Exploring B vitamin biosynthesis and exchange in understudied marine stramenopiles and their communities

Dominic Edward Absolon

Department of Plant Sciences

University of Cambridge

Darwin College

September 2022

This thesis is submitted for the degree of Doctor of Philosophy

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

The work in this thesis is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

This thesis does not exceed the prescribed word limit set by the Biology Degree Committee.

Abstract - Exploring B vitamin biosynthesis and exchange in understudied marine stramenopiles and their communities

Dominic Edward Absolon

Microbes are the dominant form of life on Earth. Environmental sequencing has revealed the extent of the diversity of microbial life in all environments, not least the ocean. In this vast aquatic environment microbes, both prokaryotic and eukaryotic, play essential roles in the cycling of nutrients, primary productivity and form the basis of all food webs. These single-celled organisms do not live in isolation, rather they form complex communities, a dominant feature of which is the exchange of nutrients. In many areas of the ocean the concentration of some macro and micronutrients are limiting to the growth of some organisms. In addition, many organisms appear to have dispensed with the ability to biosynthesise various micronutrients, for example, B vitamins. Instead, they rely on an external source to satisfy their requirement. Here I explore the role of B vitamin exchange in microbial communities, with a particular focus on the marine protists of the stramenopile group of eukaryotes.

An initial investigation took environmental sea water samples and applied different B vitamin and macronutrient amendments to assess if any of these compounds were limiting for the growth of the microbial community as a whole in this area of the ocean (English Channel), or if growth was co-limited by both B vitamin and macronutrient concentration. The results suggested that none of the B vitamins tested were a limiting factor for growth of the majority of species, either prokaryotic or eukaryotic. However, some changes to the microbial assemblage were observed when assessing the relative transcriptional activities of the organisms in the samples. For example, the bacterial genus *Litorivivens* was more transcriptionally active (high ratio of ribosomal RNA reads to ribosomal DNA reads) under the addition of vitamins B_2, B_3, B_5, B_6 and B_9. The eukaryotic genera *Ostreococcus* and *Picochlorum* were also more active in the same condition.

The stramenopiles display considerable variety in morphology and lifestyle, from parasitic soil-based oomycetes to photosynthetic planktonic species such as the diatoms. A comparative genomics approach was employed to investigate the B vitamin biosynthetic capabilities of this diverse group and therefore deduce any likely auxotrophies for these nutrients. Identifying auxotrophic species hints at potential requirements for nutrient exchange with an external partner.

The results of this study showed that the metabolism of stramenopile species is as diverse as the physical attributes of the members of the group. Of the species analysed in this experiment 64% were predicted to be auxotrophic for B_{12} and this trait was spread across the group in a mosaic fashion. For the vitamin B_3 the analysis shows a clear dichotomy in the utilisation of one of two alternative pathway branches. This indicates multiple gene transfer events have occurred across the group to allow for this change in metabolic routes. Biosynthesis for some of the B vitamins, namely B_5 and B_6 appears to be a universal trait in the stramenopiles. The results also demonstrate that B vitamin auxotrophy, in general, is more common in the heterotrophic species of the group than in the phototrophs.

Finally, metagenomics was employed to probe the community surrounding an enigmatic marine stramenopile, *Incisomonas marina,* one of only a few MAST (MArine STramenopile) species that have been cultured. MASTs represent the hitherto understudied and unknown heterotrophic cohort of the stramenopiles. This is in contrast to the comparatively well studied and understood ochrophytes, the photosynthetic members of the stramenopiles, who have thousands of representatives in culture. *I. marina* was originally isolated from Nova Scotia, Canada with a cohort of bacterial species, with which it has been maintained in culture. This provides a closed community to interrogate for nutrient exchange and interaction. Analysis of pathways for B vitamin and amino acid biosynthesis revealed that no one single species of the one eukaryote and 23 bacterial species in the community encodes a complete biosynthesis pathway for the full suite of either the B vitamins or the amino acids. This suggests some level of exchange for these essential nutrients indicating a community-based approach to metabolism.

Overall, the work outlined in this thesis suggests that B vitamins play an essential role in shaping the communities within which stramenopiles reside. Given the ecological prevalence of this group, these types of interactions between members of the community to share nutrients and exchange metabolites are essential to the health and function of the environments in which they live. Furthering our understanding of these processes will aid our understanding of microbial communities as a whole and the biogeochemical processes that they dictate.

