0.11	0.408	0.283	1.232	2.458	2011	3	14	spring
BL110412	13.16	NA	16	NA	0.45	0.26	0.108	0.687	0.077	0.469	1.009	2011	4	12	spring
BL110517	14.57	17.6	18	NA	0.49	0.14	0.096	0.037	0.016	0.442	0.686	2011	5	17	spring
BL110615	15.18	21	13.5	37.7	0.54	0.18	0.182	0.629	0.102	1.007	1.854	2011	6	15	summer
BL110704	15.12	22.4	16	37.83	0.2	0.1	0.076	0.029	0.01	0.423	0.404	2011	7	5	summer
BL110802	14.4	22.88	15	37.82	0.6	0.38	0.105	0.687	0.038	0.8	1.806	2011	8	2	summer
BL110913	12.64	23	19	NA	0.32	0.19	0.092	0.095	0.004	0.493	0.9	2011	9	13	autumn
BL111010	11.4	20.64	14	38.09	0.37	0.2	0.083	0.026	0.006	0.511	1.02	2011	10	10	autumn
BL111129	9.46	16.76	8.5	37.21	2.88	0.36	0.101	0.052	0.06	0.546	0.303	2011	11	29	winter
BL111219	9.15	15.46	14	38.05	0.89	0.28	0.074	0.027	0.106	0.54	0.845	2011	12	19	winter
BL120110	9.35	14.39	14	38.09	0.69	0.36	0.078	1.389	0.263	0.188	1.218	2012	1	10	winter
BL120214	10.54	12.24	15	38.24	0.69	0.35	0.13	0.39	0.222	2.177	2.492	2012	2	14	winter
BL120313	11.81	13.18	18.5	38.28	0.67	0.36	0.103	0.666	0.259	2.384	2.451	2012	3	13	spring
BL120411	13.16	13.82	16	38.22	1.21	0.85	0.035	0.574	0.092	0.166	1.724	2012	4	11	spring
BL120518	14.63	16.8	16	38.24	0.36	0.04	0.036	0.913	0.085	0.114	0.502	2012	5	11	spring

BL120620	15.2	20.06	16	38.23	0.76	0.21	0.042	0.909	0.042	0.009	0.513	2012	6	20	summer
BL120703	15.13	22.01	16	38.07	0.26	0.12	0.053	0.531	0.061	0.064	0.508	2012	7	3	summer
BL120807	14.18	25.35	20	38.14	0.19	0.07	0.087	1.579	0.04	0.258	0.796	2012	8	7	summer
BL120913	12.59	22.46	19	38.16	0.16	0.08	0.027	0.611	0.016	0.069	1.052	2012	9	13	autumn
BL121009	11.4	18.31	18	37.91	0.34	0.16	0.026	0.832	0.036	0.038	0.932	2012	10	9	autumn
BL121106	10.18328661	16.66	8	38.04	0.46	0.26	0.221	1.702	0.122	3.523	2.036	2012	11	6	winter
BL121211	9.210932501	14.25	13	38	0.48	0.29	0.083	0.861	0.319	1.36	2.023	2012	12	11	winter
BL130115	9.466265572	13.27	17	38.1	0.89	0.43	0.124	1.174	0.401	1.212	1.564	2013	1	15	winter
BL130206	10.20611037	12.76	12	38.15	0.72	0.37	0.107	5.897	0.698	1.728	1.464	2013	2	6	winter
BL130312	11.71698304	12.78	9	38.14	1.08	0.42	0.164	0.177	0.219	3.289	2.475	2013	3	12	spring
BL130417	13.38122371	14.51	11	37.87	0.49	0.29	0.084	0.576	0.175	0.646	1.356	2013	4	17	spring
BL130507	14.21220887	14.96	19	37.73	0.75	0.16	0.097	0.669	0.213	1.374	1.451	2013	5	7	spring
BL130604	15.03462517	16.92	20	37.62	0.31	0.13	0.103	1.962	0.168	0.34	1.249	2013	6	4	spring
BL130709	14.97620219	21.61	14	37.87	0.5	0.17	0.088	0.853	0.109	0.134	0.456	2013	7	9	summer
BL130801	14.43787618	23.58	17	37.81	0.21	0.1	0.068	3.483	0.221	0.239	0.459	2013	8	1	summer
BL130917	12.45581855	23.41	20	38.12	0.31	0.15	0.082	0.125	0.061	0.316	0.712	2013	9	17	autumn
BL131015	11.16777513	21.4	20	38.05	0.28	0.16	0.077	0.433	0.099	0.218	0.816	2013	10	15	autumn
BL131105	10.26268825	18.06	18	38.04	0.42	0.18	0.066	0.316	0.11	0.316	0.81	2013	11	5	autumn
BL131204	9.179683856	14.42	8.5	38.13	0.6	0.3	0.173	0.474	0.375	1.53	1.878	2013	12	15	winter
BL140114	9.441090391	14.28	15	38.19	1.02	0.66	0.071	0.448	0.157	0.772	0.905	2014	1	14	winter
BL140211	10.41078106	13.4	8	NA	0.91	0.39	0.097	0.836	0.269	1.043	0.973	2014	2	11	winter
BL140310	11.62350701	13.76	15	37.92	0.6	0.24	0.085	0.199	0.128	0.425	0.768	2014	3	10	spring
BL140407	12.9298971	14.12	15.5	37.89	0.3	0.16	0.097	0.906	0.199	0.775	1.303	2014	4	7	spring
BL140505	14.13538227	16.22	16	37.85	0.29	0.15	0.046	0.156	0.239	0.104	0.986	2014	5	5	spring
BL140602	14.99582855	17.53	15	37.81	0.26	0.12	0.058	1.556	0.08	0.095	0.935	2014	6	2	summer
BL140707	15.08620564	21.99	14	37.89	0.29	0.09	0.099	0.838	0.027	7.286	0.915	2014	7	7	summer
BL140804	14.33191829	24.77	18	37.91	0.36	0.14	0.092	0.114	0.021	0.148	0.904	2014	8	4	autumn
BL140916	12.50182572	23	20	NA	0.17	0.12	0.082	0.3	0.014	0.114	0.881	2014	9	16	autumn
BL141007	11.5325262	21.3	12	NA	0.3	0.18	0.081	0.172	0.027	0.205	0.87	2014	10	7	autumn
BL141111	10.02970408	18.94	15	38.17	0.2	0.12	0.045	0.067	0.029	0.189	0.814	2014	11	11	winter
BL141216	9.171535126	15.88	11	37.78	0.44	0.18	0.074	0.028	0.315	0.811	1.233	2014	12	16	winter
BL150113	9.416821946	14.47	13	37.82	0.6	0.28	0.119	2.624	0.288	3.001	1.742	2015	1	13	winter
BL150210	10.36909419	13.05	13.5	38.09	0.7	0.25	0.085	2.013	0.345	1.603	1.395	2015	2	10	winter
BL150310	11.62350701	13.27	16.5	38.11	0.61	0.31	0.058	0.396	0.23	1.447	1.628	2015	3	10	spring
BL150415	13.29	14.38	16	37.95	0.33	0.22	0.035	0.497	0.166	0.849	1.805	2015	4	15	spring
BL150512	14.40	17.71	16	37.79	0.37	0.12	0.048	0.772	0.057	0.208	0.354	2015	5	12	spring
BL150609	15.12	18.97	18	37.7	0.33	0.13	0.037	0.205	0.053	0.232	0.672	2015	6	9	summer
BL150707	15.09	18.9	15	37.87	0.51	0.06	0.084	1.266	0.157	0.396	1.036	2015	7	7	summer
BL150804	14.33	25.48	18	38.02	0.23	0.12	0.051	0.291	0.023	0.209	0.77	2015	8	4	summer
BL150916	12.50	21.3	11.5	38.09	0.31	0.17	0.052	0.147	0.049	0.656	1.846	2015	9	16	autumn
BL151006	11.58	19.47	11	38.08	0.35	0.23	0.056	0.365	0.046	0.344	1.24	2015	10	6	autumn
BL151110	10.07	17.64	16	38.2	0.62	0.27	0.053	0.375	0.063	0.24	0.994	2015	11	10	winter
BL151215	9.18	15.38	18	38.27	0.6	0.37	0.058	0.352	0.175	0.893	1.756	2015	12	15	winter

BL160119	9.49	13.91	15	38.27	1.14	0.34	0.038	9.979	0.217	0.723	0.907	2016	1	19	winter
BL160210	10.37	13.61	14	38.28	1.52	0.32	0.066	0.815	0.206	0.794	0.911	2016	2	10	winter
BL160308	11.58	13.33	14	38.27	0.98	0.32	0.053	3.529	0.384	1.180	1.311	2016	3	8	spring
BL160406	12.88	13.56	16	38.15	0.55	0.38	0.048	1.726	0.309	0.797	1.223	2016	4	6	spring
BL160504	14.14	15.57	20	37.89	0.20	0.15	0.028	0.407	0.057	0.454	1.530	2016	5	4	spring
BL160607	15.10	20.11	17	38.03	0.20	0.1	0.022	0.547	0.036	0.083	0.453	2016	6	7	summer
BL160705	15.10	23.64	20	37.86	0.16	0.09	0.033	0.727	0.051	0.156	0.212	2016	7	5	summer
BL160802	14.37	24	18	NA	0.17	0.09	0.049	2.398	0.092	0.257	0.318	2016	8	2	summer
BL160913	12.59	22.72	20	37.93	0.27	0.15	0.025	1.518	0.084	0.187	0.499	2016	9	13	autumn
BL161018	10.99	19.09	16	38.16	0.35	0.32	0.033	2.208	0.115	0.976	0.874	2016	10	18	autumn
BL161108	10.11	17.74	17	38.25	0.44	0.19	0.030	0.539	0.107	0.123	0.751	2016	11	8	winter
BL161213	9.19	13.98	10	36.34	1.27	0.22	0.038	0.811	0.266	1.941	0.147	2016	12	13	winter
BL170124	9.73	12.99	5	38.1	0.43	0.14	0.125	0.286	0.335	1.386	1.782	2017	1	24	winter
BL170220	10.80	12.8	8	38.01	1.20	0.65	0.044	0.214	0.280	1.167	1.507	2017	2	20	winter
BL170314	11.81	13.53	10	38.07	1.13	0.72	0.020	0.295	0.042	0.065	1.696	2017	3	14	winter
BL170404	12.79	13.96	20	38.12	0.34	0.24	0.028	0.506	0.153	0.480	1.634	2017	4	4	spring
BL170509	14.29	16.81	13	38.06	0.35	0.18	0.032	0.291	0.039	0.148	0.837	2017	5	9	spring
BL170606	15.07	20.58	13	38.17	0.26	0.18	0.012	2.150	0.034	0.086	0.576	2017	6	6	summer
BL170704	15.13	23.13	20	38.02	0.13	0.06	0.015	0.431	0.036	0.034	0.690	2017	7	4	summer
BL170801	14.44	26.4	20	38.02	0.23	0.22	0.022	0.361	0.029	0.071	0.702	2017	8	1	summer
BL170913	12.64	23.98	20	37.81	0.19	0.13	0.025	0.350	0.039	0.054	0.800	2017	9	13	autumn
BL171010	11.40	22.16	19	37.88	0.36	0.17	0.021	0.764	0.043	0.277	0.634	2017	10	10	autumn
BL171106	10.22	19.54	19	37.7	0.46	0.17	0.025	0.200	0.040	0.155	0.663	2017	11	6	autumn
BL171212	9.21	13.89	9	38.35	0.34	0.10	0.155	0.502	0.217	3.616	2.569	2017	12	12	winter
BL180116	9.49	13.27	11	38.31	1.16	0.19	0.105	0.658	0.260	3.076	1.188	2018	1	16	winter
BL180213	10.50	12.75	15	38.24	0.34	0.21	0.082	0.256	0.315	2.927	1.543	2018	2	13	winter
BL180306	11.44	12.76	14	38.24	0.45	0.33	0.061	0.321	0.272	2.665	1.649	2018	3	6	spring
BL180410	13.07	12.55	12	37.76	0.52	0.36	0.039	0.767	0.235	1.977	1.189	2018	4	10	spring
BL180508	14.25	15.43	12	37.77	0.52	0.27	0.037	0.490	0.146	0.961	0.911	2018	5	8	spring
BL180613	15.16	21.02	12	37.57	0.27	0.12	0.040	0.137	0.114	0.238	0.263	2018	6	13	summer
BL180704	15.13	23.37	20	37.81	0.35	0.10	0.024	0.076	0.049	0.135	0.389	2018	7	4	summer
BL180801	14.44	25.77	19	38	0.17	0.10	0.034	0.336	0.059	0.176	0.617	2018	8	1	summer
BL180913	12.64	25.05	19	38.15	0.11	0.05	0.067	0.427	0.073	0.229	0.655	2018	9	13	autumn
BL181009	11.44	19.99	17	38.2	0.26	0.20	0.033	0.220	0.088	0.260	1.035	2018	10	9	autumn
BL181105	10.26	18.34	15	38.23	0.36	0.14	0.035	0.158	0.079	0.229	1.176	2018	11	5	autumn
BL181211	9.22	16.76	17	38	0.28	0.15	0.027	0.078	0.352	0.530	1.527	2018	12	11	winter
BL190115	9.47	14.23	16	38.19	0.50	0.20	0.026	3.798	0.197	1.185	1.515	2019	1	15	winter
BL190219	NA	12.92	13	38.3	1.84	1.18	0.026	0.403	0.187	1.318	0.724	2019	2	19	winter
BL190312	NA	13.85	17	38.23	0.57	0.25	0.039	0.508	0.111	5.325	1.300	2019	3	12	spring
BL190514	NA	15.4	12	38.19	0.39	0.17	0.031	7.542	0.197	0.403	0.973	2019	5	14	spring
BL190612	NA	18.05	16	38.08	0.33	0.16	0.022	4.923	0.127	0.258	0.953	2019	6	12	spring
BL190709	NA	20.09	16	38.09	0.31	0.11	0.018	0.295	0.027	0.237	0.623	2019	7	9	summer
BL190805	NA	25.16	16	37.88	0.41	0.18	0.029	0.316	0.032	0.135	0.480	2019	8	5	summer

BL190917	NA	24	20	NA	0.29	0.13	0.027	1.591	0.150	0.299	0.791	2019	9	17	summer
BL191203	NA	15.27	13	38.23	0.33	0.17	0.019	0.179	0.183	0.620	0.957	2019	12	3	winter
BL200130	NA	13.66	10	37.75	NA	NA	0.042	1.745	0.499	2.715	2.570	2020	1	30	winter
BL200211	NA	13.99	12	37.89	NA	NA	0.034	2.044	0.362	2.307	1.761	2020	2	11	winter
BL200310	NA	13.79	10	37.84	NA	NA	0.040	1.017	0.210	1.239	0.842	2020	3	10	winter
BL200512	NA	15.02	16	38.29	0.32	0.18	0.024	0.353	0.097	0.518	1.050	2020	5	12	spring
BL200609	NA	19.58	NA	38.14	0.07	0.04	0.022	0.512	0.051	0.122	0.440	2020	6	9	summer
BL200707	NA	24.91	16	38.42	0.13	0.05	0.017	2.261	0.158	0.361	0.579	2020	7	7	summer
BL200804	NA	26.72	17	37.97	0.44	0.17	0.074	0.733	0.107	1.592	1.221	2020	8	4	summer
BL200915	NA	24.03	20	37.86	0.20	0.07	0.030	0.142	0.005	0.500	0.302	2020	9	15	summer
BL201013	NA	19.81	19.5	38.06	0.19	0.10	0.031	0.230	0.023	0.447	0.738	2020	10	13	autumn
BL201110	NA	15.99	19.5	37.92	0.34	0.17	0.029	0.616	0.131	0.421	0.622	2020	11	10	winter
BL201215	NA	14.9	19.5	37.81	0.41	0.18	0.031	0.135	0.143	0.860	0.768	2020	12	15	winter

Table 2. Metadata for SOLA samples from January 2009 to December 2015.

Samples	month_n	year	year_site	date	site	month	Temperature	S	NH4	NO3	NO2	PO4	Chla_total	TotBaccells.mL	Syncell.mL	Day_length	season
SO090105	1	9	9_SO	5/1/09	SO	Jan	11.05	36.39	0.11	5.73	0.8	0.07	1.07	1260000	5570	9.2	winter
SO090209	2	9	9_SO	9/2/09	SO	Feb	10.56	36.88	0.37	3.31	0.29	0.03	1.03	975000	1290	10.33	winter
SO090323	3	9	9_SO	23/3/09	SO	Mar	12.06	37.83	0.39	1.54	0.09	0.08	0.76	1230000	28300	12.28	spring
SO090420	4	9	9_SO	20/4/09	SO	Apr	12.82	36.68	0.51	1.29	0.08	0.09	1.36	791000	7480	13.6	spring
SO090518	5	9	9_SO	18/5/09	SO	May	16.31	37.14	0.34	0.13	0.02	0.05	0.65	689000	10300	14.7	spring
SO090617	6	9	9_SO	17/6/09	SO	Jun	18.85	37.04	0.69	0.02	0.02	0.06	0.52	753000	42600	15.2	summer
SO090727	7	9	9_SO	27/7/09	SO	Jul	21.22	37.8	1.14	0.36	0.02	0.03	0.13	NA	NA	14.63	summer
SO090824	8	9	9_SO	24/8/09	SO	Aug	24.32	37.83	0.67	0.02	0.02	0.03	NA	115000	28800	13.52	autumn
SO090909	9	9	9_SO	9/9/09	SO	Sep	22.56	38.02	0.61	0.02	0.02	0.03	0.07	354000	21700	12.78	autumn
SO091013	10	9	9_SO	13/10/09	SO	Oct	20.84	38.13	0.29	0.05	0.02	0.03	0.3	763000	43600	11.18	autumn
SO091116	11	9	9_SO	16/11/09	SO	Nov	15.79	38.19	0.23	0.26	0.09	0.03	0.29	794000	16900	9.75	winter
SO091216	12	9	9_SO	16/12/09	SO	Dec	15.87	37.99	0.19	0.22	0.04	0.03	0.32	625000	6260	9.08	winter
SO100111	1	10	10_SO	11/1/10	SO	Jan	11.78	38.01	0.15	1.08	0.29	0.03	0.24	761000	3280	9.32	winter
SO100215	2	10	10_SO	15/2/10	SO	Feb	8.53	37.47	0.19	1.74	0.26	0.03	1.54	509000	1020	10.58	winter
SO100315	3	10	10_SO	15/3/10	SO	Mar	10.09	36.9	0.4	3.62	0.28	0.01	1.34	697000	3540	11.9	spring
SO100426	4	10	10_SO	26/4/10	SO	Apr	14.66	37.25	0.25	0.6	0.07	0.03	0.39	560000	28200	13.85	spring
SO100526	5	10	10_SO	26/5/10	SO	May	16.19	37.12	0.01	0.07	0.01	0.06	0.33	809000	37800	14.93	spring
SO100607	6	10	10_SO	7/6/10	SO	Jun	17.63	37.44	NA	0.1	0.02	0.02	0.16	769000	19300	15.18	spring
SO100705	7	10	10_SO	5/7/10	SO	Jul	19.06	37.54	0.01	0.06	0.01	0.01	0.14	719000	19900	15.2	summer
SO100802	8	10	10_SO	2/8/10	SO	Aug	20.98	37.76	0.01	0.06	0.01	0.01	0.09	692000	35400	14.23	summer
SO100913	9	10	10_SO	13/9/10	SO	Sep	19.39	38	0.03	0.09	0.01	0.01	0.14	566000	33400	12.6	autumn
SO101027	10	10	10_SO	27/10/10	SO	Oct	15.4	38.08	0.24	0.6	0.09	0.1	0.48	807000	8060	10.25	winter
SO101115	11	10	10_SO	15/11/10	SO	Nov	14.8	38.06	0.12	0.6	0.25	0.03	0.4	607000	13700	9.35	winter
SO101206	12	10	10_SO	6/12/10	SO	Dec	14.23	38.13	0.06	1.06	0.18	0.01	0.24	558000	8670	9.22	winter
SO110117	1	11	11_SO	17/1/11	SO	Jan	11.91	37.3	0.02	2.98	0.33	0.09	0.78	809000	5960	9.48	winter
SO110207	2	11	11_SO	7/2/11	SO	Feb	10.71	37.44	0.21	2.28	0.28	0.07	1.22	718000	3620	10.23	winter

SO110309	3	11	11_SO	9/3/11	SO	Mar	12.05	38.11	0.01	1.43	0.17	0.01	0.61	747000	24800	11.62	spring
SO110426	4	11	11_SO	26/4/11	SO	Apr	15.2	37.13	0.14	0.47	0.05	0.01	1.2	986000	14100	13.85	spring
SO110523	5	11	11_SO	23/5/11	SO	May	18.2	37.59	0.01	0.06	0.01	0.01	0.14	504000	14000	14.85	summer
SO110607	6	11	11_SO	7/6/11	SO	Jun	18.07	38.11	0.01	0.03	0.02	0.01	0.21	376000	22900	15.18	summer
SO110711	7	11	11_SO	11/7/11	SO	Jul	20.79	37.89	0.01	0	0.01	0.01	0.04	401000	14400	15.08	summer
SO110727	7	11	11_SO	27/7/11	SO	Jul	19.65	37.95	0.02	0.2	0.02	0.01	0.09	304000	21800	14.63	summer
SO110912	9	11	11_SO	12/9/11	SO	Sep	22.18	37.25	0.01	0.02	0.01	0.01	0.12	480000	41400	12.92	autumn
SO111011	10	11	11_SO	11/10/11	SO	Oct	21.41	38.08	0.01	0.05	0.01	0.09	0.12	441000	29600	11.6	autumn
SO111123	11	11	11_SO	23/11/11	SO	Nov	16.22	34.29	0.38	4.49	0.23	0.36	1.52	919000	5410	9.52	winter
SO111206	12	11	11_SO	6/12/11	SO	Dec	16.32	37.81	0.04	0.23	0.09	0.04	1.05	NA	NA	9.22	winter
SO120103	1	12	12_SO	3/1/12	SO	Jan	13.8	38.13	0.02	0.61	0.18	0.03	0.68	NA	NA	9.17	winter
SO120131	1	12	12_SO	31/1/12	SO	Jan	11.82	38.04	0.02	1.59	0.41	0.04	0.29	806000	7430	9.95	winter
SO120221	2	12	12_SO	21/2/12	SO	Feb	10.48	38.16	0.01	2.28	0.16	0.04	0.51	525000	5250	10.85	winter
SO120307	3	12	12_SO	7/3/12	SO	Mar	10.91	38.2	0.09	1.06	0.14	0.05	0.8	648000	3250	11.55	spring
SO120313	3	12	12_SO	13/3/12	SO	Mar	11.19	38.19	0.1	0.53	0.15	0.02	0.6	473000	1240	11.85	spring
SO120404	4	12	12_SO	4/4/12	SO	Apr	13.84	37.95	0.04	0.43	0.06	0.03	0.32	680000	12600	12.9	spring
SO120423	4	12	12_SO	23/4/12	SO	Apr	13.25	38.23	0.02	1.07	0.14	0.06	0.33	765000	52900	13.77	spring
SO120509	5	12	12_SO	9/5/12	SO	May	15.45	36.61	0.01	0.94	0.14	0.05	1.8	317000	5040	14.42	spring
SO120607	6	12	12_SO	7/6/12	SO	Jun	19.5	37.64	0.01	0.23	0.01	0.03	0.14	313000	7510	15.2	summer
SO120712	7	12	12_SO	12/7/12	SO	Jul	20.09	37.86	0.01	0.02	0.02	0.03	0.12	462000	18300	15.03	summer
SO120806	8	12	12_SO	6/8/12	SO	Aug	21.75	38.01	0.09	0.06	0.01	0.02	0.16	550000	21100	14.23	summer
SO120820	8	12	12_SO	20/8/12	SO	Aug	22.87	38.18	0.06	0.12	0.01	0.02	0.2	479000	17700	13.65	autumn
SO121022	10	12	12_SO	22/10/12	SO	Oct	18.24	38.15	0.27	0.29	0.07	0.05	0.48	472000	5520	10.72	autumn
SO121105	11	12	12_SO	5/11/12	SO	Nov	16.58	37.87	0.2	0.62	0.11	0.07	0.57	478000	6100	10.13	winter
SO121119	11	12	12_SO	19/11/12	SO	Nov	15.41	37.89	0.31	0.91	0.18	0.06	0.36	506000	9060	9.62	winter
SO121212	12	12	12_SO	12/12/12	SO	Dec	13.09	38.02	0.01	1.39	0.24	0.04	0.48	562000	5760	9.12	winter
SO130115	1	13	13_SO	15/1/13	SO	Jan	12.71	37.63	0.02	0.94	0.27	0.06	1.15	580000	5080	9.43	winter
SO130204	2	13	13_SO	4/2/13	SO	Feb	11.08	38.05	0.05	1.79	0.3	0.06	1.2	549000	4200	10.12	winter
SO130311	3	13	13_SO	11/3/13	SO	Mar	11.47	34.71	0.1	5.9	0.24	0.18	2.57	667000	1180	11.7	spring
SO130422	4	13	13_SO	22/4/13	SO	Apr	13.13	37.11	0.19	1.81	0.21	0.04	0.47	767000	23700	13.68	spring
SO130506	5	13	13_SO	6/5/13	SO	May	13.92	37.34	0.07	1.6	0.17	0.05	0.18	348000	26500	14.28	spring
SO130603	6	13	13_SO	3/6/13	SO	Jun	14.96	37.76	0.03	0.24	0.03	0.06	1.41	1020000	20300	15.12	spring
SO130701	7	13	13_SO	1/7/13	SO	Jul	19.14	37.94	0.02	0.06	0.01	0.03	0.46	740000	20400	15.25	summer
SO130826	8	13	13_SO	26/8/13	SO	Aug	22.57	37.89	0.05	0.16	0.01	0.01	0.14	623000	19400	13.43	autumn
SO130923	9	13	13_SO	23/9/13	SO	Sep	19.03	38.36	0.06	0.02	0.01	0.02	0.14	NA	NA	12.12	autumn
SO131028	10	13	13_SO	28/10/13	SO	Oct	19.05	36.76	0.53	1.62	0.05	0.04	1.05	1140000	9190	10.5	autumn
SO131113	11	13	13_SO	13/11/13	SO	Nov	16.47	37.95	0.08	0.17	0.01	0.04	0.95	984000	14500	9.85	winter
SO131212	12	13	13_SO	12/12/13	SO	Dec	12.72	38.19	0.05	2.49	0.2	0.09	0.51	585000	8760	9.12	winter
SO140113	1	14	14_SO	13/1/14	SO	Jan	12.44	34.28	0.13	9.52	0.69	0.08	1.75	2570000	11900	9.37	winter
SO140224	2	14	14_SO	24/2/14	SO	Feb	12.66	37.95	0.09	1.73	0.35	0.06	0.74	590000	3640	11	spring
SO140324	3	14	14_SO	24/3/14	SO	Mar	12.76	37.68	0.22	0.73	0.15	0.01	1.09	582000	1450	12.33	spring
SO140407	4	14	14_SO	7/4/14	SO	Apr	13.52	37.07	0.16	1.06	0.14	0.03	2.53	962000	8820	13	spring
SO140422	4	14	14_SO	22/4/14	SO	Apr	15.16	37.4	0.02	0.11	0.05	0.01	1.22	485000	18500	13.68	spring

SO140519	5	14	14_SO	19/5/14	SO	May	16.02	37.77	0.11	0.12	0.05	0.03	0.39	386000	20300	14.73	spring
SO140610	6	14	14_SO	10/6/14	SO	Jun	17.7	37.73	0.01	0.02	0.01	0.01	0.33	717000	29800	15.23	spring
SO140721	7	14	14_SO	21/7/14	SO	Jul	20.5	37.88	0.05	0.2	0.01	0.01	0.27	573000	31500	14.83	summer
SO140804	8	14	14_SO	4/8/14	SO	Aug	21.99	37.83	0.02	0.17	0.01	0.01	0.26	853000	50800	14.35	summer
SO140901	9	14	14_SO	1/9/14	SO	Sep	21.67	37.94	0.01	0.03	0.08	0.01	0.15	766000	34400	13.17	autumn
SO141112	11	14	14_SO	12/11/14	SO	Nov	18.23	38.08	0.09	0.3	0.11	0.01	0.3	NA	NA	9.88	autumn
SO141124	11	14	14_SO	24/11/14	SO	Nov	17.33	37.67	0.44	0.66	0.07	0.17	0.6	NA	NA	9.5	autumn
SO141208	12	14	14_SO	8/12/14	SO	Dec	16.15	37.53	0.24	1.37	0.31	0.04	0.33	NA	NA	9.17	winter
SO150108	1	15	15_SO	8/1/15	SO	Jan	13.27	37.79	0.05	NA	NA	NA	0.69	NA	6090	9.25	winter
SO150122	1	15	15_SO	22/1/15	SO	Jan	12.67	37.83	0.08	NA	NA	NA	1.1	NA	6880	9.63	winter
SO150202	2	15	15_SO	2/2/15	SO	Feb	12.6	38.06	0.09	1.45	0.22	0.02	0.64	NA	6060	10.03	winter
SO150309	3	15	15_SO	9/3/15	SO	Mar	11.46	37.83	0.034	1.161	0.27	0.01	1.495	NA	6320	11.62	spring
SO150413	4	15	15_SO	13/4/15	SO	Apr	13.8	37.8	0.043	1.069	0.245	0.038	0.667	1107149.25	68600	13.28	spring
SO150511	5	15	15_SO	11/5/15	SO	May	17.59	36.39	0.036	0.02	0.022	0.086	0.511	NA	12000	14.47	spring
SO150608	6	15	15_SO	8/6/15	SO	Jun	NA	NA	0.085	0.02	0.01	0.016	0.457	737000	42500	15.2	spring
SO150803	8	15	15_SO	3/8/15	SO	Aug	22.17	37.88	0.013	0.02	0.01	0.01	0.202	601209.875	43700	14.38	summer
SO150831	8	15	15_SO	31/8/15	SO	Aug	20.84	37.88	0.008	0.041	0.01	0.01	0.262	740270	55600	13.22	autumn
SO150921	9	15	15_SO	21/9/15	SO	Sep	20.04	37.77	0.09	0.066	0.01	0.013	0.37	637714.25	30000	12.22	autumn
SO151005	10	15	15_SO	5/10/15	SO	Oct	18.81	37.9	0.119	NA	0.01	0.04	0.66	920847.0625	46202.94141	10.9	autumn
SO151109	11	15	15_SO	9/11/15	SO	Nov	17.66	37.57	0.194	0.39	0.05	0.07	1.62	1276370.75	13249.08	9.5	autumn
SO151214	12	15	15_SO	14/12/15	SO	Dec	14.87	38.25	0.092	0.27	0.08	0.02	0.46	NA	NA	8.43	winter

Table 3. Accession numbers for *Tara Oceans* metagenomic samples.

PANGAEA sample id	BioSamples_ID	ENA_ID	MetaG/MetaT	Station	Layer	Size_fraction	Polar	Sample ID (registered at the BioSamples ...)	Sample ID (registered at the European Nu...)	Date/Time	Latitude	Longitude	Depth, nominal	ENA_Run_ID	Shortcut
TARA_Y20000002	SAMEA2619388	ERS487911	MetaG	4	SUR	0.2-1.6	Non polar	SAMEA2619388	ERS487911	2009-09-15T11:30:00Z	36.5533	-6.5669	9	ERR598955	ERR598
TARA_Y20000002	SAMEA2619388	ERS487911	MetaG	4	SUR	0.2-1.6	Non polar	SAMEA2619388	ERS487911	2009-09-15T11:30:00Z	36.5533	-6.5669	9	ERR599003	ERR599
TARA_A20000113	SAMEA2591057	ERS477931	MetaG	7	SUR	0.2-1.6	Non polar	SAMEA2591057	ERS477931	2009-09-23T12:50:00Z	37.051	1.9378	9	ERR315857	ERR315
TARA_X00000950	SAMEA2619531	ERS488119	MetaG	9	SUR	0.2-1.6	Non polar	SAMEA2619531	ERS488119	2009-09-28T12:18:00Z	39.1633	5.916	9	ERR594288	ERR594
TARA_X00000950	SAMEA2619531	ERS488119	MetaG	9	SUR	0.2-1.6	Non polar	SAMEA2619531	ERS488119	2009-09-28T12:18:00Z	39.1633	5.916	9	ERR594316	ERR594
TARA_X00000950	SAMEA2619531	ERS488119	MetaG	9	SUR	0.2-1.6	Non polar	SAMEA2619531	ERS488119	2009-09-28T12:18:00Z	39.1633	5.916	9	ERR594317	ERR594
TARA_A10000164	SAMEA2619667	ERS488330	MetaG	18	SUR	0.2-1.6	Non polar	SAMEA2619667	ERS488330	2009-11-02T08:13:00Z	35.759	14.2574	5	ERR598993	ERR598
TARA_A10000164	SAMEA2619667	ERS488330	MetaG	18	SUR	0.2-1.6	Non polar	SAMEA2619667	ERS488330	2009-11-02T08:13:00Z	35.759	14.2574	5	ERR599140	ERR599
TARA_E50000075	SAMEA2591084	ERS477979	MetaG	23	SUR	0.2-1.6	Non polar	SAMEA2591084	ERS477979	2009-11-18T08:41:00Z	42.2038	17.715	5	ERR315858	ERR315
TARA_E50000075	SAMEA2591084	ERS477979	MetaG	23	SUR	0.2-1.6	Non polar	SAMEA2591084	ERS477979	2009-11-18T08:41:00Z	42.2038	17.715	5	ERR315861	ERR315
TARA_E50000178	SAMEA2619766	ERS488486	MetaG	25	SUR	0.2-1.6	Non polar	SAMEA2619766	ERS488486	2009-11-23T09:12:00Z	39.3888	19.3905	5	ERR598951	ERR598
TARA_E50000178	SAMEA2619766	ERS488486	MetaG	25	SUR	0.2-1.6	Non polar	SAMEA2619766	ERS488486	2009-11-23T09:12:00Z	39.3888	19.3905	5	ERR599043	ERR599
TARA_A10001015	SAMEA2591108	ERS478017	MetaG	30	SUR	0.2-1.6	Non polar	SAMEA2591108	ERS478017	2009-12-15T10:41:00Z	33.9179	32.898	5	ERR315862	ERR315
TARA_A10001015	SAMEA2591108	ERS478017	MetaG	30	SUR	0.2-1.6	Non polar	SAMEA2591108	ERS478017	2009-12-15T10:41:00Z	33.9179	32.898	5	ERR315863	ERR315
TARA_A10001388	SAMEA2619808	ERS488551	MetaG	31	SUR	0.2-1.6	Non polar	SAMEA2619808	ERS488551	2010-01-09T07:15:00Z	27.16	34.835	5	ERR598969	ERR598
TARA_A10001388	SAMEA2619808	ERS488551	MetaG	31	SUR	0.2-1.6	Non polar	SAMEA2619808	ERS488551	2010-01-09T07:15:00Z	27.16	34.835	5	ERR599106	ERR599
TARA_A10001035	SAMEA2619818	ERS488569	MetaG	32	SUR	0.2-1.6	Non polar	SAMEA2619818	ERS488569	2010-01-11T07:21:00Z	23.36	37.2183	5	ERR599041	ERR599

TARA_A100001035	SAMEA2619818	ERS488569	MetaG	32	SUR	0.2-1.6	Non polar	SAMEA2619818	ERS488569	2010-01-11T07:21:00Z	23.36	37.2183	5	ERR599116	ERR599
TARA_A100001035	SAMEA2619818	ERS488569	MetaG	32	SUR	0.2-1.6	Non polar	SAMEA2619818	ERS488569	2010-01-11T07:21:00Z	23.36	37.2183	5	ERR599155	ERR599
TARA_A100001234	SAMEA2619857	ERS488621	MetaG	33	SUR	0.2-1.6	Non polar	SAMEA2619857	ERS488621	2010-01-13T07:16:00Z	21.9467	38.2517	5	ERR599049	ERR599
TARA_A100001234	SAMEA2619857	ERS488621	MetaG	33	SUR	0.2-1.6	Non polar	SAMEA2619857	ERS488621	2010-01-13T07:16:00Z	21.9467	38.2517	5	ERR599134	ERR599
TARA_B100000003	SAMEA2619879	ERS488649	MetaG	34	SUR	0.2-1.6	Non polar	SAMEA2619879	ERS488649	2010-01-20T04:27:00Z	18.3967	39.875	5	ERR598959	ERR598
TARA_B100000003	SAMEA2619879	ERS488649	MetaG	34	SUR	0.2-1.6	Non polar	SAMEA2619879	ERS488649	2010-01-20T04:27:00Z	18.3967	39.875	5	ERR598991	ERR598
TARA_Y100000022	SAMEA2619936	ERS488723	MetaG	36	SUR	0.2-1.6	Non polar	SAMEA2619936	ERS488723	2010-03-12T06:06:00Z	20.8183	63.5047	5	ERR598966	ERR598
TARA_Y100000022	SAMEA2619936	ERS488723	MetaG	36	SUR	0.2-1.6	Non polar	SAMEA2619936	ERS488723	2010-03-12T06:06:00Z	20.8183	63.5047	5	ERR599143	ERR599
TARA_Y100000287	SAMEA2620005	ERS488804	MetaG	38	SUR	0.2-1.6	Non polar	SAMEA2620005	ERS488804	2010-03-15T03:35:00Z	19.0393	64.4913	5	ERR599102	ERR599
TARA_Y100000287	SAMEA2620005	ERS488804	MetaG	38	SUR	0.2-1.6	Non polar	SAMEA2620005	ERS488804	2010-03-15T03:35:00Z	19.0393	64.4913	5	ERR599158	ERR599
TARA_B100000282	SAMEA2620194	ERS489043	MetaG	41	SUR	0.2-1.6	Non polar	SAMEA2620194	ERS489043	2010-03-30T02:47:00Z	14.6059	69.9776	5	ERR599011	ERR599
TARA_B100000282	SAMEA2620194	ERS489043	MetaG	41	SUR	0.2-1.6	Non polar	SAMEA2620194	ERS489043	2010-03-30T02:47:00Z	14.6059	69.9776	5	ERR599074	ERR599
TARA_B100000123	SAMEA2620230	ERS489087	MetaG	42	SUR	0.2-1.6	Non polar	SAMEA2620230	ERS489087	2010-04-04T02:47:00Z	6.0001	73.8955	5	ERR599075	ERR599
TARA_B100000123	SAMEA2620230	ERS489087	MetaG	42	SUR	0.2-1.6	Non polar	SAMEA2620230	ERS489087	2010-04-04T02:47:00Z	6.0001	73.8955	5	ERR599141	ERR599
TARA_B100000161	SAMEA2620339	ERS489236	MetaG	45	SUR	0.2-1.6	Non polar	SAMEA2620339	ERS489236	2010-04-13T03:21:00Z	0.0033	71.6428	5	ERR599045	ERR599
TARA_B100000161	SAMEA2620339	ERS489236	MetaG	45	SUR	0.2-1.6	Non polar	SAMEA2620339	ERS489236	2010-04-13T03:21:00Z	0.0033	71.6428	5	ERR599054	ERR599
TARA_B100000242	SAMEA2620404	ERS489315	MetaG	48	SUR	0.2-1.6	Non polar	SAMEA2620404	ERS489315	2010-04-19T07:56:00Z	-9.3921	66.4228	5	ERR599019	ERR599
TARA_B100000242	SAMEA2620404	ERS489315	MetaG	48	SUR	0.2-1.6	Non polar	SAMEA2620404	ERS489315	2010-04-19T07:56:00Z	-9.3921	66.4228	5	ERR599138	ERR599
TARA_B100000212	SAMEA2620542	ERS489529	MetaG	52	SUR	0.2-1.6	Non polar	SAMEA2620542	ERS489529	2010-05-17T04:10:00Z	-16.957	53.9801	5	ERR599098	ERR599
TARA_B100000212	SAMEA2620542	ERS489529	MetaG	52	SUR	0.2-1.6	Non polar	SAMEA2620542	ERS489529	2010-05-17T04:10:00Z	-16.957	53.9801	5	ERR599139	ERR599
TARA_B000000609	SAMEA2620651	ERS489712	MetaG	56	SUR	0.22-3	Non polar	SAMEA2620651	ERS489712	2010-06-26T07:05:00Z	-15.3424	43.2965	5	ERR599057	ERR599
TARA_B000000565	SAMEA2620672	ERS489733	MetaG	57	SUR	0.22-3	Non polar	SAMEA2620672	ERS489733	2010-06-27T12:05:00Z	-17.0248	42.7401	5	ERR599058	ERR599
TARA_B000000532	SAMEA2620756	ERS489877	MetaG	62	SUR	0.22-3	Non polar	SAMEA2620756	ERS489877	2010-07-03T08:09:00Z	-22.3368	40.3412	5	ERR599012	ERR599
TARA_B100000401	SAMEA2620786	ERS489917	MetaG	64	SUR	0.22-3	Non polar	SAMEA2620786	ERS489917	2010-07-07T04:48:00Z	-29.5019	37.9889	5	ERR598970	ERR598
TARA_B100000401	SAMEA2620786	ERS489917	MetaG	64	SUR	0.22-3	Non polar	SAMEA2620786	ERS489917	2010-07-07T04:48:00Z	-29.5019	37.9889	5	ERR599088	ERR599
TARA_B100000401	SAMEA2620786	ERS489917	MetaG	64	SUR	0.22-3	Non polar	SAMEA2620786	ERS489917	2010-07-07T04:48:00Z	-29.5019	37.9889	5	ERR599150	ERR599
TARA_B000000437	SAMEA2620855	ERS490029	MetaG	65	SUR	0.22-3	Non polar	SAMEA2620855	ERS490029	2010-07-12T05:59:00Z	-35.1728	26.2868	5	ERR598979	ERR598
TARA_B000000437	SAMEA2620855	ERS490029	MetaG	65	SUR	0.22-3	Non polar	SAMEA2620855	ERS490029	2010-07-12T05:59:00Z	-35.1728	26.2868	5	ERR599146	ERR599
TARA_B000000475	SAMEA2620929	ERS490124	MetaG	66	SUR	0.22-3	Non polar	SAMEA2620929	ERS490124	2010-07-15T12:22:00Z	-34.9449	17.9189	5	ERR598973	ERR598
TARA_B000000475	SAMEA2620929	ERS490124	MetaG	66	SUR	0.22-3	Non polar	SAMEA2620929	ERS490124	2010-07-15T12:22:00Z	-34.9449	17.9189	5	ERR599068	ERR599
TARA_B000000475	SAMEA2620929	ERS490124	MetaG	66	SUR	0.22-3	Non polar	SAMEA2620929	ERS490124	2010-07-15T12:22:00Z	-34.9449	17.9189	5	ERR599173	ERR599
TARA_B100000497	SAMEA2620970	ERS490183	MetaG	67	SUR	0.22-3	Non polar	SAMEA2620970	ERS490183	2010-09-07T06:19:00Z	-32.2401	17.7103	5	ERR598994	ERR598
TARA_B100000497	SAMEA2620970	ERS490183	MetaG	67	SUR	0.22-3	Non polar	SAMEA2620970	ERS490183	2010-09-07T06:19:00Z	-32.2401	17.7103	5	ERR599144	ERR599
TARA_B100000475	SAMEA2621013	ERS490265	MetaG	68	SUR	0.22-3	Non polar	SAMEA2621013	ERS490265	2010-09-14T06:55:00Z	-31.0266	4.665	5	ERR599129	ERR599
TARA_B100000475	SAMEA2621013	ERS490265	MetaG	68	SUR	0.22-3	Non polar	SAMEA2621013	ERS490265	2010-09-14T06:55:00Z	-31.0266	4.665	5	ERR599171	ERR599
TARA_B100000475	SAMEA2621013	ERS490265	MetaG	68	SUR	0.22-3	Non polar	SAMEA2621013	ERS490265	2010-09-14T06:55:00Z	-31.0266	4.665	5	ERR599174	ERR599
TARA_B100000459	SAMEA2621066	ERS490327	MetaG	70	SUR	0.22-3	Non polar	SAMEA2621066	ERS490327	2010-09-21T06:55:00Z	-20.4091	-3.1759	5	ERR599135	ERR599
TARA_B100000459	SAMEA2621066	ERS490327	MetaG	70	SUR	0.22-3	Non polar	SAMEA2621066	ERS490327	2010-09-21T06:55:00Z	-20.4091	-3.1759	5	ERR599165	ERR599
TARA_B100000424	SAMEA2621132	ERS490433	MetaG	72	SUR	0.22-3	Non polar	SAMEA2621132	ERS490433	2010-10-05T08:00:00Z	-8.7789	-17.9099	5	ERR598984	ERR598
TARA_B100000424	SAMEA2621132	ERS490433	MetaG	72	SUR	0.22-3	Non polar	SAMEA2621132	ERS490433	2010-10-05T08:00:00Z	-8.7789	-17.9099	5	ERR599105	ERR599

TARA_B100000513	SAMEA2621198	ERS490542	MetaG	76	SUR	0.22-3	Non polar	SAMEA2621198	ERS490542	2010-10-16T09:55:00Z	-20.9354	-35.1803	5	ERR599010	ERR599
TARA_B100000513	SAMEA2621198	ERS490542	MetaG	76	SUR	0.22-3	Non polar	SAMEA2621198	ERS490542	2010-10-16T09:55:00Z	-20.9354	-35.1803	5	ERR599126	ERR599
TARA_B100000524	SAMEA2621254	ERS490659	MetaG	78	SUR	0.22-3	Non polar	SAMEA2621254	ERS490659	2010-11-04T10:24:00Z	-30.1367	-43.2899	5	ERR599006	ERR599
TARA_B100000524	SAMEA2621254	ERS490659	MetaG	78	SUR	0.22-3	Non polar	SAMEA2621254	ERS490659	2010-11-04T10:24:00Z	-30.1367	-43.2899	5	ERR599022	ERR599
TARA_B100000768	SAMEA2621401	ERS490885	MetaG	82	SUR	0.22-3	Non polar	SAMEA2621401	ERS490885	2010-12-06T10:33:00Z	-47.1863	-58.2902	5	ERR599009	ERR599
TARA_B100000768	SAMEA2621401	ERS490885	MetaG	82	SUR	0.22-3	Non polar	SAMEA2621401	ERS490885	2010-12-06T10:33:00Z	-47.1863	-58.2902	5	ERR599035	ERR599
TARA_B100000780	SAMEA2621487	ERS491001	MetaG	84	SUR	0.22-3	Polar	SAMEA2621487	ERS491001	2011-01-03T11:05:00Z	-60.2287	-60.6476	5	ERR598945	ERR598
TARA_B100000780	SAMEA2621487	ERS491001	MetaG	84	SUR	0.22-3	Polar	SAMEA2621487	ERS491001	2011-01-03T11:05:00Z	-60.2287	-60.6476	5	ERR599059	ERR599
TARA_B100000787	SAMEA2621509	ERS491044	MetaG	85	SUR	0.22-3	Polar	SAMEA2621509	ERS491044	2011-01-06T10:38:00Z	-62.0385	-49.529	5	ERR599090	ERR599
TARA_B100000787	SAMEA2621509	ERS491044	MetaG	85	SUR	0.22-3	Polar	SAMEA2621509	ERS491044	2011-01-06T10:38:00Z	-62.0385	-49.529	5	ERR599176	ERR599
TARA_B100001063	SAMEA2621779	ERS491421	MetaG	93	SUR	0.22-3	Non polar	SAMEA2621779	ERS491421	2011-03-12T11:34:00Z	-34.0614	-73.1066	5	ERR599064	ERR599
TARA_B100001057	SAMEA2621839	ERS491492	MetaG	94	SUR	0.22-3	Non polar	SAMEA2621839	ERS491492	2011-03-18T11:57:00Z	-32.7971	-87.0693	5	ERR599050	ERR599
TARA_B100000989	SAMEA2621859	ERS491525	MetaG	96	SUR	0.22-3	Non polar	SAMEA2621859	ERS491525	2011-03-24T13:00:00Z	-29.7238	-101.1604	5	ERR598967	ERR598
TARA_B100001027	SAMEA2621990	ERS491699	MetaG	98	SUR	0.22-3	Non polar	SAMEA2621990	ERS491699	2011-04-03T13:44:00Z	-25.8051	-111.7202	5	ERR599093	ERR599
TARA_B100001027	SAMEA2621990	ERS491699	MetaG	98	SUR	0.22-3	Non polar	SAMEA2621990	ERS491699	2011-04-03T13:44:00Z	-25.8051	-111.7202	5	ERR599120	ERR599
TARA_B100000886	SAMEA2622074	ERS491804	MetaG	99	SUR	0.22-3	Non polar	SAMEA2622074	ERS491804	2011-04-09T13:56:00Z	-21.146	-104.787	5	ERR599024	ERR599
TARA_B100000963	SAMEA2622097	ERS491836	MetaG	100	SUR	0.22-3	Non polar	SAMEA2622097	ERS491836	2011-04-15T12:45:00Z	-13.023	-95.9759	5	ERR599063	ERR599
TARA_B100000963	SAMEA2622097	ERS491836	MetaG	100	SUR	0.22-3	Non polar	SAMEA2622097	ERS491836	2011-04-15T12:45:00Z	-13.023	-95.9759	5	ERR599163	ERR599
TARA_B100000963	SAMEA2622097	ERS491836	MetaG	100	SUR	0.22-3	Non polar	SAMEA2622097	ERS491836	2011-04-15T12:45:00Z	-13.023	-95.9759	5	ERR599169	ERR599
TARA_B100000900	SAMEA2622173	ERS491938	MetaG	102	SUR	0.22-3	Non polar	SAMEA2622173	ERS491938	2011-04-21T20:07:00Z	-5.2529	-85.1545	5	ERR598943	ERR598
TARA_B100000900	SAMEA2622173	ERS491938	MetaG	102	SUR	0.22-3	Non polar	SAMEA2622173	ERS491938	2011-04-21T20:07:00Z	-5.2529	-85.1545	5	ERR598978	ERR598
TARA_B100000925	SAMEA2622316	ERS492145	MetaG	109	SUR	0.22-3	Non polar	SAMEA2622316	ERS492145	2011-05-12T14:00:00Z	1.9928	-84.5766	5	ERR598997	ERR598
TARA_B100000925	SAMEA2622316	ERS492145	MetaG	109	SUR	0.22-3	Non polar	SAMEA2622316	ERS492145	2011-05-12T14:00:00Z	1.9928	-84.5766	5	ERR599118	ERR599
TARA_B100001109	SAMEA2622376	ERS492228	MetaG	110	SUR	0.22-3	Non polar	SAMEA2622376	ERS492228	2011-05-21T12:27:00Z	-2.0133	-84.589	5	ERR599039	ERR599
TARA_B100000575	SAMEA2622452	ERS492321	MetaG	111	SUR	0.22-3	Non polar	SAMEA2622452	ERS492321	2011-05-31T14:25:00Z	-16.9601	-100.6335	5	ERR599077	ERR599
TARA_B100000941	SAMEA2622518	ERS492408	MetaG	112	SUR	0.22-3	Non polar	SAMEA2622518	ERS492408	2011-06-14T16:45:00Z	-23.2811	-129.3947	5	ERR599054	ERR598
TARA_B100001115	SAMEA2622652	ERS492642	MetaG	122	SUR	0.22-3	Non polar	SAMEA2622652	ERS492642	2011-07-26T17:10:00Z	-8.9971	-139.1963	5	ERR599892	ERR598
TARA_B100000683	SAMEA2622710	ERS492733	MetaG	123	SUR	0.22-3	Non polar	SAMEA2622710	ERS492733	2011-07-31T17:20:00Z	-8.9068	-140.283	5	ERR599160	ERR599
TARA_B100000674	SAMEA2622759	ERS492814	MetaG	124	SUR	0.22-3	Non polar	SAMEA2622759	ERS492814	2011-08-04T18:33:00Z	-9.1504	-140.5216	5	ERR588857	ERR588
TARA_B100000674	SAMEA2622759	ERS492814	MetaG	124	SUR	0.22-3	Non polar	SAMEA2622759	ERS492814	2011-08-04T18:33:00Z	-9.1504	-140.5216	5	ERR599036	ERR599
TARA_B100000674	SAMEA2622759	ERS492814	MetaG	124	SUR	0.22-3	Non polar	SAMEA2622759	ERS492814	2011-08-04T18:33:00Z	-9.1504	-140.5216	5	ERR599069	ERR599
TARA_B100000674	SAMEA2622759	ERS492814	MetaG	124	SUR	0.22-3	Non polar	SAMEA2622759	ERS492814	2011-08-04T18:33:00Z	-9.1504	-140.5216	5	ERR599080	ERR599
TARA_B100000674	SAMEA2622759	ERS492814	MetaG	124	SUR	0.22-3	Non polar	SAMEA2622759	ERS492814	2011-08-04T18:33:00Z	-9.1504	-140.5216	5	ERR599151	ERR599
TARA_B100001121	SAMEA2622817	ERS492888	MetaG	125	SUR	0.22-3	Non polar	SAMEA2622817	ERS492888	2011-08-08T17:33:00Z	-8.9111	-142.5571	5	ERR599066	ERR599
TARA_B100001121	SAMEA2622817	ERS492888	MetaG	125	SUR	0.22-3	Non polar	SAMEA2622817	ERS492888	2011-08-08T17:33:00Z	-8.9111	-142.5571	5	ERR599091	ERR599
TARA_B100001121	SAMEA2622817	ERS492888	MetaG	125	SUR	0.22-3	Non polar	SAMEA2622817	ERS492888	2011-08-08T17:33:00Z	-8.9111	-142.5571	5	ERR599114	ERR599
TARA_B100001121	SAMEA2622817	ERS492888	MetaG	125	SUR	0.22-3	Non polar	SAMEA2622817	ERS492888	2011-08-08T17:33:00Z	-8.9111	-142.5571	5	ERR599119	ERR599
TARA_B100000609	SAMEA2622901	ERS493044	MetaG	128	SUR	0.22-3	Non polar	SAMEA2622901	ERS493044	2011-09-04T18:00:00Z	3.00E-04	-153.6759	5	ERR599038	ERR599
TARA_B100001248	SAMEA2623059	ERS493300	MetaG	132	SUR	0.22-3	Non polar	SAMEA2623059	ERS493300	2011-10-04T17:46:00Z	31.5213	-158.9958	5	ERR599142	ERR599
TARA_B100001093	SAMEA2623116	ERS493390	MetaG	133	SUR	0.22-3	Non polar	SAMEA2623116	ERS493390	2011-10-18T15:35:00Z	35.3671	-127.7422	5	ERR599052	ERR599