Acknowledgments

Completing a PhD is a significant undertaking. For myself, it would not have been possible to complete without the support of the following people, to whom I'd now like to formally express my appreciation and gratitude.

First and foremost, I would like to thank Professor Alison Smith and Dr Katherine Helliwell for supervising my work so attentively. In what was a journey of discovery for all three of us, they provided a rudder of experience whilst allowing me to explore the avenues that interested me the most. I am honoured to have passed through the Smith lab and I wish Alison a very enjoyable retirement, she deserves it. I am also pleased to be one of Katherines very first PhD students and I wish her all the best of luck as she becomes a mother for the first time and continues to establish her already impressive academic career. I also thank both for securing the funding for my PhD which was provided by the Natural Earth Research Council (NERC) to which I also express my thanks for making my project a possibility.

Included as part of my project was a CASE partnership with the Marine Biological Association (MBA). I would like to thank everyone at the MBA for being so welcoming when I visited in the summer of 2019. I would like to thank the Cunliffe group especially who provided such good support and advice on the set up of my environmental experiments. In particular I would like to thank Ro Allen. Despite leaving the MBA shortly after my visit in 2019, Ro freely offered his assistance and support for the analysis of my amplicon sequencing data for which there was no other support available in Cambridge. Ro took a significant amount of his own time to help me for no other reason than kindness. For that I am extremely grateful.

Adam Monier and Vicky Jackson, connections that were made by Katherine through her lectureship at the University of Exeter, were kind enough to provide sequencing runs for me which allowed me to pursue the completion of the *Incisomonas marina* genome (Chapter 4). I thank them for their time and advice in this department.

Richard Dorrell provided a fascinating introduction to the world of protists and the use of bioinformatics and sequencing data to understand evolution. An academic visit to Richard in Paris in 2019 really set me up for success going forward in my PhD and for that I am thankful to him.

A significant proportion of my thesis is built on the work performed by Professor Tom Cavalier-Smith, including the isolation and classification of *Insicomonas marina* (Chapter 4). Professor

Cavalier-Smith, who sadly passed away in 2021, played a significant role in progressing the science of protistology, championing the understanding of the wider tree of life as well as establishing many taxonomic groups that are mentioned throughout this thesis. I thank him for providing the framework and basis on which I could build this project.

I was extremely fortunate to undertake this project in such a welcoming and friendly research group. Officially the Plant Metabolism group but know as Lab 220. I have fond memories with every member that has overlapped with my time. I thank all of them for these positive interactions. In particular I'd like to acknowledge the lab postdocs during my time: Payam, Pawel, Gonzalo and Katrin. Also, Lorraine for being the best sound board, emotional vent, source of entertainment and comedy gold. The lab techs during my time, Monika, Kitty and Katie I thank for providing entertainment as well as fielding my inane questions on the rare occasions I found myself doing wet lab work.

I was lucky enough to spend time with a fantastic group of students whilst in Lab 220. Those that have already left: Sam, Stephan, Ian, Marcel and Andy. I would like to make a particular acknowledgement to Andy for providing so much technical support and teamwork in our niche field of identifying B vitamin biosynthesis pathways. My thesis would not be what it is without him. Those who are still around, Andre and Paddy. These two I would also like to extend special thanks to. Andre also for providing great support, in particular with bioinformatics. Paddy I thank for being a particularly good drinking partner, something I hope continues in the future. A special mention must go to Aleix Gorchs-Rovira, a special person who was sadly taken from us far too early. I am grateful to have known Aleix and will remember him always. When I started my project two others began at the same time: Shelby and Ellen. Shortly after An then arrived, followed by Carrie. These four I would like to thank for being such a good support team and friendship group, you guys are great.

Finally, I'd like to thank my friends and family outside of the lab. Tim Recaldin was one of the most influential people in my decision to pursue a PhD, cheers mate. Some equally influential people from BAT: Laurel, Jen, Luke and Tijs. My family: Beth, Joff, Mum and Dad, thank you for all your help and support. Mum and Dad in particular for the support throughout my education (and to Dad for enduring the task of proofing much of this thesis). Last but not least I thank my partner Laura for her unwavering support and encouragement, you made a difficult task easier.