TARA_B10 0001287	SAMEA 2623275	ERS49 3636	MetaG	137	SU R	0.22-3	Non polar	SAMEA2623275	ERS493636	2011-12-02T14:12:00Z	14.2035	-116.6261	5	ERR598989	ERR598
TARA_B10 0001989	SAMEA 2623350	ERS49 3752	MetaG	138	SU R	0.22-3	Non polar	SAMEA2623350	ERS493752	2011-12-10T14:08:00Z	6.3332	-102.9432	5	ERR599030	ERR599
TARA_B10 0002019	SAMEA 2623426	ERS49 3877	MetaG	140	SU R	0.22-3	Non polar	SAMEA2623426	ERS493877	2011-12-21T16:20:00Z	7.4122	-79.3017	5	ERR599162	ERR599
TARA_B10 0001939	SAMEA 2623446	ERS49 3914	MetaG	141	SU R	0.22-3	Non polar	SAMEA2623446	ERS493914	2011-12-30T13:35:00Z	9.8481	-80.0454	5	ERR599029	ERR599
TARA_B10 0002051	SAMEA 2623463	ERS49 3938	MetaG	142	SU R	0.22-3	Non polar	SAMEA2623463	ERS493938	2012-01-09T13:41:00Z	25.5264	-88.394	5	ERR599136	ERR599
TARA_B10 0001142	SAMEA 2623627	ERS49 4170	MetaG	145	SU R	0.22-3	Non polar	SAMEA2623627	ERS494170	2012-02-02T11:52:00Z	39.2305	-70.0377	5	ERR598983	ERR598
TARA_B10 0001540	SAMEA 2623673	ERS49 4236	MetaG	146	SU R	0.22-3	Non polar	SAMEA2623673	ERS494236	2012-02-15T12:54:00Z	34.6712	-71.3093	5	ERR598968	ERR598
TARA_B10 0001741	SAMEA 2623734	ERS49 4332	MetaG	148	SU R	0.22-3	Non polar	SAMEA2623734	ERS494332	2012-02-24T11:40:00Z	31.6948	-64.2489	5	ERR599123	ERR599
TARA_B10 0001758	SAMEA 2623774	ERS49 4394	MetaG	149	SU R	0.22-3	Non polar	SAMEA2623774	ERS494394	2012-03-01T10:48:00Z	34.1132	-49.9181	5	ERR598963	ERR598
TARA_B10 0001769	SAMEA 2623808	ERS49 4445	MetaG	150	SU R	0.22-3	Non polar	SAMEA2623808	ERS494445	2012-03-05T09:25:00Z	35.9346	-37.3032	5	ERR599170	ERR599
TARA_B10 0001564	SAMEA 2623850	ERS49 4518	MetaG	151	SU R	0.22-3	Non polar	SAMEA2623850	ERS494518	2012-03-09T08:59:00Z	36.1715	-29.023	5	ERR598976	ERR598
TARA_B10 0001173	SAMEA 2623886	ERS49 4579	MetaG	152	SU R	0.22-3	Non polar	SAMEA2623886	ERS494579	2012-03-19T08:18:00Z	43.6792	-16.8344	5	ERR599078	ERR599
TARA_B11 0000003	SAMEA 4396424	ERS13 07873	MetaG	155	SU R	0.22-3	Polar	SAMEA4396424	ERS1307873	2013-05-24T05:36:00Z	54.5305	-16.9377	5	ERR3589591	ERR358
TARA_B11 0001450	SAMEA 4396538	ERS13 07987	MetaG	158	SU R	0.22-3	Polar	SAMEA4396538	ERS1307987	2013-06-03T07:31:00Z	67.141	0.2355	5	ERR3589592	ERR358
TARA_B11 0001469	SAMEA 4397034	ERS13 08483	MetaG	163	SU R	0.22-3	Polar	SAMEA4397034	ERS1308483	2013-06-09T07:39:00Z	76.1825	1.3918	5	ERR3589586	ERR358
TARA_B11 0000027	SAMEA 4397101	ERS13 08550	MetaG	168	SU R	0.22-3	Polar	SAMEA4397101	ERS1308550	2013-07-01T03:20:00Z	72.5128	44.0775	5	ERR3589588	ERR358
TARA_B11 0000090	SAMEA 4397239	ERS13 08688	MetaG	173	SU R	0.22-3	Polar	SAMEA4397239	ERS1308688	2013-07-08T04:12:00Z	78.9564	79.4201	5	ERR3589566	ERR358
TARA_B11 0000114	SAMEA 4397311	ERS13 08760	MetaG	175	SU R	0.22-3	Polar	SAMEA4397311	ERS1308760	2013-07-10T03:58:00Z	79.2233	66.3435	5	ERR3589577	ERR358
TARA_B11 0000208	SAMEA 4397426	ERS13 08875	MetaG	178	SU R	0.22-3	Polar	SAMEA4397426	ERS1308875	2013-07-15T02:11:00Z	77.1604	73.2057	5	ERR3589590	ERR358
TARA_B11 0000503	SAMEA 4397472	ERS13 08921	MetaG	180	SU R	0.22-3	Polar	SAMEA4397472	ERS1308921	2013-07-18T02:04:00Z	74.8023	76.1478	5	ERR3589568	ERR358
TARA_B11 0000238	SAMEA 4397569	ERS13 09018	MetaG	188	SU R	0.22-3	Polar	SAMEA4397569	ERS1309018	2013-08-15T00:30:00Z	78.2518	91.8557	5	ERR3589554	ERR358
TARA_B11 0000259	SAMEA 4397649	ERS13 09098	MetaG	189	SU R	0.22-3	Polar	SAMEA4397649	ERS1309098	2013-08-27T01:53:00Z	77.9028	117.1545	5	ERR3589555	ERR358
TARA_B11 0000285	SAMEA 4397756	ERS13 09205	MetaG	191	SU R	0.22-3	Polar	SAMEA4397756	ERS1309205	2013-09-02T22:37:00Z	71.5955	160.9383	5	ERR3589580	ERR358
TARA_B11 0000977	SAMEA 4397798	ERS13 09247	MetaG	193	SU R	0.22-3	Polar	SAMEA4397798	ERS1309247	2013-09-08T02:45:00Z	71.0704	174.9916	5	ERR3589581	ERR358
TARA_B11 0000971	SAMEA 4397842	ERS13 09291	MetaG	194	SU R	0.22-3	Polar	SAMEA4397842	ERS1309291	2013-09-11T20:00:00Z	73.3833	-168.1333	5	ERR3589556	ERR358
TARA_B11 0000305	SAMEA 4397930	ERS13 09379	MetaG	196	SU R	0.22-3	Polar	SAMEA4397930	ERS1309379	2013-09-14T17:30:00Z	71.8895	-154.9101	5	ERR3589557	ERR358
TARA_B11 0000902	SAMEA 4398009	ERS13 09458	MetaG	201	SU R	0.22-3	Polar	SAMEA4398009	ERS1309458	2013-09-30T15:02:00Z	74.2987	-85.7806	5	ERR3589558	ERR358
TARA_B11 0000879	SAMEA 4398094	ERS13 09543	MetaG	205	SU R	0.22-3	Polar	SAMEA4398094	ERS1309543	2013-10-08T12:15:00Z	72.4693	-71.892	5	ERR3589559	ERR358
TARA_B11 0000483	SAMEA 4398176	ERS13 09625	MetaG	206	SU R	0.22-3	Polar	SAMEA4398176	ERS1309625	2013-10-12T11:09:00Z	70.9574	-53.5989	5	ERR3589561	ERR358
TARA_B11 0000858	SAMEA 4398247	ERS13 09696	MetaG	208	SU R	0.22-3	Polar	SAMEA4398247	ERS1309696	2013-10-20T10:34:00Z	69.1136	-51.5086	5	ERR3589575	ERR358
TARA_B11 0000459	SAMEA 4398296	ERS13 09745	MetaG	209	SU R	0.22-3	Polar	SAMEA4398296	ERS1309745	2013-10-23T10:25:00Z	64.7127	-53.0106	5	ERR3589576	ERR358
TARA_B11 0000444	SAMEA 4398370	ERS13 09819	MetaG	210	SU R	0.22-3	Polar	SAMEA4398370	ERS1309819	2013-10-27T10:16:00Z	61.5427	-55.9869	5	ERR3589582	ERR358

Table 4. Overall genomic statistic and taxonomy for all 495 prokaryotic MAGs.

BIN	completeness	contamination	strain heterogeneity	GC	lineage	N50	size	Domain	Phylum	Class	Order	Family	Genus	Species
bin.G1.103	96.69	0.717	20	0.336	algicola	24471	2639777	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Croceibacter	Croceibacter atlanticus
bin.G1.106	51.72	0	0	0.328	Bacteria	15905	658022	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SCGC-AAA076-P13	
bin.G1.114	89.27	1.801	0	0.358	Bacteria	16092	2339769	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952	
bin.G1.125	99.46	0.806	50	0.372	Bacteria	141733	2625145	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	GCA-002722245		
bin.G1.135	56.76	1.925	0	0.335	Gammaproteobacteria	35295	980712	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	MED-G78	MED-G78 sp902514105
bin.G1.136	78.26	1.333	100	0.382	Euryarchaeota	53460	1378590	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O2	MGIIb-O2 sp902593945

bin.G1.144	94.64	8.435	0	0.315	Bacteria	46186	1790856	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2711125	
bin.G1.148	88.44	0.358	0	0.308	Bacteria	117853	1380930	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7312		
bin.G1.15	80.75	4.8	85.71	0.359	Euryarchaeota	42432	1413622	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O3	
bin.G1.157	55.51	0.577	33.33	0.301	algicola	31207	915877	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	MED-G13 sp902519075
bin.G1.160	75.56	1.081	0	0.415	Bacteria	7519	2340787	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA2040	
bin.G1.161	80.57	2.688	0	0.286	Bacteria	87923	1350714	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	GCA-002707245	
bin.G1.180	79.61	3.499	57.89	0.427	algicola	13345	1752583	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	BACL21	
bin.G1.182	52.06	3.448	0	0.291	Bacteria	11507	1076879	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	SP256	
bin.G1.187	85.33	0.861	0	0.343	Euryarchaeota	86590	1806852	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-I	
bin.G1.188	90.21	2.15	25	0.296	Bacteria	50166	1455887	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2716345	GCA-2716345 sp002716345
bin.G1.197	71.73	0	0	0.457	Euryarchaeota	50278	1354925	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O1	MGIIb-O1 sp002457555
bin.G1.201	59.31	5.172	100	0.294	Bacteria	7307	1397803	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	MED-G20	
bin.G1.203	75.82	2.298	25	0.361	Gammaproteobacteria	47463	1323048	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	SAR86	GCA-2707915	
bin.G1.204	89.67	1.621	0	0.329	Bacteria	8993	3248917	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	Putridiphycobacter	
bin.G1.207	52.81	8.75	71.74	0.347	Gammaproteobacteria	12881	1043784	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	CACNYO01	
bin.G1.211	56.50	1.746	20	0.398	Alphaproteobacteria	5179	1400825	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA5951	
bin.G1.215	97.40	0.268	0	0.435	Bacteria	45256	2841208	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16		
bin.G1.219	82.49	1.612	100	0.58	Bacteria	12445	2081693	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA8752	
bin.G1.22	79.20	0	0	0.347	Euryarchaeota	49494	1711685	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-I	MGIIa-I sp002699515
bin.G1.221	64.63	1.454	16.67	0.358	Flavobacteriaceae	4482	1328972	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	PolaribacterA	
bin.G1.222	84.76	2.956	66.67	0.561	Bacteria	27675	1890419	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663	
bin.G1.232	83.33	0	0	0.321	Bacteria	54685	1417116	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11		
bin.G1.234	74.49	5.105	11.11	0.299	Bacteria	7838	1345660	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	SP256	
bin.G1.236	51.97	2.465	10	0.341	algicola	4062	1532722	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-2733415	
bin.G1.24	51.88	1.724	0	0.488	Bacteria	4541	1334643	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	Puniceispirillum	
bin.G1.244	95.27	0.954	50	0.366	Flavobacteriaceae	29672	2978286	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	
bin.G1.25	92.16	0	0	0.527	Bacteria	21608	1922456	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	RFXV01	
bin.G1.251	68.24	1.921	87.5	0.604	Alphaproteobacteria	6143	1634178	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	
bin.G1.271	88.92	4.301	55.56	0.414	Bacteria	41646	1801768	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA974	
bin.G1.277	84.43	1.361	12.5	0.427	algicola	40206	1742303	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	BACL21	
bin.G1.282	80.13	4.8	66.67	0.431	Euryarchaeota	22019	1891471	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1	
bin.G1.291	68.31	2.419	20	0.28	Bacteria	13287	989504	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	GCA-2718035	
bin.G1.295	82.88	3.366	73.68	0.365	algicola	11852	1628202	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	UBA3537 sp001735715
bin.G1.296	92.12	3.347	54.55	0.474	Alphaproteobacteria	26555	1639116	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G1.297	70.78	0.574	0	0.364	Gammaproteobacteria	65237	909753	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112	TMED112 sp003331605
bin.G1.298	74.03	0.353	0	0.284	Flavobacteriaceae	110955	1409023	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8	MAG-121220-bin8 sp902511755
bin.G1.304	85.97	2.4	75	0.459	Euryarchaeota	118520	2066193	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp002687075
bin.G1.306	90.39	9.826	2.78	0.297	Alphaproteobacteria	57419	2294487	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	AAA536-G10		
bin.G1.308	56.55	6.077	59.09	0.493	Alphaproteobacteria	7564	950348	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G1.311	77.77	2.795	23.08	0.321	Bacteria	38057	1459076	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11		
bin.G1.314	60.87	4.086	52.94	0.333	Gammaproteobacteria	23231	898442	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	AG-339-G14	AG-339-G14 sp004213955
bin.G1.317	75.32	0.8	0	0.358	Euryarchaeota	44651	1441996	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O3	
bin.G1.319	79.82	0.537	100	0.291	Bacteria	50347	1178088	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	GCA-2696965	GCA-2696965 sp002171975

bin.G1.320	82.80	1.6	75	0.429	Euryarchaeota	17982	1981385	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1	MGIIa-K1 sp003602415
bin.G1.321	94.08	6.989	0	0.302	Bacteria	30193	1648628	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-002697625	
bin.G1.322	92.74	0.358	50	0.333	Bacteria	41429	1521709	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	TMED123	
bin.G1.332	81.46	7.594	82.35	0.372	algicola	9121	1583954	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	UBA3537 sp002684355
bin.G1.334	50.90	7.787	51.85	0.497	Alphaproteobacteria	4465	948170	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G1.341	54.64	2.777	33.33	0.364	Gammaproteobacteria	8976	924044	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	D2472	
bin.G1.348	55.02	0	0	0.417	Euryarchaeota	5452	708393	Archaea	Halobacteriota	Halobacteria	Halobacteriales	UBA12382	UBA12382	
bin.G1.35	76.13	0	0	0.516	Euryarchaeota	41698	1277228	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O5	MGIIb-O5 sp002498925
bin.G1.351	66.09	9.587	75	0.305	Bacteria	6934	1451067	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-002697625	
bin.G1.36	70.19	6.362	5	0.31	Bacteria	26781	1412792	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11		
bin.G1.364	51.72	4.31	80	0.438	Bacteria	5915	1222454	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	
bin.G1.368	76.07	0	0	0.485	Euryarchaeota	37391	1988283	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	SP79	SP79	
bin.G1.373	77.79	4.139	77.78	0.442	Bacteria	6410	1559180	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	UBA10364 sp002387615
bin.G1.378	91.31	1.26	0	0.34	Bacteria	106583	2992080	Bacteria	SAR324	SAR324	SAR324	NAC60-12	Arctic96AD-7	
bin.G1.380	72.49	1.036	0	0.381	Flavobacteriaceae	19279	1538684	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	MS024-2A sp902546725
bin.G1.389	77.31	0	0	0.31	Bacteria	39207	1303089	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11		
bin.G1.392	61.76	4.779	77.78	0.293	Bacteria	11154	1048094	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-002723295	GCA-002723295 sp002711185
bin.G1.396	76.91	0.582	0	0.47	Alphaproteobacteria	7770	1334080	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispiralaceae	UBA3439	
bin.G1.398	60.39	8.726	91.49	0.315	Flavobacteriaceae	2913	1368927	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02	
bin.G1.4	83.73	0	0	0.422	Euryarchaeota	200587	1917671	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K2	MGIIa-K2 sp002699425
bin.G1.409	58.90	9.61	23.91	0.314	Gammaproteobacteria	22646	813061	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SAR86A	SAR86A sp002169625
bin.G1.41	85.94	1.344	100	0.315	Bacteria	16071	1480510	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7312	UBA8444	
bin.G1.415	80.81	7.036	75	0.339	Flavobacteriaceae	9712	1710880	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	
bin.G1.429	76.88	1.612	0	0.323	Bacteria	71403	1292763	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2683775	
bin.G1.436	85.75	4.301	87.5	0.322	Bacteria	9143	2343416	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	GCA-2716065	
bin.G1.44	74.85	1.075	50	0.318	Bacteria	54047	1342826	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	BACL11	
bin.G1.440	91.29	4.733	50	0.326	algicola	29213	1883172	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-002733185	GCA-002733185 sp004213605
bin.G1.444	63.79	0	0	0.342	Bacteria	76620	889860	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	CACEJU01	
bin.G1.451	87.50	3.676	70	0.295	Bacteria	32866	1370432	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	MED-G14 sp902574335
bin.G1.453	80.40	0	0	0.517	Euryarchaeota	23493	1342139	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	Thalassarchaeum	Thalassarchaeum sp002698225
bin.G1.457	92.43	1.075	50	0.395	Bacteria	41844	1864397	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA974	
bin.G1.46	72.68	2.609	86.67	0.331	algicola	4072	1835445	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Mesonia	Mesonia algae
bin.G1.462	52.83	5.66	0	0.282	Bacteria	69535	676764	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	TMED165	
bin.G1.466	80.51	0.16	0	0.511	Euryarchaeota	49931	1779741	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	Poseidonia	Poseidonia sp002494645
bin.G1.468	62.33	2.676	0	0.347	Gammaproteobacteria	46316	1190629	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	MEDG-81	MEDG-81 sp002689405
bin.G1.470	97.58	0.985	25	0.407	Bacteria	36227	2036507	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	UBA10364 sp003023665
bin.G1.473	58.26	3.308	40	0.297	Bacteria	4832	810992	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G11	MED-G11 sp004214015
bin.G1.477	86.45	4.301	46.15	0.278	Bacteria	85659	1298457	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	GCA-2718035	
bin.G1.480	84.13	3.607	80.95	0.369	Euryarchaeota	92572	1590658	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O3	MGIIb-O3 sp002731195
bin.G1.481	99.01	1.075	0	0.322	Bacteria	45112	3182517	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	GCA-002722245	GCA-002722245	
bin.G1.485	53.46	2.72	100	0.415	Euryarchaeota	10114	1107641	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1	
bin.G1.491	89.53	1.504	75	0.367	Bacteria	33184	2306971	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3478	UBA3478 sp002691685
bin.G1.493	91.17	1.47	100	0.292	Bacteria	107920	1741584	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-002723295	

bin.G1.495	51.45	6.518	0	0.289	Bacteria	72534	633560	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	TMED170	
bin.G1.503	60.10	0	0	0.35	Gammaproteobacteria	77883	737989	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112	TMED112 sp902529055
bin.G1.509	69.82	3.448	66.67	0.338	Bacteria	63427	752166	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112	TMED112 sp002170245
bin.G1.526	93.22	6.379	56.52	0.415	Bacteria	31352	188476 0	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA974	
bin.G1.528	71.03	2.4	0	0.464	Euryarchaeota	30346	171418	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp002170315
bin.G1.53	86.26	0.8	0	0.48	Euryarchaeota	76188	190289 8	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	Poseidonia	Poseidonia sp002726495
bin.G1.545	96.95	3.371	37.5	0.358	algicola	93997	193028 2	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Hel1-33-131	
bin.G1.546	83.97	0.672	100	0.564	Bacteria	15610	255876 3	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16		
bin.G1.551	60.86	1.724	100	0.371	Bacteria	7178	121988 7	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UlvibacterB	
bin.G1.561	55.24	5.308	21.21	0.29	algicola	25659	120722 2	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	
bin.G1.564	76.34	2.419	0	0.293	Bacteria	76849	143779 9	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2715565	
bin.G1.566	84.03	2.688	100	0.328	Bacteria	7561	210351 2	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	SP287	
bin.G1.570	88.15	0	0	0.29	Bacteria	63595	150762 7	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2697505	
bin.G1.576	85.21	1.075	100	0.332	Bacteria	23575	230132 0	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	GCA-002722245	GCA-002722245	
bin.G1.577	81.72	2.15	0	0.318	Bacteria	38915	150800 8	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	BACL11	
bin.G1.582	70.39	3.846	80	0.411	Bacteria	8290	190782 5	Bacteria	SAR324	SAR324	SAR324	NAC60-12	Arctic96AD-7	
bin.G1.583	85.48	3.046	25	0.288	Bacteria	64618	152059 4	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2711125	
bin.G1.586	78.13	1.6	50	0.434	Euryarchaeota	19076	187010 9	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-Q1	
bin.G1.590	77.49	4.119	58.33	0.444	Bacteria	13414	366735 4	Bacteria	SAR324	SAR324	SAR324	NAC60-12	JCVI-SCAAA005	JCVI-SCAAA005 sp000224765
bin.G1.592	76.82	8.064	89.47	0.626	Bacteria	6114	224829 5	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA8752	
bin.G1.596	92.91	7.432	0	0.316	Bacteria	18134	230570 7	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952	UBA952 sp002696025
bin.G1.607	90.46	8.287	4.76	0.406	Bacteria	11471	304090 3	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Luteibaculaceae		
bin.G1.614	52.63	5.263	28.57	0.284	Bacteria	18782	142723 9	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	AAA536-G10	TMED54	TMED54 sp002691795
bin.G1.615	79.12	2.688	16.67	0.297	Bacteria	45234	144652 7	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11		
bin.G1.620	58.10	0.8	0	0.376	Euryarchaeota	11789	114469 1	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O2	MGIIb-O2 sp002495525
bin.G1.622	98.85	0.982	0	0.355	algicola	87174	345373 8	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Leeuwenhoekiella	Leeuwenhoekiella aequorea
bin.G1.623	52.47	0.376	100	0.309	Bacteria	4381	891840	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	GCA-002701365	
bin.G1.625	83.73	2.4	75	0.365	Euryarchaeota	22313	122484 5	Archaea	Thermoplasmatota	PoseidoniiaB	MGIII	CG-Epi1	CG-Epi1	
bin.G1.627	62.29	3.76	22.22	0.288	Bacteria	46166	606388	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	MED-G40	MED-G40 sp902567685
bin.G1.630	91.24	3.883	87.88	0.374	algicola	23576	215742 9	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3478	UBA3478 sp002691645
bin.G1.632	54.54	9.191	5.26	0.271	Bacteria	9456	117524 2	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-2719315	
bin.G1.633	91.82	1.118	66.67	0.365	Flavobacteriaceae	15158	213700 1	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02	
bin.G1.634	60.98	9.312	25	0.333	algicola	6077	144185 3	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	
bin.G1.635	77.36	5.28	22.22	0.421	Euryarchaeota	16404	178816 3	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1	
bin.G1.639	86.66	0	0	0.48	Euryarchaeota	60462	186299 7	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L2	
bin.G1.643	78.85	5.105	25	0.331	Bacteria	33398	139732 6	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	BACL11	BACL11 sp002730985
bin.G1.647	70.06	6.71	45.16	0.302	algicola	18782	131873 1	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	MED-G13 sp008081325
bin.G1.651	78.97	6.397	53.33	0.297	Bacteria	10804	191271 9	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	MED-G20	MED-G20 sp002457645
bin.G1.652	52.63	3.508	100	0.309	Bacteria	5580	135366 2	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02	
bin.G1.656	74.33	0.8	100	0.49	Euryarchaeota	39324	135847 7	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-N1	MGIIb-N1 sp002170775
bin.G1.66	64.74	6.559	6.25	0.289	Bacteria	7811	141453 9	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2715565	
bin.G1.67	87.52	0.537	100	0.301	Bacteria	45276	149275 9	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	GCA-002701365	GCA-002701365 sp002701365
bin.G1.671	100	0	0	0.41	Bacteria	97061	310981 1	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	UBA10066	UBA10066 sp003448535

bin.G1.673	99.37	1.075	50	0.374	Bacteria	85096	3248101	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066			
bin.G1.675	50.78	1.851	14.29	0.35	Gammaproteobacteria	28908	982010	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	SAR86	GCA-2730855		
bin.G1.678	93.24	0.54	0	0.376	Bacteria	39922	2525914	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952		
bin.G1.679	89.85	3.899	25	0.385	algicola	21450	2149915	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3478		
bin.G1.684	51.36	7.98	43.9	0.36	Gammaproteobacteria	9618	1141202	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	D2472		
bin.G1.685	82.33	2.111	11.11	0.353	algicola	6403	2417701	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Nonlabens	Nonlabens dokdonensisB	
bin.G1.686	56.00	8	42.86	0.475	Euryarchaeota	5683	1746189	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-P		
bin.G1.687	95.25	1.612	0	0.409	Bacteria	52146	1897189	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA974	UBA974 sp002292405	
bin.G1.689	85.46	0.8	0	0.427	Euryarchaeota	34396	1744490	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1	MGIIa-K1 sp002689345	
bin.G1.698	66.53	0.533	0	0.46	Euryarchaeota	21288	1199414	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O3	MGIIb-O3 sp002172185	
bin.G1.7	99.46	0.537	0	0.422	Bacteria	278187	2550753	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Salibacteraceae	SHAN690		
bin.G1.700	86.06	2.06	100	0.321	Bacteria	11865	2111364	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	GCA-2723085		
bin.G1.715	93.10	2.347	66.67	0.318	Bacteria	24780	2447006	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	GCA-2723085		
bin.G1.719	81.73	3.2	100	0.49	Euryarchaeota	34678	1784640	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-N2	MGIIb-N2 sp002713585	
bin.G1.73	76.00	0	0	0.516	Euryarchaeota	21489	1450105	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O1	MGIIb-O1 sp002497895	
bin.G1.733	87.36	0.537	0	0.317	Bacteria	54076	1549190	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2712715		
bin.G1.737	72.88	0.179	0	0.419	Bacteria	6445	2382819	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	Owenweeksia		
bin.G1.746	80.93	1.6	0	0.397	Euryarchaeota	65012	1862210	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-N2	MGIIb-N2 sp002502625	
bin.G1.748	72.56	5.125	90.91	0.563	Bacteria	17720	1570526	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16	UA16 sp002448555	
bin.G1.755	95.21	0.825	66.67	0.329	algicola	49094	2105771	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Winogradskyella		
bin.G1.758	60.34	0	0	0.337	Bacteria	15425	713163	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SCGC-AAA076-P13		
bin.G1.759	92.12	0.938	60	0.382	algicola	71630	2033811	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7428		
bin.G1.760	71.60	3.243	85.71	0.319	Bacteria	5470	1969031	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952		
bin.G1.764	73.60	1.075	100	0.415	Bacteria	7163	2465705	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	UBA10066	UBA10066 sp014239715	
bin.G1.767	72.22	8.888	90	0.4	Bacteria	5514	1408074	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA974		
bin.G1.769	78.13	7.2	100	0.41	Euryarchaeota	27218	1821344	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1		
bin.G1.77	87.45	0.537	0	0.309	Bacteria	57681	1476012	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2711125		
bin.G1.774	83.81	2.756	27.27	0.313	algicola	35178	1648652	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-002733185	GCA-002733185 sp004214175	
bin.G1.775	96.05	1.612	75	0.319	Bacteria	55566	1816531	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	GCA-2707145		
bin.G1.778	98.74	0.806	50	0.302	Bacteria	133522	2716623	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	GCA-002722245			
bin.G1.780	97.32	1.521	20	0.571	Alphaproteobacteria	55535	1987223	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	UBA1172	UBA1172	UBA1172 sp002457135	
bin.G1.783	94.08	0.537	100	0.317	Bacteria	85267	1652532	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2705205		
bin.G1.786	86.13	0.8	0	0.435	Euryarchaeota	45967	1942891	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K2		
bin.G1.787	83.62	4.019	80.77	0.397	algicola	23079	1518977	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	UBA7446 sp002470745	
bin.G1.795	86.61	1.612	25	0.399	Bacteria	12180	1618692	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA974		
bin.G1.797	84.46	0.537	0	0.31	Bacteria	57278	1247754	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	TMED113	SP256		
bin.G1.799	67.66	5.645	88.57	0.334	algicola	7211	1388127	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	UBA8316 sp902584445	
bin.G1.80	89.89	2.659	37.5	0.611	Alphaproteobacteria	18998	2039986	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	UBA8309 sp002457745	
bin.G1.802	68.84	1.075	100	0.554	Bacteria	9931	1528196	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16		
bin.G1.808	90.86	0	0	0.356	Bacteria	129961	1516736	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7312	UBA7312		
bin.G1.812	83.82	0.663	33.33	0.317	Bacteria	7897	1269836	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2705205	GCA-2705205 sp002705205	
bin.G1.815	78.76	5.021	6.67	0.299	Bacteria	10749	1556518	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11			
bin.G1.817	50.17	1.724	100	0.355	Bacteria	3990	1139195	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Hel1-33-131		

bin.G1.818	60.19	5.792	70	0.293	algicola	10368	1105139	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-2700405	GCA-2700405 sp002700405	
bin.G1.821	52.64	0.74	50	0.331	Gammaproteobacteria	16701	914045	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SCGC-AAA076-P13		
bin.G1.825	81.81	3.2	80	0.446	Euryarchaeota	23299	1825599	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L2	MGIIa-L2 sp002171315	
bin.G1.826	70.00	1.724	100	0.552	Bacteria	20818	1691898	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA8752	UBA8752 sp002336475	
bin.G1.830	59.76	0	0	0.503	Euryarchaeota	11347	1393212	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp8160u	
bin.G1.832	64.64	4.366	90	0.612	Alphaproteobacteria	5231	1640423	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309		
bin.G1.834	89.18	0.672	50	0.361	Bacteria	30493	1648950	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	MED-G21		
bin.G1.843	92.25	1.075	0	0.312	Bacteria	37255	2470979	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	GCA-2716065		
bin.G1.844	51.39	6.054	77.14	0.402	algicola	9367	8267 25	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446		
bin.G1.847	54.80	2.688	40	0.324	Bacteria	5842	1837057	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	GCA-2716065		
bin.G1.849	60.44	9.195	38.89	0.356	Gammaproteobacteria	8959	1354382	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	SAR86	GCA-2707915		
bin.G1.850	60.20	8.77	15	0.361	Archaea	6403	7877 29	Archaea	Thermoplasmatota	PoseidoniiaB	MGIII	CG-Epi1	CG-Epi1		
bin.G1.851	80.85	5.676	63.16	0.469	Alphaproteobacteria	11490	1344834	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439		
bin.G1.852	64.77	4.4	42.86	0.363	Euryarchaeota	8439	940549	Archaea	Thermoplasmatota	PoseidoniiaB	MGIII	CG-Epi1	CG-Epi1		
bin.G1.865	60.52	7.009	90	0.371	Archaea	5278	582107	Archaea	Nanoarchaeota	Nanoarchaeia	Woesearchaeales	GW2011-AR9	CABZYC01	CABZYC01 sp902529885	
bin.G1.873	73.44	1.892	33.33	0.476	Alphaproteobacteria	16164	1097823	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439		
bin.G1.875	55.42	7.228	28.57	0.298	Bacteria	38518	629562	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	Pelagibacter A		
bin.G1.877	96.32	2.205	0	0.295	Bacteria	110124	1859792	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-002723295	GCA-002723295 sp002690805	
bin.G1.878	72.41	2.205	100	0.307	Bacteria	8005	1017330	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8	MAG-121220-bin8 sp002705485	
bin.G1.88	68.16	1.891	50	0.362	Bacteria	5625	1824177	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952		
bin.G1.880	89.02	0	40	0.358	Bacteria	23106	2346465	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952		
bin.G1.882	81.46	9.247	78.57	0.347	Bacteria	6574	2836675	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066			
bin.G1.890	67.51	1.499	30	0.361	algicola	4262	2720391	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Zunongwangia	Zunongwangia profunda	
bin.G1.893	66.38	6.4	90	0.472	Euryarchaeota	9290	1644138	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L2	MGIIa-L2 sp013911465	
bin.G1.895	80.64	1.971	50	0.501	Bacteria	11697	2156639	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16		
bin.G1.900	61.40	7.017	83.33	0.314	Bacteria	7348	1569014	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02		
bin.G1.901	87.29	3.423	57.14	0.357	Bacteria	20976	2352196	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952	UBA952 sp003331365	
bin.G1.902	92.89	1.366	0	0.383	Bacteria	49604	3228274	Bacteria	Bacteroidota	Rhodothermia	Balneolales				
bin.G1.905	55.00	0.436	100	0.617	Alphaproteobacteria	5523	1154191	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309		
bin.G1.910	72.59	1.077	33.33	0.388	algicola	5800	2144930	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Gramella		
bin.G1.916	81.25	4.178	8.33	0.294	Bacteria	7873	1546369	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2712715		
bin.G1.94	98.19	0.217	0	0.473	Alphaproteobacteria	156846	2325163	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	UBA1172	UBA12202	UBA12202 sp002691725	
bin.G1.98	61.46	3.584	33.33	0.333	Bacteria	5753	1147794	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11	GCA-2711125		
bin.G2.103	86.27	1.006	75	0.357	Flavobacteriaceae	37121	2245717	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flavicella		
bin.G2.106	92.04	0.747	60	0.407	Flavobacteriaceae	78577	1693194	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA11891	UBA11891 sp003533785	
bin.G2.109	52.40	6.626	0	0.282	Bacteria	49342	660848	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae			
bin.G2.114	67.24	9.482	0	0.303	Bacteria	33767	1552708	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	AAA536-G10	AAA536-G10		
bin.G2.116	60.50	0	0	0.278	Bacteria	7973	935376	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	MED-G14 sp003331885	
bin.G2.12	91.78	0.918	100	0.36	Bacteria	28624	2509915	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952		
bin.G2.122	69.15	0	0	0.473	Euryarchaeota	49290	1173336	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O5	MGIIb-O5 sp002506825	
bin.G2.124	91.32	1.379	37.5	0.336	Flavobacteriaceae	19318	2810155	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flavicella		
bin.G2.127	63.60	1.724	100	0.362	Bacteria	6986	1392123	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UlvibacterB		
bin.G2.130	68.00	0.8	100	0.515	Euryarchaeota	30134	1803008	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp009887095	

bin.G2.131	55.17	3.448	100	0.417	Bacteria	4698	845254	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA724	
bin.G2.135	81.78	0.99	20	0.306	algicola	122701	1472238	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	
bin.G2.147	91.77	2.765	77.78	0.484	Alphaproteobacteria	12280	2459850	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	Puniceispirillum	
bin.G2.150	66.66	8.602	30	0.294	Bacteria	12359	1302577	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	
bin.G2.153	71.79	1.612	100	0.55	Bacteria	11162	1442513	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663	
bin.G2.155	74.34	4.961	18.75	0.386	algicola	13834	1266307	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	UBA7446 sp002698745
bin.G2.161	94.97	0	0	0.481	Alphaproteobacteria	43909	2096047	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae		
bin.G2.164	89.73	8.167	93.55	0.553	Bacteria	10318	2168569	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16	
bin.G2.166	50.65	7.243	59.26	0.347	Gammaproteobacteria	7140	984310	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	D2472	D2472 sp902599315
bin.G2.17	92.58	0.537	100	0.309	Bacteria	33504	2327718	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	SP287	
bin.G2.170	58.94	0	0	0.344	Bacteria	23739	1467824	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	
bin.G2.171	66.72	3.773	0	0.288	Bacteria	43903	619881	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	PelagibacterA	PelagibacterA sp002170125
bin.G2.172	90.26	0.614	83.33	0.36	algicola	17423	1741999	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Hel1-33-131	
bin.G2.174	62.93	3.448	100	0.45	Bacteria	15977	1396161	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	UBA10364 sp003045825
bin.G2.175	60.23	1.194	71.43	0.378	Flavobacteriaceae	4383	1285967	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae		
bin.G2.177	54.02	0.74	50	0.327	Gammaproteobacteria	32989	856327	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SAR86A	
bin.G2.178	64.09	0.426	0	0.412	Alphaproteobacteria	5706	1323164	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA5951	UBA5951 sp003332015
bin.G2.181	97.04	0	0	0.435	Bacteria	39878	2051625	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA974	
bin.G2.183	66.09	1.724	0	0.313	Bacteria	20615	922090	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	MED-G85	MED-G85 sp003331505
bin.G2.184	74.69	3.245	78.57	0.365	algicola	9692	1517099	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	UBA3537 sp001735715
bin.G2.191	99.26	7.352	61.11	0.382	Bacteria	91994	2115146	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	
bin.G2.194	53.85	1.965	16.67	0.472	Alphaproteobacteria	6446	1058622	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G2.20	56.89	6.896	0	0.292	Bacteria	8315	1300536	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	MED-G13 sp002697255
bin.G2.202	81.01	0.476	100	0.39	Bacteroidetes	23607	1905677	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	
bin.G2.203	57.68	8.191	29.63	0.362	Gammaproteobacteria	7267	1277112	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	SAR86	GCA-2707915	
bin.G2.219	61.41	1.075	100	0.457	Bacteria	6488	1216719	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	
bin.G2.222	72.45	6.559	67.65	0.34	algicola	10340	1741671	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	
bin.G2.225	65.96	4.385	100	0.314	Bacteria	8506	1709781	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02	
bin.G2.228	82.75	5.172	95.65	0.362	Bacteria	43141	1351629	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	
bin.G2.23	71.22	0	0	0.317	Bacteria	16130	1832187	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter	
bin.G2.233	61.37	2.712	47.06	0.309	Flavobacteriaceae	13045	1345211	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02	
bin.G2.237	64.82	1.724	100	0.406	Bacteria	32419	1132981	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	UBA7446 sp002862645
bin.G2.239	58.66	4.693	50	0.331	Gammaproteobacteria	9867	953677	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SAR86A	SAR86A sp902557965
bin.G2.24	69.85	7.352	9.09	0.294	Bacteria	10910	1641834	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-002723295	
bin.G2.245	53.53	6.451	93.33	0.56	Bacteria	4569	1514021	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663	
bin.G2.25	93.58	3.676	100	0.382	Bacteria	36164	1966194	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	MS024-2A sp002167945
bin.G2.252	90.05	1.075	100	0.346	Bacteria	36084	1873405	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	MED-G21	
bin.G2.256	85.55	2.379	55.56	0.608	Alphaproteobacteria	8136	2192929	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	UBA8309 sp002457745
bin.G2.258	82.16	9.705	76.92	0.367	algicola	7731	1834712	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UlvibacterB	
bin.G2.259	98.25	0	0	0.573	Alphaproteobacteria	70840	2574005	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	
bin.G2.261	95.85	1.746	100	0.471	Alphaproteobacteria	23213	1996968	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA4588	
bin.G2.263	83.95	4.478	71.43	0.323	Bacteria	9802	1686672	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	
bin.G2.267	50.07	2.253	20	0.326	Gammaproteobacteria	28499	908769	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SAR86A	

bin.G2.270	72.40	0.4	100	0.437	Euryarchaeota	14038	1515679	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1	MGIIa-K1 sp002701145
bin.G2.276	54.45	0.88	0	0.361	algicola	8434	1023158	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	
bin.G2.277	75.16	1.594	41.67	0.289	Flavobacteriaceae	34702	1486976	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8	MAG-121220-bin8 sp002700465
bin.G2.285	56.65	0.862	100	0.34	Gammaproteobacteria	16252	713225	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112	TMED112 sp004321845
bin.G2.287	51.51	7.959	16.42	0.334	algicola	5971	1402416	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	
bin.G2.290	83.33	0.8	100	0.489	Euryarchaeota	85404	1889099	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp002506275
bin.G2.292	97.84	7.078	26.32	0.309	Bacteria	54678	2330987	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	SP287	
bin.G2.295	80.31	4.213	100	0.411	Bacteria	6557	1865651	Bacteria	Bacteroidota	Rhodothermia	Balneolales	Balneolaceae	UBA1275	UBA1275 sp002457365
bin.G2.308	61.86	1.213	57.14	0.29	Flavobacteriaceae	4616	1214352	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter	
bin.G2.310	97.84	0	0	0.575	Bacteria	75889	2159506	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663	UBA11663 sp002469765
bin.G2.315	65.51	1.724	100	0.296	Bacteria	16430	1722023	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	MED-G20	MED-G20 sp002457645
bin.G2.318	58.69	8.771	88.89	0.399	algicola	5278	1188258	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA724	UBA724 sp002430545
bin.G2.32	66.03	7.547	0	0.285	Bacteria	51138	6925 34	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	MED-G40	
bin.G2.321	52.83	9.433	16.67	0.293	Bacteria	89785	664604	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae		
bin.G2.323	96.21	3.213	93.33	0.393	Bacteria	21235	1920998	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	UBA8316 sp002390455
bin.G2.324	75.41	1.881	100	0.355	Bacteria	5760	1431131	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	UBA6772	
bin.G2.325	78.40	0	0	0.417	Euryarchaeota	23673	1834496	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	
bin.G2.327	70.77	6.699	64.29	0.293	Bacteria	27889	1163883	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G11	MED-G11 sp004213645
bin.G2.329	60.47	1.149	0	0.33	Gammaproteobacteria	19658	817220	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	AG-339-G14	AG-339-G14 sp902522825
bin.G2.333	51.40	9.244	10.31	0.396	algicola	11456	2163573	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	
bin.G2.339	51.81	9.326	20.93	0.421	Bacteria	8310	2898391	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	
bin.G2.34	70.85	2.573	50	0.273	Bacteria	9520	1412557	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-2719315	
bin.G2.38	81.32	5.112	81.25	0.4	Flavobacteriaceae	16709	1326453	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	UBA7446 sp002470745
bin.G2.39	62.65	6.024	42.86	0.298	Bacteria	21201	675132	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	GCA-2704625	
bin.G2.4	87.18	2.162	75	0.364	Bacteria	11910	2022009	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952	
bin.G2.41	85.38	3.439	66.67	0.307	Flavobacteriaceae	9467	2400375	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter	
bin.G2.45	52.63	5.263	0	0.298	Bacteria	26814	551477	Bacteria	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Actinomarina	
bin.G2.49	96.37	2.183	80	0.567	Alphaproteobacteria	54229	2531354	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	UBA8309 sp002683535
bin.G2.50	50.45	0.144	0	0.296	Alphaproteobacteria	19532	731807	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	AAA536-G10	AAA536-G10	
bin.G2.52	55.21	0.873	66.67	0.477	Alphaproteobacteria	5342	781644	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G2.55	99.46	1.124	0	0.315	Bacteria	103716	1693870	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7312		
bin.G2.60	70.27	2.074	18.75	0.304	Flavobacteriaceae	19393	1651433	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02	
bin.G2.64	50.15	0.862	100	0.443	Bacteria	5847	1119228	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	
bin.G2.66	71.73	0.806	0	0.464	Bacteria	4487	1596486	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16	
bin.G2.7	93.96	0.165	100	0.385	algicola	52533	1923444	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-120531	
bin.G2.71	51.36	1.724	100	0.388	Bacteria	10643	1296058	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7428	
bin.G2.80	87.72	2.15	75	0.32	Bacteria	9601	1401895	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	BACL11		
bin.G2.91	70.06	2.197	28.57	0.494	Alphaproteobacteria	10190	1027821	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G2.92	59.83	3.932	60	0.3	algicola	12019	1149251	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	
bin.G2.94	50.36	6.406	60	0.49	Alphaproteobacteria	6320	875988	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G2.95	59.78	4.43	54.17	0.358	Gammaproteobacteria	13924	1327204	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	SAR86	GCA-2707915	GCA-2707915 sp004214065
bin.G2.97	91.67	1.965	80	0.573	Alphaproteobacteria	25271	2197176	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	UBA8309 sp002457765
bin.G3.101	89.78	1.192	66.67	0.376	algicola	20840	1818450	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	UBA3537 sp002709185

bin.G3.104	57.39	0	0	0.386	Bacteria	12350	1647041	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7428		
bin.G3.114	57.28	1.971	63.64	0.335	algicola	6166	1182637	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A		
bin.G3.116	60.34	0	0	0.277	Bacteria	12780	890837	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	MED-G14 sp003331885	
bin.G3.118	74.91	7.72	55.93	0.393	algicola	19832	1388911	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446		
bin.G3.119	99.6	0	0	0.368	Bacteria	139922	2359679	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A		
bin.G3.122	94.66	0	0	0.311	Archaea	23705	1136179	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosopumilus	Nitrosopumilus sp002690535	
bin.G3.142	62.86	1.293	100	0.338	Gammaproteobacteria	16114	700511	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112		
bin.G3.145	69.61	0.33	0	0.373	algicola	56855	1477997	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316		
bin.G3.146	88.14	0.11	0	0.362	algicola	66639	1832636	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	MS024-2A sp009886625	
bin.G3.147	83.85	0.687	60	0.372	algicola	22573	2249597	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A		
bin.G3.15	51.89	3.793	33.33	0.409	Bacteria	6276	1022281	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446		
bin.G3.158	58.14	1.8	28.57	0.339	Gammaproteobacteria	36252	929064	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	AG-339-G14	AG-339-G14 sp003282105	
bin.G3.184	94.11	0	0	0.368	Bacteria	166929	2469628	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A		
bin.G3.185	85.76	2.084	64.29	0.37	algicola	9784	2099446	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3478		
bin.G3.189	91.07	0.132	25	0.383	algicola	74851	1739187	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	UBA3537 sp002725015	
bin.G3.19	98.65	0	0	0.61	Bacteria	122182	2337366	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA8752	UBA8752 sp002172485	
bin.G3.194	59.23	2.557	37.5	0.334	algicola	11851	1190757	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316		
bin.G3.195	63.76	0	0	0.503	Bacteria	12351	1659291	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	Puniceispirillum		
bin.G3.197	51.92	5.422	5.56	0.328	Bacteria	11532	558019	Bacteria	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Actinomarina	Actinomarina sp902555055	
bin.G3.199	90.77	5.427	44.44	0.457	Bacteria	21844	4686669	Bacteria	SAR324	SAR324	SAR324	NAC60-12	JCVI-SCAAA005	JCVI-SCAAA005 sp002450295	
bin.G3.217	66.61	0.856	42.86	0.299	algicola	15461	1101052	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	MED-G13 sp002457735	
bin.G3.233	75.97	0	0	0.293	Bacteria	11304	1336518	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	GCA-002723295	GCA-002723295 sp002711185	
bin.G3.25	82.65	1.033	80	0.494	Alphaproteobacteria	15832	1991589	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	Puniceispirillum	Puniceispirillum marinum	
bin.G3.250	98.16	0.762	33.33	0.359	Bacteria	47868	2140731	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	MS024-2A sp002457295	
bin.G3.254	81.40	4.63	81.48	0.314	Flavobacteriaceae	10556	1885709	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02		
bin.G3.256	83.44	2.662	66.67	0.403	Flavobacteriaceae	10666	1744584	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	UBA8316 sp003538555	
bin.G3.262	81.20	0	0	0.514	Euryarchaeota	34008	1810132	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp002502605	
bin.G3.268	72.41	0	0	0.586	Bacteria	11471	1503775	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663		
bin.G3.269	86.59	7.132	85	0.409	Alphaproteobacteria	8381	2066679	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA5951	UBA5951 sp003332015	
bin.G3.272	69.46	7.864	73.33	0.341	algicola	10101	1702170	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316		
bin.G3.275	78.31	2.667	91.67	0.325	algicola	11735	1922946	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MedPE-SWsnd-G2		
bin.G3.276	58.11	4.25	72.73	0.329	Bacteria	9522	826857	Bacteria	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Actinomarina	Actinomarina sp004213405	
bin.G3.285	79.16	0.353	0	0.29	Flavobacteriaceae	50420	1445048	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8		
bin.G3.29	58.66	0	0	0.306	Bacteria	9748	1076601	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8		
bin.G3.296	90.94	0	0	0.375	Bacteria	12995	1835407	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A		
bin.G3.298	90.90	0.367	100	0.292	Bacteria	42643	1444810	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8	MAG-121220-bin8 sp004214185	
bin.G3.299	96.39	0	88.57	0.523	Bacteria	112198	2313733	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16	UA16 sp002690915	
bin.G3.300	91.39	5.483	76.92	0.298	Bacteria	16497	2097759	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	MED-G20	MED-G20 sp002457645	
bin.G3.303	95.39	0.537	100	0.559	Bacteria	26656	2032508	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663		
bin.G3.304	72.46	8.652	62.96	0.361	Gammaproteobacteria	12939	1414525	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	D2472		
bin.G3.310	97.42	0	0	0.291	Bacteria	77151	1744737	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8		
bin.G3.314	79.41	4.806	66.67	0.285	Bacteria	18841	1240729	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	MED-G14 sp002457715	
bin.G3.322	91.53	1.637	72.73	0.365	algicola	20883	1697384	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	UBA3537 sp001735715	