Table of Contents

Table of Tables

Table number	Section	Content
2.1	2.3.2.1	Relative abundances of dominant prokaryotic taxa
2.2	2.3.2.2	Taxonomic classification of top 20 most abundant prokaryotic genera
2.3	2.3.4.1	Relative abundances of dominant eukaryotic taxa
2.4	2.3.4.2	Taxonomic classification of top 20 most abundant eukaryotic species
3.1	3.2.1, 3.3.1.1	List of species used in in initial assessment
3.2	3.2.2.3, 3.3.1.1	List of B vitamin biosynthesis pathway enzymes
3.3	3.3.1.2	BUSCO completeness of species of interest
3.4	3.3.1.10	Presence and absence of B_{12} related enzymes in stramenopile species
3.5	3.3.1.11	Biosynthesis prediction summary
4.1	4.3.2	Taxonomic classifications of 16S sequences yielded from colony PCR
4.2	4.3.2	Taxonomic classifications of 16S sequences yielded from amplicon sequencing of *I. marina* cultures
4.3	4.3.4	Summary statistics of draft assemblies of long-read data, produced with quast
4.4	4.3.5	Metagenome bin summary statistics 1
4.5	4.3.5	Metagenome bin summary statistics 2
4.6	4.3.5	Anvi'o taxonomic classification predictions of metagenome bins
4.7	4.3.6.1	SINA taxonomic classification predictions of metagenome bins using extracted 16S sequences from each bin
4.8	4.3.6.1	Summary of bacterial identification methods
4.9	4.3.6.4	Summary of BLAST results from nitrogen assimilation enzymes in *I. marina*
4.10	4.3.6.6	List of red-algal annotated genes in the *I. marina* genome
4.11	4.3.6.7	BLAST results from *I. marina* query against mavirus

Table of Figures

Figure number	Section	Content
1.1	1.1.1	Eukaryotic tree of life
1.2	1.1.4	Microbial loop
1.3	1.2.1	Stramenopile phylogeny
1.4	1.2.1	Examples of stramenopiles
1.5	1.3.1	B vitamin molecular structures
2.1	2.1.2	Vitamin amendment experimental design
2.2	2.3.1.1	16S Alpha diversity
2.3	2.3.1.2	16S Beta diversity
2.4	2.3.2.1	16S Relative abundance bar-chart
2.5	2.3.2.2	16S Ratios and relative abundance (top 20 most abundant)
2.6	2.3.2.2	16S ratios as boxplot of top 20 most abundant genera
2.7	2.3.2.2	16S ratios as boxplot of top 20 most abundant families
2.8	2.3.3.1	18S Alpha diversity
2.9	2.3.3.2	18S Beta diversity
2.10	2.3.4.1	18S Relative abundance bar-chart
2.11	2.3.4.2	18S Ratios and relative abundance (top 20 most abundant)
2.12	2.3.4.2	18S ratios as boxplot of top 20 most abundant genera
2.13	2.3.4.2	18S ratios as boxplot of top 20 most abundant ASVs
3.1	3.3.1.1	Phylogeny of the stramenopiles