bin.G3.323	93.65	5.862	47.83	0.542	Alphaproteobacteria	18319	2519938	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	MED-G116	MED-G116 sp004212735
bin.G3.325	97.42	1.006	25	0.354	Flavobacteriaceae	129835	2216348	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter	
bin.G3.330	66.71	7.784	0	0.329	Bacteria	9129	859612	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SAR86A	SAR86A sp902557965
bin.G3.333	66.35	0.605	33.33	0.374	algicola	6001	1460501	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7428	
bin.G3.336	85.39	0.58	50	0.353	Bacteria	6304	2093679	Bacteria	Bacteroidota	Rhodothermia	Balneolales	Balneolaceae	RHLJ01	
bin.G3.337	65.91	4.139	7.14	0.302	algicola	13696	1449828	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	MED-G13 sp002691265
bin.G3.346	54.73	2.604	66.67	0.334	Gammaproteobacteria	16864	1120592	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	MED-G78	
bin.G3.353	93.10	7.915	12.5	0.294	Bacteria	29833	1390123	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	MED-G14 sp004321735
bin.G3.354	93.54	0.806	50	0.349	Bacteria	159265	1479739	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7312	UBA8444	UBA8444 sp003454845
bin.G3.356	80.54	1.881	100	0.447	Bacteria	6077	2703233	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16		
bin.G3.360	80.74	1.756	50	0.33	Gammaproteobacteria	18813	1126352	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	MEDG-81	MEDG-81 sp003331625
bin.G3.369	65.48	2.586	100	0.477	Bacteria	14108	1477460	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae		
bin.G3.37	87.60	0.873	100	0.603	Alphaproteobacteria	21945	2142942	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	
bin.G3.371	88.60	1.344	100	0.457	Bacteria	15935	1713497	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	UBA10364 sp013911625
bin.G3.387	53.34	1.742	91.67	0.347	Gammaproteobacteria	12499	926117	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	D2472	D2472 sp902599315
bin.G3.39	93.14	0.561	0	0.333	algicola	17299	2332812	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Winogradskyella	
bin.G3.393	86.72	3.125	20	0.304	Bacteria	34306	1490670	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	MED-G14 sp003331875
bin.G3.398	56.89	0.862	100	0.293	Bacteria	16956	879767	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G11	
bin.G3.410	84.93	0	0	0.357	Euryarchaeota	91618	1552847	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O3	MGIIb-O3 sp002457145
bin.G3.42	70.56	3.93	41.18	0.492	Alphaproteobacteria	8240	1186249	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA3439	
bin.G3.431	85.60	0.042	100	0.51	Euryarchaeota	27411	1891689	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	Poseidonia	
bin.G3.437	75.07	0.33	0	0.404	algicola	50635	1309189	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3478	
bin.G3.443	84.13	0.8	100	0.4	Euryarchaeota	36256	1867446	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp002172355
bin.G3.463	77.61	9.872	52	0.434	Bacteria	5205	2184211	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	
bin.G3.469	84.64	5.529	3.2	0.411	Bacteria	9062	3908475	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	UBA7430	
bin.G3.61	68.16	0.596	71.43	0.608	Alphaproteobacteria	5758	1587695	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	
bin.G3.65	77.97	1.6	100	0.499	Euryarchaeota	16501	1770836	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	
bin.G3.7	80.68	1.59	27.27	0.295	Flavobacteriaceae	36185	1611557	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8	MAG-121220-bin8 sp003331265
bin.G3.70	70.97	3.312	86.36	0.427	algicola	16285	1058969	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA724	UBA724 sp002171575
bin.G3.86	89.72	0.135	100	0.379	Bacteria	23499	2155961	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952	UBA952 sp002167775
bin.G3.90	91.45	0	0	0.563	Bacteria	19602	2318363	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA8752	
bin.G4.108	51.42	4.901	35.53	0.347	Proteobacteria	16128	997800	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	SAR86	GCA-2707915	
bin.G4.116	80.28	0.655	0	0.462	Alphaproteobacteria	7938	2068653	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA4588	
bin.G4.120	93.12	4.126	85.71	0.474	Alphaproteobacteria	12371	1875581	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	HIMB100	HIMB100 sp002700485
bin.G4.123	81.25	1.06	25	0.37	Flavobacteriaceae	14237	2252534	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	
bin.G4.125	98.97	1.776	0	0.37	algicola	36078	3737024	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Salegentibacter	
bin.G4.129	70.98	2.386	69.23	0.366	Flavobacteriaceae	5898	1967129	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02	
bin.G4.143	70.53	0.806	100	0.564	Bacteria	6634	1458536	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA8752	
bin.G4.146	92.14	1.082	33.33	0.339	Flavobacteriaceae	29977	2010144	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae		
bin.G4.161	70.95	1.029	33.33	0.351	Bacteria	4731	2072100	Bacteria	Bacteroidota	Rhodothermia	Balneolales	Balneolaceae		
bin.G4.174	54.11	1.402	85.71	0.394	algicola	18442	840388	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	
bin.G4.185	55.40	5.011	78.57	0.416	algicola	5609	938169	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	
bin.G4.189	70.84	2.366	83.33	0.356	Gammaproteobacteria	13556	1298621	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	MED-G82	MED-G82 sp003331565

bin.G4.19	86.18	1.237	75	0.549	Alphaproteobacteria	21247	1975551	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	HIMB100		
bin.G4.190	76.80	5.235	73.08	0.312	Flavobacteriaceae	6961	1982988	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02		
bin.G4.191	82.00	0	0	0.379	Euryarchaeota	55439	1502541	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-O2	MGIIb-O2 sp002498985	
bin.G4.194	91.97	3.225	100	0.448	Bacteria	16615	3071314	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16			
bin.G4.196	99.46	0.716	100	0.372	Bacteria	45867	1762771	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7312	GCA-2862585		
bin.G4.199	50.86	0	0	0.613	Bacteria	5212	1265304	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309		
bin.G4.200	97.31	0.335	0	0.374	Flavobacteriaceae	122063	3601575	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	SCGC-AAA160-P02		
bin.G4.210	63.90	3.459	94.44	0.328	algicola	5552	1543518	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MedPE-SWsnd-G2		
bin.G4.212	63.48	3.611	90.91	0.471	Euryarchaeota	6760	1654517	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L2	MGIIa-L2 sp013911465	
bin.G4.230	78.53	0.8	33.33	0.449	Euryarchaeota	24108	1769246	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L2		
bin.G4.232	96.36	0.33	100	0.372	algicola	103590	1842774	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537		
bin.G4.238	75.58	4.361	0	0.336	algicola	11453	1770009	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316		
bin.G4.241	60.34	4.584	100	0.446	Euryarchaeota	2012	1910762	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1		
bin.G4.249	57.43	1.111	40	0.299	algicola	4680	1220211	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MedPE-SWsnd-G2		
bin.G4.25	58.18	0.156	100	0.394	Bacteria	8369	1641452	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7428		
bin.G4.252	66.28	5.842	38.46	0.327	Gammaproteobacteria	14426	1182089	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	MED-G78	MED-G78 sp003331645	
bin.G4.255	86.28	1.344	100	0.613	Bacteria	30080	2175123	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA8752		
bin.G4.259	88.92	0.647	0	0.388	Flavobacteriaceae	96390	1832483	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316		
bin.G4.260	98.38	0.663	78.57	0.349	Bacteria	65259	1979459	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430			
bin.G4.268	81.50	0.478	75	0.367	algicola	14914	1546473	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537		
bin.G4.27	90.92	1.155	50	0.357	algicola	32376	2087218	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae			
bin.G4.274	82.47	0.105	0	0.444	Bacteria	8387	3041520	Bacteria	SAR324	SAR324	SAR324	NAC60-12	JCVI-SCAAA005	JCVI-SCAAA005 sp000224765	
bin.G4.279	91.95	2.022	75	0.364	Bacteria	54119	1921664	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7428		
bin.G4.287	79.20	0	0	0.515	Euryarchaeota	45608	1766946	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	MGIIa-L1 sp002495535	
bin.G4.29	52.71	0.412	0	0.369	algicola	4866	1031286	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537		
bin.G4.30	80.32	0	0	0.474	Bacteria	24407	1205424	Bacteria	Bacteroidota	Rhodothermia	Balneolales	Balneolaceae	UBA8296	UBA8296 sp002338335	
bin.G4.302	99.46	0.107	0	0.368	Bacteria	40462	1623755	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7312	UBA7312		
bin.G4.305	89.29	1.463	36.36	0.361	Flavobacteriaceae	13125	2180972	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter		
bin.G4.311	81.67	0.8	100	0.5	Euryarchaeota	148554	1507167	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Thalassarchaeaceae	MGIIb-N1		
bin.G4.315	87.96	7.111	72.73	0.57	Alphaproteobacteria	14335	2466966	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	UBA8309 sp002457765	
bin.G4.318	78.42	0.158	0	0.378	Bacteroidetes	7773	1505727	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7428		
bin.G4.327	53.63	3.964	77.78	0.4	algicola	8128	8795 10	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446		
bin.G4.342	61.20	0.862	100	0.321	Bacteria	19922	950951	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SAR86A	SAR86A sp004212975	
bin.G4.347	84.81	2.347	77.78	0.358	Bacteria	11504	1533574	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	TMED14	TMED14 sp002167805	
bin.G4.349	78.84	0.537	100	0.342	Bacteria	7737	2790162	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	GCA-002722245	GCA-002722245		
bin.G4.351	93.82	3.234	80	0.332	algicola	17144	2512766	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Winogradskyella	Winogradskyella sp003335675	
bin.G4.366	85.23	5.286	75	0.444	Bacteria	10911	1599434	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	UBA10364 sp002387615	
bin.G4.367	57.46	1.218	77.78	0.351	algicola	4131	1502601	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Mesonia	Mesonia mobilis	
bin.G4.372	99.82	0.537	100	0.379	Bacteria	148185	1776784	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	UBA7430		
bin.G4.382	84.60	0.716	0	0.298	Bacteria	30078	2030093	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	MED-G20	MED-G20 sp002691605	
bin.G4.384	63.32	0.754	37.5	0.382	algicola	30271	1070602	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	UBA7446 sp002698745	
bin.G4.39	60.34	2.586	11.11	0.475	Bacteria	13723	1599644	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae			
bin.G4.397	62.06	3.448	0	0.316	Bacteria	47691	838213	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	SAR86A	SAR86A sp002690725	

bin.G4.401	89.65	6.182	100	0.343	Bacteria	10377	1834742	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	MED-G21	MED-G21 sp002457305
bin.G4.402	94.98	0	0	0.396	Bacteria	59748	2335407	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	
bin.G4.408	54.42	3.225	0	0.32	Bacteria	8886	775456	Bacteria	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Actinomarina	
bin.G4.41	76.70	6.628	82.05	0.388	algicola	9404	1870893	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3478	UBA3478 sp011525015
bin.G4.410	57.75	4.31	100	0.62	Bacteria	9653	1661765	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	
bin.G4.413	90.86	0	0	0.352	Bacteria	199048	1795844	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	UBA6772	UBA6772 sp002685115
bin.G4.414	97.58	1.075	0	0.35	Bacteria	22180	2670644	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	GCA-002722245	GCA-002722245	GCA-002722245 sp002722245
bin.G4.423	57.75	1.724	100	0.59	Bacteria	17488	1586329	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663	
bin.G4.425	95.69	1.075	100	0.314	Bacteria	13011	3354169	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Weeksellaceae	Empedobacter	Empedobacter falseniiA
bin.G4.426	68.50	9.344	82.05	0.611	Alphaproteobacteria	5177	1903564	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	UBA8309	
bin.G4.43	53.67	1.293	0	0.322	Gammaproteobacteria	11370	652849	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112	
bin.G4.438	68.96	0	0	0.338	Bacteria	23078	924277	Bacteria	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Actinomarina	Actinomarina sp002308095
bin.G4.446	82.51	0.082	0	0.379	algicola	139900	1562247	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3537	
bin.G4.450	84.55	2.941	0	0.286	Bacteria	37409	1464674	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	
bin.G4.467	66.63	0.824	100	0.393	Flavobacteriaceae	36956	956812	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	
bin.G4.471	55.83	3.448	100	0.289	Bacteria	7792	9800404	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8	MAG-121220-bin8 sp902559035
bin.G4.472	53.44	8.463	82.35	0.307	Bacteria	8000	1316556	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	
bin.G4.477	90.75	2.037	85.71	0.485	Alphaproteobacteria	23187	2310582	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales	Puniceispirillaceae	Puniceispirillum	
bin.G4.480	95.22	1.838	66.67	0.306	Bacteria	32102	1745810	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	TMED96	TMED96 sp002171475
bin.G4.481	82.00	0	0	0.42	Euryarchaeota	109918	1918923	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L1	
bin.G4.486	92.64	0.735	0	0.404	Bacteria	73393	2012515	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	
bin.G4.488	98.92	0	0	0.503	Bacteria	159917	2368972	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16	
bin.G4.491	91.69	3.046	71.43	0.355	Bacteria	11309	2428114	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	GCA-002722245		
bin.G4.493	58.61	0.687	16.67	0.413	algicola	6112	978764	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	
bin.G4.503	62.45	0	0	0.331	Bacteria	13291	860129	Bacteria	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Actinomarina	Actinomarina sp902519215
bin.G4.505	68.96	0	0	0.602	Bacteria	19595	1810060	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663	
bin.G4.508	94.62	1.075	50	0.327	Bacteria	70922	1771097	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA7430	MED-G21	
bin.G4.513	66.07	0.766	50	0.343	Gammaproteobacteria	57430	1042800	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112	TMED112 sp002716745
bin.G4.515	79.17	0.459	50	0.364	Flavobacteriaceae	9891	1674008	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	UBA8316 sp002711215
bin.G4.518	56.89	1.724	0	0.295	Bacteria	9314	1291430	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MAG-121220-bin8	MAG-121220-bin8 sp902635895
bin.G4.519	65.75	2.566	80	0.369	algicola	8077	1311778	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UlvibacterB	
bin.G4.527	84.80	0.645	100	0.554	Bacteria	14991	1809843	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UA16	
bin.G4.529	73.86	2.284	77.27	0.588	Bacteria	14096	1604187	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16	UBA11663	
bin.G4.541	88.41	0.95	33.33	0.439	algicola	20320	1798211	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	BACL21	
bin.G4.551	52.35	0	0	0.301	Bacteria	8830	918283	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	
bin.G4.554	87.09	1.075	0	0.308	Bacteria	75427	2204505	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	MED-G20	
bin.G4.557	72.75	0	0	0.281	Bacteria	21565	1164629	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G14	
bin.G4.576	82.25	1.792	0	0.511	Bacteria	8647	1805317	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	RFXV01	
bin.G4.58	82.81	9.785	74.6	0.374	algicola	14321	2118536	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA3478	UBA3478 sp002691645
bin.G4.581	97.79	7.352	28.57	0.393	Bacteria	35996	2097813	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	UBA8316 sp002390455
bin.G4.583	61.89	7.375	66.67	0.334	Gammaproteobacteria	10301	972478	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	AG-339-G14	AG-339-G14 sp902614235
bin.G4.584	68.34	1.79	50	0.33	Gammaproteobacteria	132380	1170008	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	AG-339-G14	AG-339-G14	
bin.G4.585	79.30	0.179	0	0.44	Bacteria	5376	2297220	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UA16		

bin.G4.586	53.06	0	0	0.324	Gammaproteobacteria	25558	538659	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	TMED112	TMED112	
bin.G4.590	59.40	7.502	37.5	0.326	Bacteria	14136	806978	Bacteria	Actinobacteriota	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Actinomarina	
bin.G4.602	60.20	4.032	60	0.44	Bacteria	5221	1473970	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Schleiferiaceae	UBA10364	
bin.G4.606	62.50	8.272	1.67	0.382	algicola	6056	4505437	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MS024-2A	
bin.G4.612	52.41	9.905	2.99	0.391	Bacteria	5114	2615979	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA7446	
bin.G4.62	94.35	5.645	11.76	0.311	Bacteria	26861	2449930	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	SP287	
bin.G4.63	54.67	7.009	45.45	0.468	Archaea	9819	1426846	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L2	MGIIa-L2 sp002722615
bin.G4.67	77.50	0.759	60	0.465	algicola	6126	2284066	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aureicoccus	
bin.G4.68	80.40	0.8	100	0.541	Euryarchaeota	30813	1818812	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	Poseidonia	Poseidonia sp002704515
bin.G4.70	72.40	0.8	0	0.613	Euryarchaeota	77746	1768526	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-L3	MGIIa-L3 sp11892u
bin.G4.77	65.91	0	0	0.323	Gammaproteobacteria	90079	1149437	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86	D2472	CACEJU01	CACEJU01 sp902559885
bin.G4.8	61.20	1.724	0	0.293	Bacteria	12030	967677	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	MED-G13	MED-G13 sp902510415
bin.G4.86	73.86	0	0	0.401	Bacteria	79652	1768871	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	UBA8316	
bin.G4.87	84.40	0	0	0.43	Euryarchaeota	32614	1861113	Archaea	Thermoplasmatota	PoseidoniiaA	Poseidoniales	Poseidoniaceae	MGIIa-K1	MGIIa-K1 sp002689565
bin.G4.90	67.38	7.634	42.31	0.296	Bacteria	8622	1670019	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	UBA10066	MED-G20	
bin.G4.93	92.02	2.882	90	0.364	Bacteria	11710	2309746	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales	Crocinitomicaceae	UBA952	

Table 5. Summary of the selected 169 MAGs based on horizontal coverage across BBMO and TARA samples. Seasonality was computed with the lomb-scargle periodogram algorithm.

Table too large to fit. Available on-line at: https://doi.org/10.5281/zenodo.7078952

Table 6. Number of genes with a mean pNpS > 0.8 across samples for each dataset and MAG.

MAG	nº predicted genes	nº genes BBMO	nº genes SOLA	nº genes TARA	nº genes shared in BBMO&SOLA	% shared genes in BBMO/SOLA	Taxonomy (order)	Genome size (bp)	genes/Mb BBMO	genes/Mb SOLA	genes/Mb TARA
bin.G3.250	2014	147	128	61	98	55.4	o_Flavobacteriales	2140731	68.67	59.79	28.49
bin.G3.398	935	35	26	1	20	48.8	o_Flavobacteriales	879767	39.78	29.55	1.14
bin.G1.297	941	35	29	5	12	23.1	o_SAR86	909753	38.47	31.88	5.50
bin.G4.315	2512	86	103	43	51	37.0	o_SAR116	2466966	34.86	41.75	17.43
bin.G2.97	2135	76	81	61	43	37.7	o_SAR116	2197176	34.59	36.87	27.76
bin.G3.314	1265	41	32	3	27	58.7	o_Flavobacteriales	1240729	33.05	25.79	2.42
bin.G3.142	768	21	17	0	12	46.2	o_SAR86	700511	29.98	24.27	0.00
bin.G2.80	1429	41	35	16	4	5.6	o_Flavobacteriales	1401895	29.25	24.97	11.41
bin.G3.323	2568	70	70	27	44	45.8	o_SAR116	2519938	27.78	27.78	10.71
bin.G2.327	1188	31	17	4	15	45.5	o_Flavobacteriales	1163883	26.63	14.61	3.44
bin.G3.199	4680	113	141	83	73	40.3	o_SAR324	4686669	24.11	30.09	17.71
bin.G1.197	1184	32	19	10	16	45.7	s_Archaea	1354925	23.62	14.02	7.38
bin.G4.185	1030	22	13	10	9	34.6	o_Flavobacteriales	938169	23.45	13.86	10.66
bin.G2.277	1414	34	0	4	0	0.0	o_Flavobacteriales	1486976	22.87	0.00	2.69
bin.G3.269	2256	46	59	35	29	55.2	o_SAR116	2066679	22.26	28.55	16.94
bin.G4.480	1672	38	38	4	24	63.2	o_Flavobacteriales	1745810	21.77	21.77	2.29
bin.G2.177	924	18	13	1	8	51.6	o_SAR86	856327	21.02	15.18	1.17
bin.G3.256	1743	36	33	32	25	72.5	o_Flavobacteriales	1744584	20.64	18.92	18.34

bin.G4.413	1574	37	0	9	0	0.0	o_Flavobacteriales	1795844	20.60	0.00	5.01
bin.G1.656	1224	27	20	2	12	51.1	s_Archaea	1358477	19.88	14.72	1.47
bin.G4.586	599	10	12	0	5	45.5	o_SAR86	538659	18.56	22.28	0.00
bin.G2.178	1448	24	32	10	13	46.4	o_SAR116	1323164	18.14	24.18	7.56
bin.G1.332	1580	28	21	7	9	36.7	o_Flavobacteriales	1583954	17.68	13.26	4.42
bin.G2.267	982	16	19	4	7	40.0	o_SAR86	908769	17.61	20.91	4.40
bin.G4.311	1291	26	2	3	1	7.1	s_Archaea	1507167	17.25	1.33	1.99
bin.G1.319	1100	20	0	0	0	0.0	o_Flavobacteriales	1178088	16.98	0.00	0.00
bin.G2.318	1226	20	18	4	14	73.7	o_Flavobacteriales	1188258	16.83	15.15	3.37
bin.G3.354	1356	24	22	2	3	13.0	o_Flavobacteriales	1479739	16.22	14.87	1.35
bin.G1.748	1422	25	34	0	13	44.1	o_Flavobacteriales	1570526	15.92	21.65	0.00
bin.G4.238	1683	28	18	3	14	60.9	o_Flavobacteriales	1770009	15.82	10.17	1.69
bin.G4.8	968	15	14	3	7	48.3	o_Flavobacteriales	967677	15.50	14.47	3.10
bin.G2.276	1038	15	1	5	0	0.0	o_Flavobacteriales	1023158	14.66	0.98	4.89
bin.G3.70	1086	15	13	1	3	21.4	o_Flavobacteriales	1058969	14.16	12.28	0.94
bin.G1.719	1489	25	21	0	12	52.2	s_Archaea	1784640	14.01	11.77	0.00
bin.G2.116	1029	13	11	7	8	66.7	o_Flavobacteriales	935376	13.90	11.76	7.48
bin.G3.310	1615	24	20	1	15	68.2	o_Flavobacteriales	1744737	13.76	11.46	0.57
bin.G2.181	1769	28	29	12	18	63.2	o_Flavobacteriales	2051625	13.65	14.14	5.85
bin.G2.122	1025	16	16	16	12	75.0	s_Archaea	1173336	13.64	13.64	13.64
bin.G4.590	931	11	5	3	2	25.0	o_Actinomarinales	806978	13.63	6.20	3.72
bin.G3.393	1430	20	23	8	12	55.8	o_Flavobacteriales	1490670	13.42	15.43	5.37
bin.G1.774	1557	22	18	1	8	40.0	o_Flavobacteriales	1648652	13.34	10.92	0.61
bin.G2.252	1620	24	10	3	0	0.0	o_Flavobacteriales	1873405	12.81	5.34	1.60
bin.G2.131	908	10	11	4	5	47.6	o_Flavobacteriales	845254	11.83	13.01	4.73
bin.G1.651	1856	22	11	5	7	42.4	o_Flavobacteriales	1912719	11.50	5.75	2.61
bin.G4.401	1734	21	1	3	0	0.0	o_Flavobacteriales	1834742	11.45	0.55	1.64
bin.G2.34	1457	16	8	3	0	0.0	o_Flavobacteriales	1412557	11.33	5.66	2.12
bin.G3.233	1356	15	12	7	8	59.3	o_Flavobacteriales	1336518	11.22	8.98	5.24
bin.G4.384	1003	12	14	4	7	53.8	o_Flavobacteriales	1070602	11.21	13.08	3.74
bin.G3.7	1530	18	15	6	10	60.6	o_Flavobacteriales	1611557	11.17	9.31	3.72
bin.G2.155	1216	14	9	5	5	43.5	o_Flavobacteriales	1266307	11.06	7.11	3.95
bin.G2.183	1024	10	18	3	7	50.0	o_SAR86	922090	10.84	19.52	3.25
bin.G3.300	1953	22	30	6	12	46.2	o_Flavobacteriales	2097759	10.49	14.30	2.86
bin.G2.222	1649	18	12	2	8	53.3	o_Flavobacteriales	1741671	10.33	6.89	1.15
bin.G4.557	1168	12	13	5	8	64.0	o_Flavobacteriales	1164629	10.30	11.16	4.29
bin.G2.295	1808	19	18	21	8	43.2	o_Balneolales	1865651	10.18	9.65	11.26
bin.G2.263	1717	17	14	5	12	77.4	o_Flavobacteriales	1686672	10.08	8.30	2.96
bin.G3.122	1492	11	7	1	5	55.6	s_Archaea	1136179	9.68	6.16	0.88
bin.G3.337	1475	14	15	1	8	55.2	o_Flavobacteriales	1449828	9.66	10.35	0.69
bin.G4.450	1436	14	14	3	4	28.6	o_Flavobacteriales	1464674	9.56	9.56	2.05
bin.G1.334	1077	9	8	3	5	58.8	o_SAR116	948170	9.49	8.44	3.16
bin.G1.503	768	7	0	3	0	0.0	o_SAR86	737989	9.49	0.00	4.07

bin.G4.342	1034	9	25	1	3	17.6	o_SAR86	950951	9.46	26.29	1.05
bin.G4.68	1549	17	14	9	12	77.4	s_Archaea	1818812	9.35	7.70	4.95
bin.G1.468	1253	11	15	3	7	53.8	o_SAR86	1190629	9.24	12.60	2.52
bin.G1.647	1286	12	8	2	0	0.0	o_Flavobacteriales	1318731	9.10	6.07	1.52
bin.G4.120	1946	17	31	8	5	20.8	o_SAR116	1875581	9.06	16.53	4.27
bin.G3.116	946	8	11	6	5	52.6	o_Flavobacteriales	890837	8.98	12.35	6.74
bin.G1.35	1122	11	1	4	0	0.0	s_Archaea	1277228	8.61	0.78	3.13
bin.G4.90	1649	14	17	6	3	19.4	o_Flavobacteriales	1670019	8.38	10.18	3.59
bin.G1.528	1456	14	9	5	4	34.8	s_Archaea	1714418	8.17	5.25	2.92
bin.G2.315	1625	14	14	1	6	42.9	o_Flavobacteriales	1722023	8.13	8.13	0.58
bin.G1.799	1421	11	19	8	8	53.3	o_Flavobacteriales	1388127	7.92	13.69	5.76
bin.G3.353	1410	11	12	5	8	69.6	o_Flavobacteriales	1390123	7.91	8.63	3.60
bin.G2.91	1124	8	7	4	3	40.0	o_SAR116	1027821	7.78	6.81	3.89
bin.G3.272	1642	13	14	1	7	51.9	o_Flavobacteriales	1702170	7.64	8.22	0.59
bin.G3.387	1034	7	8	4	4	53.3	o_SAR86	926117	7.56	8.64	4.32
bin.G3.276	982	6	3	2	2	44.4	o_Actinomarinales	826857	7.26	3.63	2.42
bin.G1.201	1425	10	7	2	3	35.3	o_Flavobacteriales	1397803	7.15	5.01	1.43
bin.G3.304	1521	10	9	3	7	73.7	o_SAR86	1414525	7.07	6.36	2.12
bin.G1.684	1261	8	6	3	4	57.1	o_SAR86	1141202	7.01	5.26	2.63
bin.G1.614	1627	10	5	5	1	13.3	o_SAR116	1427239	7.01	3.50	3.50
bin.G2.135	1344	10	12	3	4	36.4	o_Flavobacteriales	1472238	6.79	8.15	2.04
bin.G4.191	1349	10	11	3	4	38.1	s_Archaea	1502541	6.66	7.32	2.00
bin.G4.551	964	6	0	2	0	0.0	o_Flavobacteriales	918283	6.53	0.00	2.18
bin.G4.347	1473	10	14	4	8	66.7	o_Flavobacteriales	1533574	6.52	9.13	2.61
bin.G4.438	1024	6	11	2	3	35.3	o_Actinomarinales	924277	6.49	11.90	2.16
bin.G1.830	1216	9	2	9	2	36.4	s_Archaea	1393212	6.46	1.44	6.46
bin.G3.158	1021	6	10	3	3	37.5	o_SAR86	929064	6.46	10.76	3.23
bin.G1.415	1661	11	6	2	2	23.5	o_Flavobacteriales	1710880	6.43	3.51	1.17
bin.G1.767	1454	9	0	1	0	0.0	o_Flavobacteriales	1408074	6.39	0.00	0.71
bin.G3.217	1062	7	8	4	2	26.7	o_Flavobacteriales	1101052	6.36	7.27	3.63
bin.G1.4	1562	12	0	7	0	0.0	s_Archaea	1917671	6.26	0.00	3.65
bin.G3.298	1426	9	19	1	6	42.9	o_Flavobacteriales	1444810	6.23	13.15	0.69
bin.G1.462	730	4	2	2	1	33.3	o_SAR11	676764	5.91	2.96	2.96
bin.G3.330	988	5	6	1	3	54.5	o_SAR86	859612	5.82	6.98	1.16
bin.G1.769	1593	10	0	1	0	0.0	s_Archaea	1821344	5.49	0.00	0.55
bin.G2.45	641	3	5	2	1	25.0	o_Actinomarinales	551477	5.44	9.07	3.63
bin.G2.150	1306	7	5	4	3	50.0	o_Flavobacteriales	1302577	5.37	3.84	3.07
bin.G1.700	1993	11	2	0	1	15.4	o_Flavobacteriales	2111364	5.21	0.95	0.00
bin.G2.166	1115	5	9	0	2	28.6	o_SAR86	984310	5.08	9.14	0.00
bin.G4.252	1309	6	7	6	4	61.5	o_SAR86	1182089	5.08	5.92	5.08
bin.G3.194	1201	6	6	7	3	50.0	o_Flavobacteriales	1190757	5.04	5.04	5.88
bin.G1.825	1577	9	3	6	2	33.3	s_Archaea	1825599	4.93	1.64	3.29
bin.G4.513	1068	5	8	2	4	61.5	o_SAR86	1042800	4.79	7.67	1.92

bin.G1.392	1092	5	3	1	2	50.0	o_Flavobacteriales	1048094	4.77	2.86	0.95
bin.G2.270	1385	7	4	3	3	54.5	s_Archaea	1515679	4.62	2.64	1.98
bin.G3.346	1195	5	11	3	3	37.5	o_SAR86	1120592	4.46	9.82	2.68
bin.G1.314	995	4	8	3	1	16.7	o_SAR86	898442	4.45	8.90	3.34
bin.G1.632	1259	5	2	5	1	28.6	o_Flavobacteriales	1175242	4.25	1.70	4.25
bin.G4.408	872	3	5	0	1	25.0	o_Actinomarinales	775456	3.87	6.45	0.00
bin.G4.189	1421	5	7	0	2	33.3	o_SAR86	1298621	3.85	5.39	0.00
bin.G3.101	1720	7	3	10	0	0.0	o_Flavobacteriales	1818450	3.85	1.65	5.50
bin.G1.207	1140	4	3	1	0	0.0	o_SAR86	1043784	3.83	2.87	0.96
bin.G4.472	1289	5	7	0	4	66.7	o_Flavobacteriales	1316556	3.80	5.32	0.00
bin.G2.329	898	3	3	2	1	33.3	o_SAR86	817220	3.67	3.67	2.45
bin.G1.451	1361	5	1	3	0	0.0	o_Flavobacteriales	1370432	3.65	0.73	2.19
bin.G1.136	1230	5	0	5	0	0.0	s_Archaea	1378590	3.63	0.00	3.63
bin.G4.503	958	3	5	3	2	50.0	o_Actinomarinales	860129	3.49	5.81	3.49
bin.G1.493	1601	6	0	8	0	0.0	o_Flavobacteriales	1741584	3.45	0.00	4.59
bin.G2.94	1017	3	3	0	2	66.7	o_SAR116	875988	3.42	3.42	0.00
bin.G1.308	1038	3	1	1	1	50.0	o_SAR116	950348	3.16	1.05	1.05
bin.G4.423	1420	5	0	1	0	0.0	o_Flavobacteriales	1586329	3.15	0.00	0.63
bin.G2.239	1063	3	8	1	1	18.2	o_SAR86	953677	3.15	8.39	1.05
bin.G1.675	1026	3	0	3	0	0.0	o_SAR86	982010	3.05	0.00	3.05
bin.G2.109	727	2	0	0	0	0.0	o_SAR11	660848	3.03	0.00	0.00
bin.G1.203	1370	4	9	2	3	46.2	o_SAR86	1323048	3.02	6.80	1.51
bin.G4.108	1070	3	5	6	2	50.0	o_SAR86	997800	3.01	5.01	6.01
bin.G2.39	753	2	6	4	1	25.0	o_SAR11	675132	2.96	8.89	5.92
bin.G1.317	1283	4	1	0	1	40.0	s_Archaea	1441996	2.77	0.69	0.00
bin.G4.77	1182	3	9	0	1	16.7	o_SAR86	1149437	2.61	7.83	0.00
bin.G1.473	951	2	2	2	2	100.0	o_Flavobacteriales	810992	2.47	2.47	2.47
bin.G1.409	896	2	5	2	1	28.6	o_SAR86	813061	2.46	6.15	2.46
bin.G2.95	1426	3	12	1	2	26.7	o_SAR86	1327204	2.26	9.04	0.75
bin.G2.325	1598	4	1	1	0	0.0	s_Archaea	1834496	2.18	0.55	0.55
bin.G1.746	1525	4	31	2	3	17.1	s_Archaea	1862210	2.15	16.65	1.07
bin.G1.852	974	2	4	2	1	33.3	s_Archaea	940549	2.13	4.25	2.13
bin.G3.285	1388	3	2	2	2	80.0	o_Flavobacteriales	1445048	2.08	1.38	1.38
bin.G4.62	2214	5	1	0	0	0.0	o_Flavobacteriales	2449930	2.04	0.41	0.00
bin.G4.471	1059	2	5	1	2	57.1	o_Flavobacteriales	980004	2.04	5.10	1.02
bin.G1.485	973	2	1	1	1	66.7	s_Archaea	1107641	1.81	0.90	0.90
bin.G3.360	1231	2	6	3	1	25.0	o_SAR86	1126352	1.78	5.33	2.66
bin.G1.378	2641	5	4	2	3	66.7	o_SAR324	2992080	1.67	1.34	0.67
bin.G1.875	677	1	2	3	1	66.7	o_SAR11	629562	1.59	3.18	4.77
bin.G1.495	702	1	2	3	0	0.0	o_SAR11	633560	1.58	3.16	4.74
bin.G4.43	724	1	1	0	0	0.0	o_SAR86	652849	1.53	1.53	0.00
bin.G2.321	730	1	2	1	0	0.0	o_SAR11	664604	1.50	3.01	1.50
bin.G2.32	774	1	4	5	1	40.0	o_SAR11	692534	1.44	5.78	7.22

bin.G1.211	1534	2	1	2	1	66.7	o_SAR116	1400825	1.43	0.71	1.43
bin.G1.566	2057	3	0	0	0	0.0	o_Flavobacteriales	2103512	1.43	0.00	0.00
bin.G2.285	782	1	3	0	0	0.0	o_SAR86	713225	1.40	4.21	0.00
bin.G4.397	888	1	1	1	1	100.0	o_SAR86	838213	1.19	1.19	1.19
bin.G1.873	1134	1	1	1	1	100.0	o_SAR116	1097823	0.91	0.91	0.91
bin.G2.92	1136	1	2	2	0	0.0	o_Flavobacteriales	1149251	0.87	1.74	1.74
bin.G1.596	2132	2	1	0	0	0.0	o_Flavobacteriales	2305707	0.87	0.43	0.00
bin.G2.203	1436	1	7	0	1	25.0	o_SAR86	1277112	0.78	5.48	0.00
bin.G1.396	1433	1	1	1	0	0.0	o_SAR116	1334080	0.75	0.75	0.75
bin.G1.851	1420	1	1	0	0	0.0	o_SAR116	1344834	0.74	0.74	0.00
bin.G1.304	1681	1	1	5	0	0.0	s_Archaea	2066193	0.48	0.48	2.42
bin.G1.590	3645	1	3	11	0	0.0	o_SAR324	3667354	0.27	0.82	3.00
bin.G3.197	631	0	4	1	0	0.0	o_Actinomarinales	558019	0.00	7.17	1.79
bin.G2.171	678	0	2	2	0	0.0	o_SAR11	619881	0.00	3.23	3.23
bin.G1.849	1472	0	4	4	0	0.0	o_SAR86	1354382	0.00	2.95	2.95
bin.G4.583	1089	0	2	0	0	0.0	o_SAR86	972478	0.00	2.06	0.00
bin.G1.251	1770	0	0	8	0	0.0	o_SAR116	1634178	0.00	0.00	4.90
bin.G4.87	1600	0	0	2	0	0.0	s_Archaea	1861113	0.00	0.00	1.07
bin.G4.274	3138	0	0	3	0	0.0	o_SAR324	3041520	0.00	0.00	0.99
bin.G1.306	2482	0	0	2	0	0.0	o_SAR116	2294487	0.00	0.00	0.87
bin.G1.157	854	0	0	0	0	0.0	o_Flavobacteriales	915877	0.00	0.00	0.00
bin.G1.627	666	0	0	0	0	0.0	o_SAR11	606388	0.00	0.00	0.00

Table 7. Summary of seasonality and distribution patterns in the global ocean for the selected 169 MAGs across the BBMO, SOLA and TARA datasets.

MAG	Taxonomy	Seasonality BBMO	Seasonality SOLA	Biogeography TARA
bin.G1.136	o_Actinomarinales	annual	annual	subtropical
bin.G1.157	o_Actinomarinales	annual	annual	subtropical
bin.G1.197	o_Actinomarinales	annual	annual	subtropical/tropical
bin.G1.201	o_Actinomarinales	annual	annual	subtropical
bin.G1.203	o_Actinomarinales	not significant	not significant	subtropical
bin.G1.207	o_Actinomarinales	annual	annual	subtropical/tropical
bin.G1.211	o_Actinomarinales	annual	annual	subtropical/tropical
bin.G1.251	o_Balneolales	annual	annual	subtropical
bin.G1.297	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G1.304	o_Flavobacteriales	biannual	not significant	subtropical/tropical
bin.G1.306	o_Flavobacteriales	annual	annual	subtropical
bin.G1.308	o_Flavobacteriales	annual	annual	subtropical
bin.G1.314	o_Flavobacteriales	annual	annual	subtropical
bin.G1.317	o_Flavobacteriales	biannual	not significant	subtropical
bin.G1.319	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G1.332	o_Flavobacteriales	annual	no pattern	subtropical
bin.G1.334	o_Flavobacteriales	annual	annual	subtropical
bin.G1.35	o_Flavobacteriales	annual	annual	subtropical
bin.G1.378	o_Flavobacteriales	annual	annual	subtropical
bin.G1.392	o_Flavobacteriales	annual	annual	subtropical
bin.G1.396	o_Flavobacteriales	annual	annual	subtropical/subpolar
bin.G1.4	o_Flavobacteriales	annual	annual	subtropical
bin.G1.409	o_Flavobacteriales	biannual	annual	subtropical
bin.G1.415	o_Flavobacteriales	annual	annual	subtropical/subpolar
bin.G1.451	o_Flavobacteriales	annual	annual	subtropical

bin.G1.462	o_Flavobacteriales	not significant	annual	subtropical/subpolar
bin.G1.468	o_Flavobacteriales	annual	annual	subtropical
bin.G1.473	o_Flavobacteriales	annual	annual	subtropical
bin.G1.485	o_Flavobacteriales	not significant	not significant	subtropical/tropical
bin.G1.493	o_Flavobacteriales	not significant	not significant	subtropical
bin.G1.495	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G1.503	o_Flavobacteriales	biannual	annual	subtropical
bin.G1.528	o_Flavobacteriales	annual	not significant	subpolar
bin.G1.566	o_Flavobacteriales	biannual	annual	subtropical
bin.G1.590	o_Flavobacteriales	not significant	annual	subtropical/tropical
bin.G1.596	o_Flavobacteriales	annual	annual	subtropical/subpolar
bin.G1.614	o_Flavobacteriales	not significant	not significant	subtropical
bin.G1.627	o_Flavobacteriales	annual	annual	subpolar
bin.G1.632	o_Flavobacteriales	annual	annual	subtropical
bin.G1.647	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G1.651	o_Flavobacteriales	biannual	annual	subtropical
bin.G1.656	o_Flavobacteriales	annual	annual	subtropical
bin.G1.675	o_Flavobacteriales	annual	annual	subpolar
bin.G1.684	o_Flavobacteriales	annual	annual	subtropical
bin.G1.700	o_Flavobacteriales	annual	not significant	subtropical
bin.G1.719	o_Flavobacteriales	annual	annual	subtropical
bin.G1.746	o_Flavobacteriales	biannual	biannual	subtropical
bin.G1.748	o_Flavobacteriales	annual	annual	subtropical
bin.G1.767	o_Flavobacteriales	annual	annual	subtropical
bin.G1.769	o_Flavobacteriales	annual	annual	subtropical
bin.G1.774	o_Flavobacteriales	annual	annual	subtropical
bin.G1.799	o_Flavobacteriales	biannual	annual	subtropical
bin.G1.825	o_Flavobacteriales	annual	annual	subtropical/subpolar
bin.G1.830	o_Flavobacteriales	annual	not significant	subtropical/tropical
bin.G1.849	o_Flavobacteriales	annual	annual	subtropical
bin.G1.851	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G1.852	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G1.873	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G1.875	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G2.109	o_Flavobacteriales	not significant	not significant	subtropical/tropical
bin.G2.116	o_Flavobacteriales	annual	annual	subtropical
bin.G2.122	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G2.131	o_Flavobacteriales	annual	annual	subtropical/subpolar
bin.G2.135	o_Flavobacteriales	not significant	annual	subtropical/tropical
bin.G2.150	o_Flavobacteriales	annual	annual	subtropical
bin.G2.155	o_Flavobacteriales	biannual	not significant	subtropical
bin.G2.166	o_Flavobacteriales	annual	not significant	subtropical/tropical
bin.G2.171	o_Flavobacteriales	biannual	annual	subtropical
bin.G2.177	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G2.178	o_Flavobacteriales	not significant	not significant	subpolar/polar
bin.G2.181	o_Flavobacteriales	not significant	annual	subtropical/tropical
bin.G2.183	o_Flavobacteriales	not significant	not significant	subtropical
bin.G2.203	o_Flavobacteriales	annual	annual	subtropical
bin.G2.222	o_Flavobacteriales	annual	annual	subtropical/tropical
bin.G2.239	o_Flavobacteriales	biannual	annual	subtropical/tropical
bin.G2.252	o_Flavobacteriales	annual	annual	subtropical
bin.G2.263	o_Flavobacteriales	annual	annual	subtropical
bin.G2.267	o_Flavobacteriales	not significant	not significant	subtropical
bin.G2.270	o_Flavobacteriales	biannual	annual	subtropical
bin.G2.276	o_Flavobacteriales	biannual	not significant	subtropical/tropical
bin.G2.277	o_SAR11	no pattern	not significant	subtropical
bin.G2.285	o_SAR11	biannual	not significant	subtropical/subpolar
bin.G2.295	o_SAR11	annual	not significant	subtropical
bin.G2.315	o_SAR11	no pattern	annual	subtropical
bin.G2.318	o_SAR11	annual	annual	subtropical
bin.G2.32	o_SAR11	no pattern	no pattern	subtropical

bin.G2.321	o_SAR11	no pattern	no pattern	subtropical
bin.G2.325	o_SAR11	no pattern	annual	subtropical
bin.G2.327	o_SAR11	annual	annual	subtropical
bin.G2.329	o_SAR116	biannual	annual	subtropical
bin.G2.34	o_SAR116	annual	annual	tropical
bin.G2.39	o_SAR116	not significant	annual	subtropical
bin.G2.45	o_SAR116	biannual	biannual	subtropical/subpolar
bin.G2.80	o_SAR116	biannual	biannual	subtropical/subpolar
bin.G2.91	o_SAR116	annual	annual	subtropical
bin.G2.92	o_SAR116	annual	annual	subtropical
bin.G2.94	o_SAR116	annual	annual	subtropical/tropical
bin.G2.95	o_SAR116	biannual	annual	subtropical
bin.G2.97	o_SAR116	annual	annual	subtropical
bin.G3.101	o_SAR116	biannual	annual	subtropical/subpolar
bin.G3.116	o_SAR116	biannual	biannual	subtropical/subpolar
bin.G3.122	o_SAR116	annual	annual	subtropical
bin.G3.142	o_SAR116	annual	annual	subtropical
bin.G3.158	o_SAR116	annual	annual	subtropical
bin.G3.194	o_SAR116	annual	annual	subtropical/tropical
bin.G3.197	o_SAR116	annual	annual	subtropical
bin.G3.199	o_SAR324	annual	annual	subtropical
bin.G3.217	o_SAR324	annual	annual	subtropical
bin.G3.233	o_SAR324	annual	annual	subtropical
bin.G3.250	o_SAR324	annual	annual	subtropical
bin.G3.256	o_SAR86	annual	annual	subtropical
bin.G3.269	o_SAR86	annual	annual	subtropical/tropical
bin.G3.272	o_SAR86	annual	annual	subtropical
bin.G3.276	o_SAR86	annual	annual	subtropical
bin.G3.285	o_SAR86	annual	annual	subtropical
bin.G3.298	o_SAR86	annual	annual	subtropical/tropical
bin.G3.300	o_SAR86	annual	annual	subtropical
bin.G3.304	o_SAR86	annual	annual	subtropical
bin.G3.310	o_SAR86	annual	annual	subtropical
bin.G3.314	o_SAR86	annual	annual	subtropical
bin.G3.323	o_SAR86	annual	annual	subtropical
bin.G3.330	o_SAR86	annual	annual	subtropical/tropical
bin.G3.337	o_SAR86	annual	annual	subtropical
bin.G3.346	o_SAR86	annual	annual	subtropical/tropical
bin.G3.353	o_SAR86	annual	not significant	subtropical/tropical
bin.G3.354	o_SAR86	annual	annual	subtropical
bin.G3.360	o_SAR86	annual	annual	subtropical
bin.G3.387	o_SAR86	not significant	annual	subtropical
bin.G3.393	o_SAR86	annual	annual	subtropical
bin.G3.398	o_SAR86	annual	annual	subtropical
bin.G3.7	o_SAR86	annual	annual	subtropical
bin.G3.70	o_SAR86	annual	annual	subtropical
bin.G4.108	o_SAR86	annual	not significant	subtropical/tropical
bin.G4.120	o_SAR86	annual	annual	subtropical/subpolar
bin.G4.185	o_SAR86	biannual	not significant	subtropical
bin.G4.189	o_SAR86	annual	annual	subtropical
bin.G4.191	o_SAR86	annual	annual	subtropical
bin.G4.238	o_SAR86	not significant	annual	subtropical/tropical
bin.G4.252	o_SAR86	annual	annual	subtropical
bin.G4.274	o_SAR86	annual	annual	subtropical/tropical
bin.G4.311	o_SAR86	annual	annual	subtropical
bin.G4.315	o_SAR86	annual	annual	subtropical/tropical
bin.G4.342	o_SAR86	annual	annual	subtropical
bin.G4.347	o_SAR86	annual	annual	subtropical
bin.G4.384	o_SAR86	biannual	annual	subtropical/tropical
bin.G4.397	o_SAR86	not significant	not significant	subtropical
bin.G4.401	s_Archaea	annual	annual	subtropical

bin.G4.408	s_Archaea	biannual	annual	subtropical
bin.G4.413	s_Archaea	annual	annual	subtropical
bin.G4.423	s_Archaea	annual	annual	subtropical
bin.G4.43	s_Archaea	biannual	annual	subtropical
bin.G4.438	s_Archaea	annual	annual	subtropical
bin.G4.450	s_Archaea	biannual	annual	subtropical
bin.G4.471	s_Archaea	annual	annual	subtropical
bin.G4.472	s_Archaea	annual	annual	subtropical
bin.G4.480	s_Archaea	annual	annual	subtropical
bin.G4.503	s_Archaea	annual	annual	subtropical
bin.G4.513	s_Archaea	annual	annual	subtropical
bin.G4.551	s_Archaea	annual	annual	subtropical
bin.G4.557	s_Archaea	annual	annual	subtropical
bin.G4.583	s_Archaea	annual	annual	subtropical/tropical
bin.G4.586	s_Archaea	annual	annual	subtropical
bin.G4.590	s_Archaea	annual	annual	subtropical
bin.G4.62	s_Archaea	not significant	annual	subtropical
bin.G4.68	s_Archaea	annual	annual	subtropical/subpolar
bin.G4.77	s_Archaea	annual	annual	subtropical
bin.G4.8	s_Archaea	not significant	annual	subtropical
bin.G4.87	s_Archaea	annual	annual	subtropical
bin.G4.90	s_Archaea	not significant	annual	subtropical

ANNEX D – SUPPLEMENTARY MATERIAL FOR CHAPTER 4

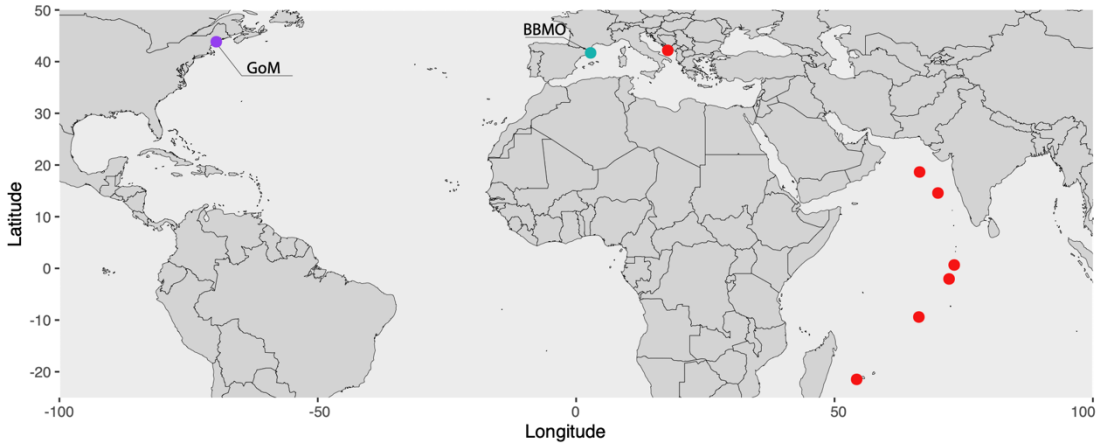

Figure 1. Locations of all the sample sites used to collect single-cell data. Dot color indicates the dataset: Blue – BBMO, Violet – GoM, Red – TARA.