3.2	3.3.1.3	Thiamine B_1 pathway and presence and absence results
3.3	3.3.1.4, 3.3.1.5	(a) Riboflavin B_2 pathway and presence and absence results, (b) Niacin B_3 pathway and presence and absence results
3.4	3.3.1.6, 3.3.1.7	(a) Pantothenate B_5 pathway and presence and absence results, (b) Pyridoxal B_6 pathway and presence and absence results
3.5	3.3.1.8, 3.3.1.9	(a) Biotin B_7 pathway and presence and absence results, (b) Folate B_9 pathway and presence and absence results
3.6	3.3.2.1	Thiamine B_1 presence and absence large scale stramenopile group
3.7	3.3.2.2, 3.3.2.3	(a) Riboflavin B_2 biosynthesis enzymes, presence and absence large scale stramenopile group, (b) Niacin B_3 biosynthesis enzymes, presence and absence large scale stramenopile group
3.8	3.3.2.4, 3.3.2.5	(a) Pantothenate B_5 biosynthesis enzymes, presence and absence large scale stramenopile group, (b) Pyridoxal B_6 biosynthesis enzymes, presence and absence large scale stramenopile group
3.9	3.3.2.6, 3.3.2.7	(a) Biotin B_7 biosynthesis enzymes, presence and absence large scale stramenopile group, (b) Folate B_9 biosynthesis enzymes, presence and absence large scale stramenopile group
3.10	3.3.2.8	B_{12} associated enzymes, presence and absence large scale stramenopile group
4.1	4.3.1.1	(a-d) Microscopy of *I. marina* cultures
4.2	4.3.1.1	Microscopy images of antibiotic treatment experiment
4.3	4.3.1.1	Marine broth agar plates spread with samples from cultures after antibiotic experiment
4.4	4.3.1.2	SEM imaging of *I. marina*
4.5	4.3.1.2	SEM image of degrading *I. marina* cell being consumed by bacteria
4.6	4.3.2	Images of marine broth plates spread with samples from *I. marina* cultures, used for colony PCR to ID bacteria
4.7	4.3.3	Quality control graphs of *I. marina* culture long-read sequencing data
4.8	4.3.4	Assembly workflow schematic
4.9	4.3.5	Metagenomic analysis workflow schematic
4.10	4.3.5	Metagenome binning, graphical representation
4.11	4.3.6.2	Schematic of *I. marina* MAG analysis workflow
4.12	4.3.6.2	Meta analysis of genome completeness for available stramenopile genomes
4.13	4.3.6.2	K-mer frequency plot of *I. marina* MAG
4.14	4.3.6.2	Summary of ghostKOALA annotation of *I. marina*
4.15	4.3.6.2	COG annotation summary
4.16	4.3.6.3	*I. marina* community B vitamin biosynthesis pathway analysis
4.17	4.3.6.4	*I. marina* community amino acid biosynthesis pathway analysis
4.18	4.3.6.5	DMSP biosynthesis pathway
4.19	4.3.6.5	Phylogenetic gene tree of putative DSYB seqences

Table of Abbreviations – in order of first use. (Enzyme name abbreviations for Chapter 3 can be found in Table 3.2)	
Abbreviation	**Definition**
EToL	Eukaryotic tree of life
OTU	Operational taxonomic unit
C:N:P:S	carbon:nitrogen:phosphorus:sulpher
DMSP	dimethylsulfoniopropionate
IAA	indole-3-acetic acid
T(SAR)	Telonemia, (Stramenopila, Alveolata and Rhizaria)
rDNA	Ribosomal DNA
rRNA	Ribosomal RNA
HAB	Harmful algal bloom
MAST	MArine STramenopile
MOCH	Marine OCHrophytes
MAG	Metagenome assembled genome
B_1	Thiamine / Vitamin B_1
B_2	Riboflavin / Vitamin B_2
B_3	Niacin / Vitamin B_3
B_5	Pantothenate / Vitamin B_5
B_6	Pyridoxal / Vitamin B_6
B_7	Biotin / Vitamin B_7
B_9	Folate / Vitamin B_9
B_{12}	Cobalamin / Vitamin B_{12}
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
CoA	Coenzyme A
ASV	Amplicon sequencing variant
GO	Gene ontology
AA	Amino acid
BUSCO	Benchmarking universal single-copy orthologues
SAG	Single-cell amplified genome
KO	KEGG orthologues
IA	Iminoaspartate
NM	Nicotinate mononucleotide
MMETSP	The Marine Microbial Eukaryotic Transcriptome Sequencing Project
EDLs	Early diverging lineages
TPP	Thiamine pyrophosphate
HGT	Horizontal gene transfer
HMM	Hidden Markov-model
SEM	Scanning Electron Microscopy
SMAGs	SAGs + MAGs from TARA
NGS	Next-generation sequencing
ONT	Oxford-nanopore technologies
PacBio	Pacific Biosciences sequencing
ASWP	Artificial sea-water for protists
QC	Quality control
GOGAT	glutamine oxoglutarate aminotransferase
DMS	dimethyl sulfide
MMT	MTHB-methyltransferase
CroV	*Cafeteria roenbergensis virus*
EBP	The Earth BioGenome Project