Table 1. Samples context for LoCoS datasets for GoM and BBMO

Datset	Plate	Cells Used	Citation	Source doi	Organism Type
GoM	AAA071	285	Martinez-Garcia 2012	https://doi.org/10.1038/ismej.2011.126	a-plastidic eukaryote
GoM	AAA072	310	Brown 2020	https://doi.org/10.3389/fmicb.2020.524828	a-plastidic eukaryote
GoM	AG-605	317	Brown 2020	https://doi.org/10.3389/fmicb.2020.524828	plastidic eukaryote
BBMO	WA170123	378	Brown 2020	https://doi.org/10.3389/fmicb.2020.524828	plastidic eukaryote
BBMO	WA170125	378	this study	this study	plastidic eukaryote
BBMO	SH171117	382	Brown 2020	https://doi.org/10.3389/fmicb.2020.524828	a-plastidic eukaryote
BBMO	SHp170809	307	this study	this study	a-plastidic eukaryote
BBMO	WH180222	372	Brown 2020	https://doi.org/10.3389/fmicb.2020.524828	a-plastidic eukaryote

Table 2. Sample context for deep sequenced SAGs from TARA and BBMO. Abbreviations: GS - Genoscope, CNAG - Centro Nacional de Análisis Genómico, OR - Oregon Health & Science University. D - Deep, S - Surface; H - Heterotrophic, P – Phototrophic.

Table too large to fit. Data available on-line at:

https://doi.org/10.5281/zenodo.7078952

Table 3. Eukaryote - Prokaryote interactions found in WA170123 LoCoS SAGs

SAG Taxonomy	nº Cells	Prokaryotic taxonomy	nº of times found
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	5
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	4
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Nisaeales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	GCA-002705445;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Lactobacillales;Bacilli;Firmicutes;Bacteria	1

Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	SG8-23;Gemmatimonadetes;Gemmatimonadota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	15	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	4
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	3
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	Planctomycetales;Planctomycetes;Planctomycetota;Bacteria	2
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	TMED109;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	9	UBA817;ZB3;Margulisbacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	2	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	2	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	2	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Chloropicon;Chloropicaceae;Chloropicales;Chloropicophyceae;Chlorophyta;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Chloropicon;Chloropicaceae;Chloropicales;Chloropicophyceae;Chlorophyta;Eukaryota	1	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Chloropicon;Chloropicaceae;Chloropicales;Chloropicophyceae;Chlorophyta;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Chrysochromulina;Chrysochromulinaceae;Prymnesiales;Haptophyta;Haptista;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Cyanobacteriales;Cyanobacteriia;Cyanobacteria;Bacteria	1

Table 4. Eukaryote - Prokaryote interactions found in WA170125 LoCoS SAGs

SAG Taxonomy	nº Cells	Prokaryotic taxonomy	nº of times found
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	5
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	4
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	3
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Nisaeales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	2
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Bacteriovoracales;Bacteriovoracia;Bdellovibrionota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	GCA-002705445;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Lactobacillales;Bacilli;Firmicutes;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	UBA1151;Dehalococcoidia;Chloroflexota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	17	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1

Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	5
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Planctomycetales;Planctomycetes;Planctomycetota;Bacteria	2
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	GCA-002705445;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Nitrososphaerales;Nitrososphaeria;Crenarchaeota;Archaea	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	TMED109;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	11	UBA817;ZB3;Margulisbacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	2	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Chloropicon;Chloropicaceae;Chloropicales;Chloropicophyceae;Chlorophyta;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Chloropicon;Chloropicaceae;Chloropicales;Chloropicophyceae;Chlorophyta;Eukaryota	1	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Chloropicon;Chloropicaceae;Chloropicales;Chloropicophyceae;Chlorophyta;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Chrysochromulina;Chrysochromulinaceae;Prymnesiales;Haptophyta;Haptista;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	1	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Cyanobacteriales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1

Table 5. Eukaryote - Prokaryote interactions found in SH171117 LoCoS SAGs

SAG Taxonomy	nº Cells	Prokaryotic taxonomy	nº of times found
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	SAR324;SAR324;SAR324;Bacteria	4
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Ectothiorhodospirales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Methylococcales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	4	Thiotrichales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	3

Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	3
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	2
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Nisaeales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	TMED127;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	UBA1146;UBA8108;Planctomycetota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	UBA7879;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	3	UBA7916;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	2	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	2	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	2	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	2	UBA1280;Alphaproteobacteria;Proteobacteria;Bacteria	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	2
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Dinobryon;Dinobryaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Dinobryon;Dinobryaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Rickettsiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Dinobryon;Dinobryaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	UBA7916;Gammaproteobacteria;Proteobacteria;Bacteria	1
Lagenoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Lagenoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Phaeocystis;Phaeocystaceae;Phaeocystales;Haptophyta;Haptista;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Puniceispirillales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1

Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	UBA11654;Gammaproteobacteria;Proteobacteria;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Sphingomonadales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	SAR324;SAR324;SAR324;Bacteria	1

Table 6. Eukaryote - Prokaryote interactions found in SHp170809 LoCoS SAGs

SAG Taxonomy	nº Cells	Prokaryotic taxonomy	nº of times found
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	3
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	UBA1146;UBA8108;Planctomycetota;Bacteria	1
Telonema;unknown;Telonemida;unknown;unknown;Eukaryota	3	UBA228;Deferribacteres;Deferribacterota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	2
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	2
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	NS11-12g;Bacteroidia;Bacteroidota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	TMED127;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	UBA1146;UBA8108;Planctomycetota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	UBA228;Deferribacteres;Deferribacterota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	UBA7879;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	2	UBA7916;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	SAR324;SAR324;SAR324;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Cyanobacteriales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Ectothiorhodospirales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Leptospirales;Leptospirae;Spirochaetota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Methylococcales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1

Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	2	Puniceispirillales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Salpingoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	2	UBA11654;Gammaproteobacteria;Proteobacteria;Bacteria	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	UBA1146;UBA8108;Planctomycetota;Bacteria	1
Acanthoeca;Acanthoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	UBA228;Deferribacteres;Deferribacterota;Bacteria	1
Cafeteria;Cafeteriaceae;Bicosoecida;Bigyra;unknown;Eukaryota	1	Cyanobacteriales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Colpodella;Colpodellaceae;unknown;unknown;unknown;Eukaryota	1	Bacteroidales;Bacteroidia;Bacteroidota;Bacteria	1
Colpodella;Colpodellaceae;unknown;unknown;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Colpodella;Colpodellaceae;unknown;unknown;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	SAR324;SAR324;SAR324;Bacteria	1
Diaphanoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Microtrichales;Acidimicrobiia;Actinobacteriota;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Lagenoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Lagenoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Lagenoeca;Salpingoecidae;Craspedida;Choanoflagellata;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Oxyrrhis;Oxyrrhinaceae;Oxyrrhinales;Dinophyceae;unknown;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Oxyrrhis;Oxyrrhinaceae;Oxyrrhinales;Dinophyceae;unknown;Eukaryota	1	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Oxyrrhis;Oxyrrhinaceae;Oxyrrhinales;Dinophyceae;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Oxyrrhis;Oxyrrhinaceae;Oxyrrhinales;Dinophyceae;unknown;Eukaryota	1	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1

Oxyrrhis;Oxyrrhinaceae;Oxyrrhinales;Dinophyceae;unknown;Eukaryota	1	Streptomycetales;Actinobacteria;Actinobacteriota;Bacteria	1
Oxyrrhis;Oxyrrhinaceae;Oxyrrhinales;Dinophyceae;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Coxiellales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Rickettsiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	UBA7916;Gammaproteobacteria;Proteobacteria;Bacteria	1
Phaeocystis;Phaeocystaceae;Phaeocystales;Haptophyta;Haptista;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	1	UBA1280;Alphaproteobacteria;Proteobacteria;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Nevskiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Schmidingerella;Rhabdonellidae;Tintinnida;Spirotrichea;Ciliophora;Eukaryota	1	Sphingomonadales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	SAR324;SAR324;SAR324;Bacteria	1

Table 7. Eukaryote - Prokaryote interactions found in WH180222 LoCoS SAGs

SAG Taxonomy	n° Cells	Prokaryotic taxonomy	n° of times found
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	7
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	SAR324;SAR324;SAR324;Bacteria	6
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	5
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	4
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	4
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	WGA-4E;WGA-4E;Poribacteria;Bacteria	4
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Cyanobacteriales;Cyanobacteriia;Cyanobacteria;Bacteria	3
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	3
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	3
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Methylococcales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Puniceispirillales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Bacillales;Bacilli;Firmicutes;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Bacteroidales;Bacteroidia;Bacteroidota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Bradymonadales;Bradimonadia;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Brocadiales;Brocadiae;Planctomycetota;Bacteria	1

Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Caldilineales;Anaerolineae;Chloroflexota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Chromatiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Dehalococcoidales;Dehalococcoidia;Chloroflexota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Francisellales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Mycobacteriales;Actinobacteria;Actinobacteriota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Myxococcales;Myxococcia;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Nannocystales;Polyangia;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Polyangiales;Polyangia;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	Rhodospirillales_A;Alphaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	TMED109;Alphaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA10353;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA1135;UBA1135;Planctomycetota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA2968;UBA2968;Latescibacterota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA3071;Anaerolineae;Chloroflexota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA4151;UBA727;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA6777;UBA6777;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA796;UBA796;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA7976;Bradimonadia;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA8231;UBA2968;Latescibacterota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA9042;UBA9042;Myxococcota;Bacteria	1
Nemacystus;Chordariaceae;Ectocarpales;Phaeophyceae;unknown;Eukaryota	10	UBA9615;UBA796;Myxococcota;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	SAR324;SAR324;SAR324;Bacteria	4
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	2
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Caldilineales;Anaerolineae;Chloroflexota;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Chromatiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Ectothiorhodospirales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	Rhodospirillales_A;Alphaproteobacteria;Proteobacteria;Bacteria	1
Thalassiosira;Thalassiosiraceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	4	UBA3071;Anaerolineae;Chloroflexota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	2
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	SAR324;SAR324;SAR324;Bacteria	2
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1

Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Peptostreptococcales;Clostridia;Firmicutes_A;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Puniceispirillales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	Tissierellales;Clostridia;Firmicutes_A;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	UBA10353;Gammaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	UBA2968;UBA2968;Latescibacterota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	UBA3071;Anaerolineae;Chloroflexota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	UBA8231;UBA2968;Latescibacterota;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	3	WGA-4E;WGA-4E;Poribacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Planctomycetales;Planctomycetes;Planctomycetota;Bacteria	3
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	WGA-4E;WGA-4E;Poribacteria;Bacteria	3
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Bacteroidales;Bacteroidia;Bacteroidota;Bacteria	2
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Acetivibrionales;Clostridia;Firmicutes_A;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Acidobacteriales;Acidobacteriae;Acidobacteriota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Actinomycetales;Actinobacteria;Actinobacteriota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Bacillales;Bacilli;Firmicutes;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Bacteriovoracales;Bacteriovoracia;Bdellovibrionota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Balneolales;Rhodothermia;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Bdellovibrionales;Bdellovibrionia;Bdellovibrionota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Beggiatoales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Brocadiales;Brocadiae;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Bryobacterales;Acidobacteriae;Acidobacteriota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	C00003060;Desulfobacteria;Desulfobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Campylobacterales;Campylobacteria;Campylobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Chlorobiales;Chlorobia;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Chloroflexales;Chloroflexia;Chloroflexota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Chromatiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Chthoniobacterales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Cyanobacteriales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Desulfatiglandales;Desulfobacteria;Desulfobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Desulfobacterales;Desulfobacteria;Desulfobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Desulfobulbales;Desulfobulbia;Desulfobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Desulfotomaculales;Desulfotomaculia;Firmicutes_B;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Desulfovibrionales;Desulfovibrionia;Desulfobacterota_A;Bacteria	1

Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Desulfuromonadales;Desulfuromonadia;Desulfuromonadota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Dissulfuribacterales;Dissulfuribacteria;Desulfobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Entotheonellales;Entotheonellia;Tectomicrobia;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Fibrobacterales;Fibrobacteria;Fibrobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	GCA-2746535;UBA11346;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Halobacteroidales;Halanaerobiia;Firmicutes_F;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Ignavibacteriales;Ignavibacteria;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	JdFR-76;UBA2214;KSB1;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Kiloniellales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Kiritimatiellales;Kiritimatiellae;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Lachnospirales;Clostridia;Firmicutes_A;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Lactobacillales;Bacilli;Firmicutes;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Lentisphaerales;Lentisphaeria;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Leptospirales;Leptospirae;Spirochaetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Methanosarcinales;Methanosarcinia;Halobacterota;Archaea	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Methylococcales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	MHYJ01;Brocadiae;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Micavibrionales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Nitrospinales;Nitrospinia;Nitrospinota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Nitrospirales;Nitrospiria;Nitrospirota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Oligoflexales;Oligoflexia;Bdellovibrionota_B;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Omnitrophales;Omnitrophia;Omnitrophota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Oscillospirales;Clostridia;Firmicutes_A;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	OXYB2-FULL-49-7;OXYB2-FULL-49-7;Fibrobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Paenibacillales;Bacilli_A;Firmicutes_I;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Parachlamydiales;Chlamydiia;Verrucomicrobiota_A;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Peptostreptococcales;Clostridia;Firmicutes_A;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Rhodospirillales_A;Alphaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Rickettsiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	SAR324;SAR324;SAR324;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	SG8-4;Phycisphaerae;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Sphingobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Sphingomonadales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Spirochaetales;Spirochaetia;Spirochaetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	SS1-B-03-39;Kiritimatiellae;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Streptomycetales;Actinobacteria;Actinobacteriota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Syntrophales;Syntrophia;Desulfobacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Syntrophorhabdales;Syntrophorhabdia;Desulfobacterota;Bacteria	1

Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	SZUA-336;UBA9160;Myxococcota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	SZUA-567;SZUA-567;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Thermodesulfovibrionales;Thermodesulfovibrionia;Nitrospirota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Thiohalomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Thiomicrospirales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Tissierellales;Clostridia;Firmicutes_A;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA10015;koll11;Omnitrophota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA10030;UBA10030;Bacteroidota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA11346;UBA11346;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA1146;UBA8108;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA12247;Lentisphaeria;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA1407;Lentisphaeria;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA2565;Lentisphaeria;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA2968;UBA2968;Latescibacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA4802;UBA4802;Spirochaetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA6191;UBA6191;UBP17;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA6919;UBA6919;Spirochaetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA8108;UBA8108;Planctomycetota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA8231;UBA2968;Latescibacterota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA8416;Kiritimatiellae;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA9160;UBA9160;Myxococcota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	UBA9983_A;Paceibacteria;Patescibacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Vicinamibacterales;Vicinamibacteria;Acidobacteriota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Victivallales;Lentisphaeria;Verrucomicrobiota;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	Xanthomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Euglypha;Euglyphidae;Euglyphida;unknown;Imbricatea;Eukaryota	3	XYD2-FULL-50-16;SAR324;SAR324;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	2
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Puniceispirillales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	SAR324;SAR324;SAR324;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	UBA10117;Nanoarchacia;Nanoarchaeota;Archaea	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	UBA10353;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	UBA2968;UBA2968;Latescibacterota;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	UBA3071;Anaerolineae;Chloroflexota;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	UBA8231;UBA2968;Latescibacterota;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	3	WGA-4E;WGA-4E;Poribacteria;Bacteria	1

Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Chromatiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Planctomycetales;Planctomycetes;Planctomycetota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	1	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Aplanochytrium;Thraustochytriaceae;Thraustochytrida;Bigyra;unknown;Eukaryota	1	SG8-23;Gemmatimonadetes;Gemmatimonadota;Bacteria	1
Calanus;Calanidae;Calanoida;Hexanauplia;Arthropoda;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Chaetoceros;Chaetocerotaceae;Chaetocerotales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Chaetoceros;Chaetocerotaceae;Chaetocerotales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Chaetoceros;Chaetocerotaceae;Chaetocerotales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Chaetoceros;Chaetocerotaceae;Chaetocerotales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	UBA7976;Bradimonadia;Myxococcota;Bacteria	1
Chrysochromulina;Chrysochromulinaceae;Prymnesiales;Haptophyta;Haptista;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Paceibacterales;Paceibacteria;Patescibacteria;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	SAR324;SAR324;SAR324;Bacteria	1
Ciona;Cionidae;Phlebobranchia;Ascidiacea;Chordata;Eukaryota	1	UBA796;UBA796;Myxococcota;Bacteria	1
Fabomonas;Planomonadidae;Ancyromonadida;unknown;unknown;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Fabomonas;Planomonadidae;Ancyromonadida;unknown;unknown;Eukaryota	1	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1
Fabomonas;Planomonadidae;Ancyromonadida;unknown;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Fabomonas;Planomonadidae;Ancyromonadida;unknown;unknown;Eukaryota	1	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Fabomonas;Planomonadidae;Ancyromonadida;unknown;unknown;Eukaryota	1	Rhodospirillales_A;Alphaproteobacteria;Proteobacteria;Bacteria	1
Fabomonas;Planomonadidae;Ancyromonadida;unknown;unknown;Eukaryota	1	SAR324;SAR324;SAR324;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Bdellovibrionales;Bdellovibrionia;Bdellovibrionota;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1

Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Geminigera;Geminigeraceae;Pyrenomonadales;Cryptophyceae;unknown;Eukaryota	1	UBA7916;Gammaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	UBA9160;UBA9160;Myxococcota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	1	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Mikrocytos;Mikrocytiidae;unknown;Ascetosporea;Endomyxa;Eukaryota	1	Rickettsiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Bacillales;Bacilli;Firmicutes;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Bradymonadales;Bradimonadia;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	GCA-2863065;UBA9042;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Mycobacteriales;Actinobacteria;Actinobacteriota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Myxococcales;Myxococcia;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Nannocystales;Polyangia;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Palsa-1104;Polyangia;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Polyangiales;Polyangia;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	UBA1135;UBA1135;Planctomycetota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	UBA4151;UBA727;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	UBA6777;UBA6777;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	UBA796;UBA796;Myxococcota;Bacteria	1
Nephroselmis;Nephroselmidaceae;Nephroselmidales;Nephroselmidophyceae;Chlorophyta;Eukaryota	1	UBA9615;UBA796;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Bacillales;Bacilli;Firmicutes;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Bradymonadales;Bradimonadia;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Myxococcales;Myxococcia;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Nannocystales;Polyangia;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Polyangiales;Polyangia;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1

Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	UBA1135;UBA1135;Planctomycetota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	UBA4151;UBA727;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	UBA6777;UBA6777;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	UBA796;UBA796;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	UBA7976;Bradimonadia;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	UBA9042;UBA9042;Myxococcota;Bacteria	1
Parvularia;unknown;Rotosphaerida;unknown;unknown;Eukaryota	1	UBA9615;UBA796;Myxococcota;Bacteria	1
Pelagomonas;unknown;Pelagomonadales;Pelagophyceae;unknown;Eukaryota	1	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Pseudo-nitzschia;Bacillariaceae;Bacillariales;Bacillariophyceae;Bacillariophyta;Eukaryota	1	Cyanobacteriales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Symbiodinium;Symbiodiniaceae;Suessiales;Dinophyceae;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Symbiodinium;Symbiodiniaceae;Suessiales;Dinophyceae;unknown;Eukaryota	1	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1

Table 8. Eukaryote - Prokaryote interactions found in GoM LoCoS SAGs

SAG Taxonomy	nº Cells	Prokaryotic taxonomy	nº of times found
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	26
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	21
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	11
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	9
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	8
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	8
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	7
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	6
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	4
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	4
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	4
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Micavibrionales;Alphaproteobacteria;Proteobacteria;Bacteria	3
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	NS11-12g;Bacteroidia;Bacteroidota;Bacteria	3
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	3
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	TMED189;Acidimicrobiia;Actinobacteriota;Bacteria	3
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Coxiellales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	UBA7916;Gammaproteobacteria;Proteobacteria;Bacteria	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Acidiferrobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Francisellales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Paceibacterales;Paceibacteria;Patescibacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Pirellulales;Planctomycetes;Planctomycetota;Bacteria	1

Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	SCGC-AAA003-L08;Marinisomatia;Marinisomatota;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	Thiomicrospirales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	TMED109;Alphaproteobacteria;Proteobacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	UBA1144;UBA1144;Dadabacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	UBA11654;Gammaproteobacteria;Proteobacteria;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	UBA228;Deferribacteres;Deferribacterota;Bacteria	1
Ostreococcus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	54	UBA817;ZB3;Margulisbacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	8
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	6
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	4
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	4
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	2
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	UBA817;ZB3;Margulisbacteria;Bacteria	2
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Bacteroidales;Bacteroidia;Bacteroidota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Coxiellales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Methanobacteriales;Methanobacteria;Euryarchaeota;Archaea	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	SCGC-AAA011-G17;Nanoarchaeia;Nanoarchaeota;Archaea	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Thiomicrospirales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	UBA7916;Gammaproteobacteria;Proteobacteria;Bacteria	1
Micromonas;Mamiellaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	18	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	4
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	4
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	2
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	NS11-12g;Bacteroidia;Bacteroidota;Bacteria	1
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Amoebophrya;Amoebophryaceae;Syndiniales;Dinophyceae;unknown;Eukaryota	12	SCGC-AAA003-L08;Marinisomatia;Marinisomatota;Bacteria	1
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	7
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	4
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	3
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	3
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	2

unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	2
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	1
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	Rhizobiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
unknown;unknown;unknown;Chrysophyceae;unknown;Eukaryota	9	UBA817;ZB3;Margulisbacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	3
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	2
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	2
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Bathycoccus;Bathycoccaceae;Mamiellales;Mamiellophyceae;Chlorophyta;Eukaryota	8	TMED189;Acidimicrobiia;Actinobacteriota;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	3
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	2
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	NS11-12g;Bacteroidia;Bacteroidota;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Paraphysomonas;Paraphysomonadaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	5	TMED189;Acidimicrobiia;Actinobacteriota;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	3
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	AKYH767;Bacteroidia;Bacteroidota;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Bacteroidales;Bacteroidia;Bacteroidota;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Micavibrionales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Mataza;unknown;unknown;Thecofilosea;Cercozoa;Eukaryota	4	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	2
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	2	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	2	Opitutales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Abollifer;unknown;Marimonadida;unknown;Imbricatea;Eukaryota	2	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Aplanochytrium;Thraustochytriaceae;Thraustochytrida;Bigyra;unknown;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Aplanochytrium;Thraustochytriaceae;Thraustochytrida;Bigyra;unknown;Eukaryota	2	SAR86;Gammaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	2	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Collodictyon;Collodictyonidae;unknown;unknown;unknown;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1

Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	UBA1146;UBA8108;Planctomycetota;Bacteria	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	UBA228;Deferribacteres;Deferribacterota;Bacteria	2
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Enterobacterales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	NS11-12g;Bacteroidia;Bacteroidota;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Parvibaculales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Rhodobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	Synechococcales;Cyanobacteriia;Cyanobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	TMED127;Alphaproteobacteria;Proteobacteria;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	TMED189;Acidimicrobiia;Actinobacteriota;Bacteria	1
Didymoeca;Stephanoecidae;Acanthoecida;Choanoflagellata;unknown;Eukaryota	2	UBA1018;Bacteriovoracia;Bdellovibrionota;Bacteria	1
Lotharella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	2	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	2
Lotharella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	2	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Lotharella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	2	Marinisomatales;Marinisomatia;Marinisomatota;Bacteria	1
Lotharella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	2	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Lotharella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	2	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Lotharella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	2	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Arcella;Arcellidae;Arcellinida;Elardia;Tubulinea;Eukaryota	1	Caulobacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Arcella;Arcellidae;Arcellinida;Elardia;Tubulinea;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Bigelowiella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Bigelowiella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Bigelowiella;unknown;unknown;Chlorarachniophyceae;Cercozoa;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Filamoeba;unknown;unknown;Variosea;Evosea;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Filamoeba;unknown;unknown;Variosea;Evosea;Eukaryota	1	Pedosphaerales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Filamoeba;unknown;unknown;Variosea;Evosea;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Filamoeba;unknown;unknown;Variosea;Evosea;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Filamoeba;unknown;unknown;Variosea;Evosea;Eukaryota	1	TMED109;Alphaproteobacteria;Proteobacteria;Bacteria	1
Filamoeba;unknown;unknown;Variosea;Evosea;Eukaryota	1	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Filamoeba;unknown;unknown;Variosea;Evosea;Eukaryota	1	Verrucomicrobiales;Verrucomicrobiae;Verrucomicrobiota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	Bacteroidales;Bacteroidia;Bacteroidota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	Cytophagales;Bacteroidia;Bacteroidota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	Pacearchaeales;Nanoarchaeia;Nanoarchaeota;Archaea	1

Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	TMED127;Alphaproteobacteria;Proteobacteria;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	UBA10117;Nanoarchaeia;Nanoarchaeota;Archaea	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	UBA1146;UBA8108;Planctomycetota;Bacteria	1
Incisomonas;unknown;Nanomonadea;Bigyra;unknown;Eukaryota	1	UBA228;Deferribacteres;Deferribacterota;Bacteria	1
Paulinella;Paulinellidae;Euglyphida;unknown;Imbricatea;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pedospumella;Chromulinaceae;Chromulinales;Chrysophyceae;unknown;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Picochlorum;unknown;unknown;Trebouxiophyceae;Chlorophyta;Eukaryota	1	Chitinophagales;Bacteroidia;Bacteroidota;Bacteria	1
Picochlorum;unknown;unknown;Trebouxiophyceae;Chlorophyta;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Picochlorum;unknown;unknown;Trebouxiophyceae;Chlorophyta;Eukaryota	1	Phycisphaerales;Phycisphaerae;Planctomycetota;Bacteria	1
Picochlorum;unknown;unknown;Trebouxiophyceae;Chlorophyta;Eukaryota	1	Pseudomonadales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Prymnesium;Prymnesiaceae;Prymnesiales;Haptophyta;Haptista;Eukaryota	1	Burkholderiales;Gammaproteobacteria;Proteobacteria;Bacteria	1
Pythium;Pythiaceae;Pythiales;unknown;Oomycota;Eukaryota	1	Pelagibacterales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Skeletonema;Skeletonemataceae;Thalassiosirales;Coscinodiscophyceae;Bacillariophyta;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
Strombidium;Strombidiidae;unknown;Spirotrichea;Ciliophora;Eukaryota	1	Micavibrionales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Strombidium;Strombidiidae;unknown;Spirotrichea;Ciliophora;Eukaryota	1	Rickettsiales;Alphaproteobacteria;Proteobacteria;Bacteria	1
Strombidium;Strombidiidae;unknown;Spirotrichea;Ciliophora;Eukaryota	1	TMED189;Acidimicrobiia;Actinobacteriota;Bacteria	1
Triparma;Triparmaceae;Parmales;Bolidophyceae;unknown;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1
unknown;unknown;unknown;unknown;Picozoa;Eukaryota	1	Flavobacteriales;Bacteroidia;Bacteroidota;Bacteria	1
unknown;unknown;unknown;unknown;Picozoa;Eukaryota	1	Poseidoniales;Poseidoniia;Thermoplasmatota;Archaea	1

Table 9. Eukaryote - Prokaryote and Eukaryote - Eukaryote interactions found in TARA deep sequenced SAGs.

Table too large to fit. Available on-line at: https://doi.org/10.5281/zenodo.7078952

Table 10. Eukaryote - Prokaryote and Eukaryote - Eukaryote interactions found in BBMO deep sequenced SAGs.

Table too large to fit. Available on-line at: https://doi.org/10.5281/zenodo.7078952

Table 11. Taxonomy of microbes represented by network nodes in Figure 4.1 – 4.3.

Node_ID	Superkingdom	Phylum	Class	Order
1	Archaea	Asgardarchaeota	Lokiarchaeia	Thorarchaeales
2	Archaea	Crenarchaeota	Nitrososphaeria	Nitrososphaerales
3	Archaea	Crenarchaeota	Thermoprotei	Sulfolobales
4	Archaea	Crenarchaeota	Thermoprotei	Desulfurococcales
5	Archaea	Euryarchaeota	Methanobacteria	Methanobacteriales
6	Archaea	Halobacterota	Archaeoglobi	Archaeoglobales
7	Archaea	Halobacterota	Halobacteria	Halobacteriales
8	Archaea	Halobacterota	Methanomicrobia	Methanomicrobiales
9	Archaea	Halobacterota	Methanosarcinia	Methanosarcinales
10	Archaea	Nanoarchaeota	Nanoarchaeia	Pacearchaeales
11	Archaea	Nanoarchaeota	Nanoarchaeia	UBA10117
12	Archaea	Nanoarchaeota	Nanoarchaeia	Woesearchaeales
13	Archaea	Thermoplasmatota	Poseidoniia	Poseidoniales
14	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales
15	Bacteria	Acidobacteriota	Acidobacteriae	Bryobacterales
16	Bacteria	Acidobacteriota	Acidobacteriae	UBA7541
17	Bacteria	Acidobacteriota	Aminicenantia	Aminicenantales
18	Bacteria	Acidobacteriota	Blastocatellia	Pyrinomonadales
19	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales
20	Bacteria	Actinobacteriota	Acidimicrobiia	TMED189
21	Bacteria	Actinobacteriota	Acidimicrobiia	Microtrichales

22	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales
23	Bacteria	Actinobacteriota	Actinobacteria	Streptomycetales
24	Bacteria	Actinobacteriota	Actinobacteria	Mycobacteriales
25	Bacteria	Actinobacteriota	Actinobacteria	Streptosporangiales
26	Bacteria	Actinobacteriota	Actinobacteria	Propionibacteriales
27	Bacteria	Actinobacteriota	Thermoleophilia	Solirubrobacterales
28	Bacteria	Aquificota	Aquificae	Aquificales
29	Bacteria	Armatimonadota	UBA5377	UBA5377
30	Bacteria	Bacteroidota	Bacteroidia	Flavobacteriales
31	Bacteria	Bacteroidota	Bacteroidia	Chitinophagales
32	Bacteria	Bacteroidota	Bacteroidia	Cytophagales
33	Bacteria	Bacteroidota	Bacteroidia	NS11-12g
34	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales
35	Bacteria	Bacteroidota	Bacteroidia	Sphingobacteriales
36	Bacteria	Bacteroidota	Chlorobia	Chlorobiales
37	Bacteria	Bacteroidota	Ignavibacteria	Ignavibacteriales
38	Bacteria	Bacteroidota	Kapabacteria	Kapabacteriales
39	Bacteria	Bacteroidota	Rhodothermia	Rhodothermales
40	Bacteria	Bacteroidota	Rhodothermia	Balneolales
41	Bacteria	Bacteroidota	UBA10030	UBA10030
42	Bacteria	Bdellovibrionota	Bacteriovoracia	Bacteriovoracales
43	Bacteria	Bdellovibrionota	Bacteriovoracia	UBA1018
44	Bacteria	Bdellovibrionota	Bdellovibrionia	Bdellovibrionales
45	Bacteria	Bdellovibrionota_B	Oligoflexia	Oligoflexales
46	Bacteria	Caldatribacteriota	JS1	SB-45
47	Bacteria	Campylobacterota	Campylobacteria	Campylobacterales
48	Bacteria	Chloroflexota	Anaerolineae	Caldilineales
49	Bacteria	Chloroflexota	Anaerolineae	Promineofilales
50	Bacteria	Chloroflexota	Anaerolineae	Anaerolineales
51	Bacteria	Chloroflexota	Anaerolineae	UBA1429
52	Bacteria	Chloroflexota	Chloroflexia	Chloroflexales
53	Bacteria	Chloroflexota	Dehalococcoidia	UBA2979
54	Bacteria	Chloroflexota	UBA5177	UBA5177
55	Bacteria	Cyanobacteria	Cyanobacteriia	Synechococcales
56	Bacteria	Cyanobacteria	Cyanobacteriia	Cyanobacteriales
57	Bacteria	Cyanobacteria	Cyanobacteriia	Phormidesmiales
58	Bacteria	Cyanobacteria	Cyanobacteriia	Leptolyngbyales
59	Bacteria	Cyanobacteria	Cyanobacteriia	Pseudanabaenales
60	Bacteria	Cyanobacteria	Cyanobacteriia	Thermosynechococcales
61	Bacteria	Dadabacteria	UBA1144	UBA2774
62	Bacteria	Deferribacterota	Deferribacteres	UBA228
63	Bacteria	Deferrisomatota	Defferisomatia	Defferisomatales
64	Bacteria	Deinococcota	Deinococci	Deinococcales
65	Bacteria	Dependentiae	Babeliae	Babeliales
66	Bacteria	Desulfobacterota	Syntrophia	Syntrophales
67	Bacteria	Desulfobacterota	Syntrophobacteria	Syntrophobacterales
68	Bacteria	Desulfobacterota_A	Desulfovibrionia	Desulfovibrionales
69	Bacteria	Elusimicrobiota	Elusimicrobia	Elusimicrobiales
70	Bacteria	Eremiobacterota	UBP9	UBA4705
71	Bacteria	Fibrobacterota	Fibrobacteria	Fibrobacterales
72	Bacteria	Firmicutes	Bacilli	Bacillales
73	Bacteria	Firmicutes	Bacilli	Lactobacillales
74	Bacteria	Firmicutes	Bacilli	Staphylococcales
75	Bacteria	Firmicutes	Bacilli	RF39
76	Bacteria	Firmicutes_A	Clostridia	Oscillospirales
77	Bacteria	Firmicutes_A	Clostridia	Tissierellales
78	Bacteria	Firmicutes_A	Clostridia	Clostridiales
79	Bacteria	Firmicutes_A	Clostridia	Lachnospirales
80	Bacteria	Firmicutes_A	Clostridia	Peptostreptococcales
81	Bacteria	Firmicutes_A	Thermoanaerobacteria	Caldanaerobiales
82	Bacteria	Firmicutes_C	Negativicutes	Selenomonadales
83	Bacteria	Firmicutes_I	Bacilli_A	Paenibacillales
84	Bacteria	Firmicutes_I	Bacilli_A	Thermoactinomycetales
85	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales
86	Bacteria	Latescibacterota	UBA2968	UBA8231
87	Bacteria	Latescibacterota	UBA2968	UBA2968
88	Bacteria	Margulisbacteria	ZB3	UBA817
89	Bacteria	Marinisomatota	Marinisomatia	Marinisomatales
90	Bacteria	Marinisomatota	Marinisomatia	SCGC-AAA003-L08
91	Bacteria	Methylomirabilota	Methylomirabilia	Rokubacteriales
92	Bacteria	Myxococcota	Myxococcia	Myxococcales
93	Bacteria	Myxococcota	Polyangia	Nannocystales
94	Bacteria	Myxococcota	Polyangia	Palsa-1104
95	Bacteria	Myxococcota	Polyangia	Polyangiales
96	Bacteria	Myxococcota	Polyangia	Haliangiales
97	Bacteria	Myxococcota	UBA6777	UBA6777

98	Bacteria	Myxococcota	UBA727	UBA727
99	Bacteria	Myxococcota	UBA796	UBA796
100	Bacteria	Myxococcota	UBA9042	PHBI01
101	Bacteria	Myxococcota	UBA9160	UBA9160
102	Bacteria	Omnitrophota	koll11	GIF10
103	Bacteria	Patescibacteria	ABY1	BM507
104	Bacteria	Patescibacteria	ABY1	UBA10025
105	Bacteria	Patescibacteria	Gracilibacteria	UBA1369
106	Bacteria	Patescibacteria	Microgenomatia	UBA1406
107	Bacteria	Patescibacteria	Paceibacteria	UBA9983_A
108	Bacteria	Patescibacteria	Paceibacteria	Paceibacterales
109	Bacteria	Planctomycetota	Brocadiae	DG-23
110	Bacteria	Planctomycetota	GCA-002687715	GCA-002687715
111	Bacteria	Planctomycetota	Phycisphaerae	Phycisphaerales
112	Bacteria	Planctomycetota	Phycisphaerae	SG8-4
113	Bacteria	Planctomycetota	Phycisphaerae	UBA1845
114	Bacteria	Planctomycetota	Planctomycetes	Planctomycetales
115	Bacteria	Planctomycetota	Planctomycetes	Isosphaerales
116	Bacteria	Planctomycetota	Planctomycetes	Pirellulales
117	Bacteria	Planctomycetota	Planctomycetes	Gemmatales
118	Bacteria	Planctomycetota	UBA1135	UBA1135
119	Bacteria	Planctomycetota	UBA1135	UBA2386
120	Bacteria	Planctomycetota	UBA8108	UBA1146
121	Bacteria	Poribacteria	WGA-4E	WGA-4E
122	Bacteria	Proteobacteria	Alphaproteobacteria	Caulobacterales
123	Bacteria	Proteobacteria	Alphaproteobacteria	Pelagibacterales
124	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales
125	Bacteria	Proteobacteria	Alphaproteobacteria	Micavibrionales
126	Bacteria	Proteobacteria	Alphaproteobacteria	Parvibaculales
127	Bacteria	Proteobacteria	Alphaproteobacteria	Nisaeales
128	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales
129	Bacteria	Proteobacteria	Alphaproteobacteria	Puniceispirillales
130	Bacteria	Proteobacteria	Alphaproteobacteria	Acetobacterales
131	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales_A
132	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales_C
133	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales
134	Bacteria	Proteobacteria	Alphaproteobacteria	UBA7985
135	Bacteria	Proteobacteria	Alphaproteobacteria	TMED109
136	Bacteria	Proteobacteria	Alphaproteobacteria	Caedimonadales
137	Bacteria	Proteobacteria	Alphaproteobacteria	HIMB59
138	Bacteria	Proteobacteria	Alphaproteobacteria	UBA998
139	Bacteria	Proteobacteria	Alphaproteobacteria	UBA1280
140	Bacteria	Proteobacteria	Alphaproteobacteria	TMED127
141	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales
142	Bacteria	Proteobacteria	Gammaproteobacteria	SAR86
143	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales
144	Bacteria	Proteobacteria	Gammaproteobacteria	Coxiellales
145	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales
146	Bacteria	Proteobacteria	Gammaproteobacteria	UBA7916
147	Bacteria	Proteobacteria	Gammaproteobacteria	Methylococcales
148	Bacteria	Proteobacteria	Gammaproteobacteria	Legionellales
149	Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales
150	Bacteria	Proteobacteria	Gammaproteobacteria	Granulosicoccales
151	Bacteria	Proteobacteria	Gammaproteobacteria	UBA9339
152	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales
153	Bacteria	Proteobacteria	Gammaproteobacteria	Beggiatoales
154	Bacteria	Proteobacteria	Gammaproteobacteria	Piscirickettsiales
155	Bacteria	Proteobacteria	Gammaproteobacteria	Thiomicrospirales
156	Bacteria	Proteobacteria	Gammaproteobacteria	UBA1113
157	Bacteria	Proteobacteria	Gammaproteobacteria	Berkiellales
158	Bacteria	Proteobacteria	Gammaproteobacteria	Diplorickettsiales
159	Bacteria	Proteobacteria	Gammaproteobacteria	Francisellales
160	Bacteria	Proteobacteria	Gammaproteobacteria	Nevskiales
161	Bacteria	Proteobacteria	Gammaproteobacteria	Steroidobacterales
162	Bacteria	Proteobacteria	Gammaproteobacteria	Thiohalobacterales
163	Bacteria	Proteobacteria	Gammaproteobacteria	Thiohalomonadales
164	Bacteria	Proteobacteria	Gammaproteobacteria	UBA5158
165	Bacteria	Proteobacteria	Gammaproteobacteria	UBA11654
166	Bacteria	Proteobacteria	Magnetococcia	Magnetococcales
167	Bacteria	Proteobacteria	Zetaproteobacteria	Mariprofundales
168	Bacteria	SAR324	SAR324	SAR324
169	Bacteria	Spirochaetota	Brachyspirae	Brachyspirales
170	Bacteria	Spirochaetota	Leptospirae	Turneriellales
171	Bacteria	Spirochaetota	Leptospirae	Leptospirales
172	Bacteria	Spirochaetota	Spirochaetia	Spirochaetales
173	Bacteria	Spirochaetota	Spirochaetia	Treponematales

174	Bacteria	Spirochaetota	UBA6919	UBA6919
175	Bacteria	Tectomicrobia	Entotheonellia	Entotheonellales
176	Bacteria	Thermotogota	Thermotogae	Thermotogales
177	Bacteria	Verrucomicrobiota	Kiritimatiellae	SS1-B-03-39
178	Bacteria	Verrucomicrobiota	Kiritimatiellae	UBA8416
179	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Opitutales
180	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Verrucomicrobiales
181	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Pedosphaerales
182	Bacteria	Verrucomicrobiota	Verrucomicrobiae	Chthoniobacterales
183	Bacteria	Verrucomicrobiota_A	Chlamydiia	Parachlamydiales
184	Eukaryota	Apicomplexa	Conoidasida	Eugregarinorida
185	Eukaryota	Arthropoda	Hexanauplia	Calanoida
186	Eukaryota	Bacillariophyta	Bacillariophyceae	Bacillariales
187	Eukaryota	Bacillariophyta	Coscinodiscophyceae	Thalassiosirales
188	Eukaryota	Bacillariophyta	Coscinodiscophyceae	Chaetocerotales
189	Eukaryota	Bacillariophyta	Mediophyceae	Cymatosirales
190	Eukaryota	Cercozoa	Chlorarachniophyceae	unknown
191	Eukaryota	Cercozoa	Thecofilosea	unknown
192	Eukaryota	Chlorophyta	Chlorodendrophyceae	Chlorodendrales
193	Eukaryota	Chlorophyta	Chlorophyceae	Chlamydomonadales
194	Eukaryota	Chlorophyta	Chloropicophyceae	Chloropicales
195	Eukaryota	Chlorophyta	Mamiellophyceae	Dolichomastigales
196	Eukaryota	Chlorophyta	Mamiellophyceae	Mamiellales
197	Eukaryota	Chlorophyta	Nephroselmidophyceae	Nephroselmidales
198	Eukaryota	Chlorophyta	Pyramimonadophyceae	Pyramimonadales
199	Eukaryota	Ciliophora	Litostomatea	Cyclotrichida
200	Eukaryota	Euglenozoa	unknown	Diplonemea
201	Eukaryota	Evosea	Eumycetozoa	Physariida
202	Eukaryota	Foraminifera	unknown	Rotaliida
203	Eukaryota	Haptista	Centroplasthelida	Pterocystida
204	Eukaryota	Haptista	Haptophyta	Prymnesiales
205	Eukaryota	Haptista	Haptophyta	Coccolithales
206	Eukaryota	Haptista	Haptophyta	Isochrysidales
207	Eukaryota	Haptista	Haptophyta	Pavlovales
208	Eukaryota	Haptista	Haptophyta	Phaeocystales
209	Eukaryota	Haptista	Haptophyta	Coccosphaerales
210	Eukaryota	Imbricatea	unknown	Euglyphida
211	Eukaryota	Imbricatea	unknown	Marimonadida
212	Eukaryota	Oomycota	unknown	Pythiales
213	Eukaryota	Oomycota	unknown	Peronosporales
214	Eukaryota	Oomycota	unknown	Saprolegniales
215	Eukaryota	Prasinodermophyta	Prasinodermophyceae	Prasinodermales
216	Eukaryota	Streptophyta	Ginkgoopsida	Ginkgoales
217	Eukaryota	unknown	Bigyra	Bicosoecida
218	Eukaryota	unknown	Bigyra	Nanomonadea
219	Eukaryota	unknown	Bigyra	Thraustochytrida
220	Eukaryota	unknown	Bolidophyceae	Parmales
221	Eukaryota	unknown	Choanoflagellata	Craspedida
222	Eukaryota	unknown	Choanoflagellata	Acanthoecida
223	Eukaryota	unknown	Chrysophyceae	Chromulinales
224	Eukaryota	unknown	Chrysophyceae	unknown
225	Eukaryota	unknown	Cryptophyceae	Cryptomonadales
226	Eukaryota	unknown	Cryptophyceae	Cyathomonadacea
227	Eukaryota	unknown	Cryptophyceae	Pyrenomonadales
228	Eukaryota	unknown	Dictyochophyceae	Florenciellales
229	Eukaryota	unknown	Dictyochophyceae	Pedinellales
230	Eukaryota	unknown	Dictyochophyceae	Dictyochales
231	Eukaryota	unknown	Dinophyceae	Gonyaulacales
232	Eukaryota	unknown	Dinophyceae	Prorocentrales
233	Eukaryota	unknown	Dinophyceae	Dinophysiales
234	Eukaryota	unknown	Dinophyceae	Gymnodiniales
235	Eukaryota	unknown	Dinophyceae	Peridiniales
236	Eukaryota	unknown	Dinophyceae	Suessiales
237	Eukaryota	unknown	Dinophyceae	Syndiniales
238	Eukaryota	unknown	Dinophyceae	Thoracosphaerales
239	Eukaryota	unknown	Glaucocystophyceae	unknown
240	Eukaryota	unknown	Ichthyosporea	Ichthyophonida
241	Eukaryota	unknown	Pelagophyceae	Pelagomonadales
242	Eukaryota	unknown	Pelagophyceae	Sarcinochrysidales
243	Eukaryota	unknown	Pelagophyceae	unknown
244	Eukaryota	unknown	Phaeophyceae	Ectocarpales
245	Eukaryota	unknown	Pinguiophyceae	Pinguiochrysidales
246	Eukaryota	unknown	Raphidophyceae	Chattonellales
247	Eukaryota	unknown	Rhodelphea	Rhodelphida
248	Eukaryota	unknown	Synchromophyceae	unknown
249	Eukaryota	unknown	Synurophyceae	Ochromonadales

250	Eukaryota	unknown	Synurophyceae	Synurales
251	Eukaryota	unknown	unknown	Ancyromonadida
252	Eukaryota	unknown	unknown	Telonemida
253	Eukaryota	unknown	unknown	Rotosphaerida
254	Eukaryota	Bacillariophyta	Coscinodiscophyceae	Thalassiosirales
255	NA	Bathycoccus prasinos	NA	NA
256	Eukaryota	Cercozoa	Chlorarachniophyceae	unknown
257	Eukaryota	Cercozoa	Thecofilosea	unknown
258	NA	Chlorarachniophyta-sp1	NA	NA
259	Eukaryota	Chlorophyta	Mamiellophyceae	Mamiellales
260	Eukaryota	Chlorophyta	Mamiellophyceae	Mamiellales
261	Eukaryota	Chlorophyta	Mamiellophyceae	Mamiellales
262	NA	ChrysophyceaeG-sp2	NA	NA
263	NA	Chrysophyte-G	NA	NA
264	NA	Chrysophyte-H	NA	NA
265	NA	Dictyochophyceae_SAG	NA	NA
266	Eukaryota	Imbricatea	unknown	Marimonadida
267	Eukaryota	Imbricatea	unknown	Euglyphida
268	NA	MAST-1C-sp1	NA	NA
269	NA	MAST-1D-sp2	NA	NA
270	NA	MAST-11	NA	NA
271	NA	MAST-1D	NA	NA
272	NA	MAST-3C-sp1	NA	NA
273	NA	MAST-3C-sp2	NA	NA
274	NA	MAST-3A	NA	NA
275	NA	MAST-3F	NA	NA
276	NA	MAST-4A-sp1	NA	NA
277	NA	MAST-4A	NA	NA
278	NA	MAST-4B	NA	NA
279	NA	MAST-4C	NA	NA
280	NA	MAST-4E	NA	NA
281	NA	MAST-7	NA	NA
282	NA	MAST-8B-sp1	NA	NA
283	NA	MAST-9	NA	NA
284	NA	Micromonas-sp1	NA	NA
285	Eukaryota	Oomycota	unknown	Pythiales
286	NA	Pelagomonas Calceolata	NA	NA
287	NA	Picozoa-sp1	NA	NA
288	NA	Prymnesiophyceae-sp1	NA	NA
289	Eukaryota	unknown	Dinophyceae	Syndiniales
290	Eukaryota	unknown	Choanoflagellata	Acanthoecida
291	Eukaryota	unknown	Chrysophyceae	Chromulinales
292	Eukaryota	unknown	Chrysophyceae	unknown
293	Eukaryota	unknown	Bigyra	Nanomonadea
294	Eukaryota	unknown	Phaeophyceae	Ectocarpales
295	Eukaryota	unknown	unknown	Telonemida
296	Eukaryota	unknown	Pelagophyceae	Pelagomonadales
297	Eukaryota	unknown	unknown	unknown