Table of species. Species referred to at multiple points throughout the thesis.	
Full name	**Abbreviation**
Ruegeria pomeroyi	*R. pomeroyi*
Thallasiosira pseudonana	*T. pseudonana*
Asterionellopsis glacialis	*A. glacialis*
Pseudo-Nitzschia multiseries	*P. multiseries*
Cafeteria roenbergensis	*C. roenbergensis*
Fragilariopsis cylindrus	*F. cylindrus*
Pseudophyllomitus vesiculosus	*P. vesiculosus*
Phytophthora sojae	*P. sojae*
Phytophthora infestans	*P. infestans*
Nannochloropsis gaditana	*N. gaditana*
Ostreococcus tauri	*O. tauri*
Dinoroseobacter shibae	*D. shibae*
Incisomonas marina	*I. marina*
Escherichia coli	*E. coli*
Nannochloropsis oceanica	*N. oceanica*
Aureococcus anophagefferens	*A. anophagefferens*
Phytophthora cinnamomi var cinnamomi	*P. cinnamomi var cinnamomi*
Platysulcus tardus	*P. tardus*
Euglena gracilis	*E. gracilis*
Schizochytrium aggregatum	*S. aggregatum*
Aplanochytrium kerguelense	*A. kerguelense*
Aurantiochytrium limacinum	*A. limacinum*
Porphyridium purpureum	*P. purpureum*
Winogradskyella sediminis	*W. sediminis*
Marinobacter hydrocarbonoclasticus	*M. hydrocarbonoclasticus*
Chrysocampanula spinifera	*C. spinifera*

1 – Introduction

1.1 The evolution of microbial eukaryotes and their importance

1.1.1 The eukaryotes

The eukaryotic cell evolved around 2.7 billion years ago (Cooper 2000). A single, unlikely event paved the way for the plethora of diversity in species seen on earth today. One prokaryotic cell engulfed another, but instead of being digested the engulfed cell persisted and was subsequently maintained within the host cell. The engulfed cell became reduced in function and developed into the organelle known to exist in virtually every eukaryotic cell on earth today, the mitochondrion. This process, endosymbiosis, occurred again *c.* 1.5 billion years ago. A heterotrophic eukaryote engulfed a photosynthetic prokaryote, a β–cyanobacterium (Yoon *et al.,* 2004). Again, the prey persisted in the host, gradually decaying into the first plastid (photosynthetic organelle) and leading to the first photosynthetic eukaryote. This is referred to as the primary endosymbiosis event and gave rise to glaucophytes, green and red algae (and subsequently all extant plants) (Stephens *et al.,* 2021). Since these two events the process of evolution has resulted in the enormous proliferation of both non-photosynthetic and photosynthetic eukaryotes. The vast majority of these lineages have remained as single-celled microbes (Grattepanche *et al.,* 2018; Massana *et al.,* 2014). Microbial life is complex and diverse and is found in every super-group of the eukaryotic tree of life (EToL –Figure 1.1). Despite over 165 years of the microbiology discipline (Opal, 2009), and huge advancements in our understanding, many of the intricacies and nuances displayed by this diversity remain to be discovered.

1.1.2 Diversity within the eukaryotes and research biases

Algae is a generic term used to describe photosynthetic taxa that are not plant life. It is a polyphyletic group with representatives from across the EToL. Some members of the bacterial kingdom, the cyanobacteria, are also frequently included in the algal group due to their photosynthetic capabilities, but this is strictly functional rather than phylogenetic. Algal representatives can be found in the following groups of note: Stramenopila (e.g. diatoms), Haptista (e.g. coccolithophores), Archaeplastida (e.g. green and red algae) and Discoba (e.g. Euglenozoa). The polyphyletic nature of the algae is demonstrated in Figure 1.1 where the groups containing algae in a schematic of the EToL are indicated by a circle at the group leaf. The evolution of these algal groups has been a source of contentious debate for many years. One hypothesis, the chromalveolate hypothesis, suggested a common secondary endosymbiosis event that gave rise to all red-algal plastid containing algae which includes members of the alveolates, stramenopiles and

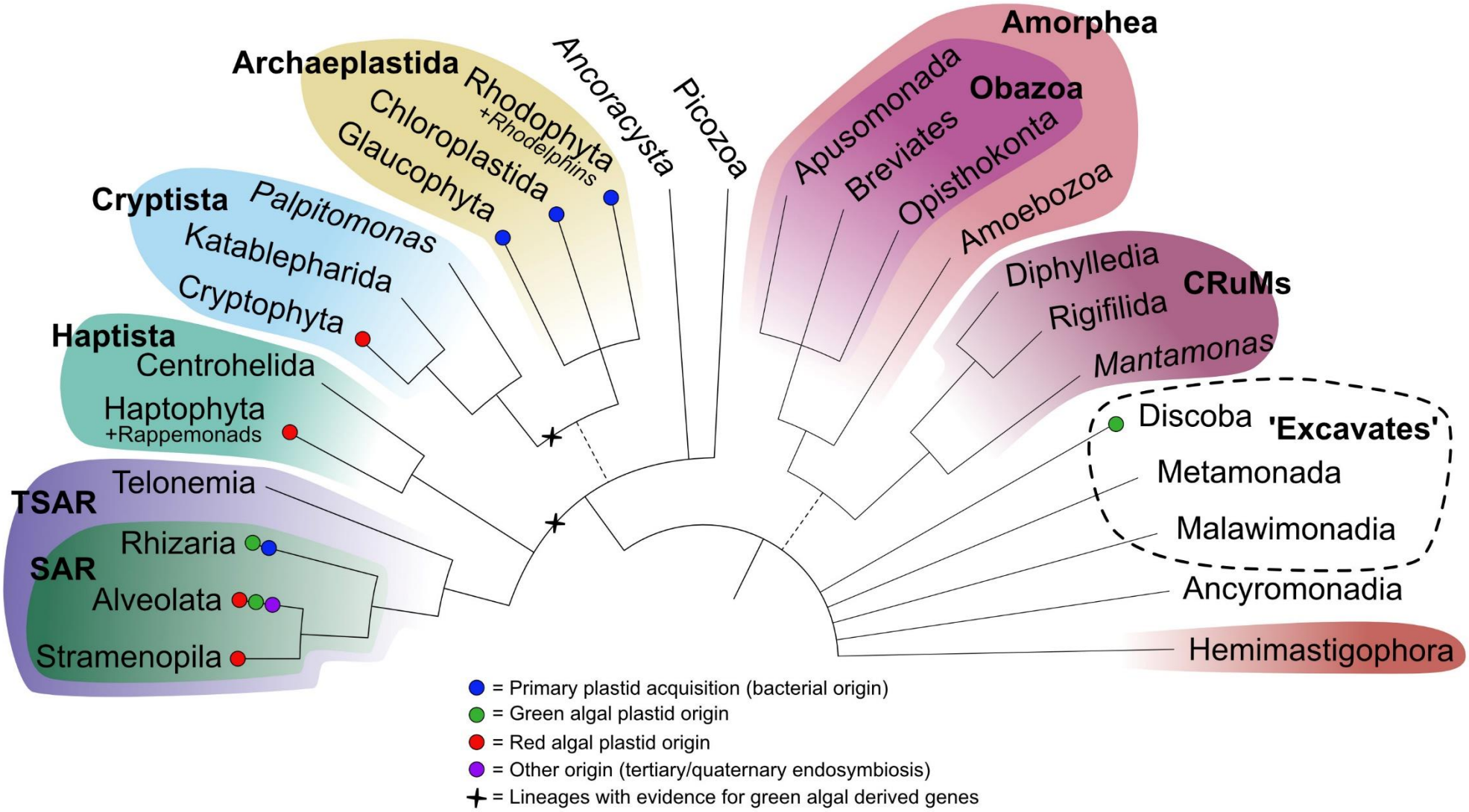


Figure 1.1 The Eukaryotic Tree of Life. *A summary of the latest consensus eukaryotic phylogeny as presented by Burki et al. (2020). Circles at the end of branches indicate that the group contains photosynthetic species. The colour represents the likely origin of the plastid in the lineages in each group of phototrophs (Dorrell and Smith 2011, Marin et al., 2005). A blue circle indicates organisms of this lineage have plastids derived from the primary endosymbiosis event. Green circles indicates the group contains organisms with plastids deriving from green algae. Red circles show members of the group have plastids derived from a red alga and purple circles show groups that have acquiredired a plastid from another source, via tertiary of quarternary symbiosis. Crosses represent lineages which contain evidence for cryptic green algal derived genes as discusssed by Dorrell and Smith (2011). Multifurcations indicate unresolved branching at those points and dotted lines indicate lesser uncertainties of the monophyly of those groups. T(SAR) = Telonemia, (Stramenopila, Alveolata, Rhizaria). CRuMs = Collodictyonids (syn. diphylleids) + Rigifilida + Mantamonas.*

haptophytes (Cavalier-Smith, 1999; Keeling, 2009). This however has been disputed in recent years with strong evidence from recent advancements in phylogenomics suggesting that these groups are distant in evolutionary history (Burki *et al.,* 2020; Burki 2017) .There is also little evidence of any remnant red-algal plastids in the non-photosynthetic members of these algae containing groups (Stiller *et al.,* 2009; Wang, Sun, and Huang 2017), which would be required to explain the close relationships for example between the oomycetes and diatoms in the Stramenopila. The current consensus view of the evolution of photosynthesis in these separate groups is by a series of endosymbiosis events of varying levels (secondary, tertiary and quaternary) (Archibald, 2009; Dorrell and Smith, 2011). Secondary endosymbiosis refers to an event in which a primary endosymbiosis derived algal cell is phagotrophically engulfed and the photosynthetic apparatus is maintained. Tertiary and quaternary events are phagotrophy of a secondary or tertiary endosymbiosis derived algal cell respectively. A model for the instances of these events is proposed by Stiller et al. (2014). This model suggests that cryptophyte algae underwent secondary endosymbiosis of a red alga, followed by tertiary endosymbiosis of a cryptophyte to give rise to the ochrophytes (within the stramenopiles) and a subsequent quaternary event resulting in the haptophytes (Stiller *et al.,* 2014). The instances of these events are yet to be proven however the origins of plastids from the various algal groups are known (Dorrell and Smith, 2011) and can be seen mapped on to the EToL in Figure 1.1 in the form of coloured circles representing the various plastid origins. There is however some evidence of a cryptic green algal endosymbiosis event that may have occurred prior to the acquisition of the red algal endo-symbiont. This was first identified by Moustafa et al. (2009) who found green-algal derived genes in the genome of two diatoms. Although these result have since been disputed, the consensus remains that the presence of such green-algal-derived genes hints at a further complication to the story of plastid acquisition (Dorrell and Smith 2011).

Despite displaying a huge range of diversity themselves, the algae represent only a fraction of the diversity seen in the whole of the EToL. The diversity displayed by heterotrophic lineages is equal to, if not surpasses, that of the algae. In recent years sequencing has become more accessible and has been greatly reduced in cost. This, combined with ambitious global expeditions, has seen a boom in metagenomic sequencing information (Heidelberg, Gilbert, and Joint 2010; Falkowski and De Vargas 2004). Experiments such as the Global Oceans Sampling expedition (Venter *et al.,* 2004; Rusch *et al.,* 2007) and the TARA Oceans expedition (Bork *et al.,* 2015) as well as many more localised sequencing efforts (e.g. Massana *et al.,* 2002; 2014; 2004; del Campo and Massana 2011;

Taylor and Cunliffe 2014) has revealed to a greater extent the diversity of protist life in the global oceans. de Vargas et al. (2015) estimate a total of around 150,000 planktonic eukaryote operational taxonomic units (OTUs) in the marine photic zone (measured by 18S amplicon sequencing) and contrast this to the number of formally described species that match these parameters, which is ~11,000.

An almost incomprehensible level of diversity exists in these environments and awaits classification and subsequent research. However, a traditional research bias has prevented the discovery of much of this diversity. The anthropogenic orientated research curriculum has focused on microbial life that either is photosynthetic, or is pathogenic so has medical implications (such as the human pathogens belonging to the Alveolata, e.g. *Plasmodium falciparum*, which causes malaria), or affects the growth of crop species (e.g. the oomycetes of the stramenopile group including *Phytophthora*, the causative agent of potato blight and soy bean stem rot) (Burki and Keeling 2014; del Campo *et al.,* 2014). Historical, and to a significant extent modern, research has neglected the unseen and unculturable microorganisms, which happen to be a huge range of taxa. Thanks to the endeavours of the recent environmental sequencing efforts we now at least appreciate the enormity of the gulf in our understanding and can make efforts to address it.

1.1.3 Microbial communities

Microbial cells are seldom found in isolation. Community dynamics of single celled organisms adds a layer of complexity of which we have even less understanding. This is despite the acute importance of these microbial communities to the health of every environment on the planet as well as the macro inhabitants of those environments.

Microbial communities in the guts of megafauna are key to their survival, aiding digestion, fighting disease and providing nutrients (Shreiner *et al.,* 2015). Over the last few decades our understanding of gut microbiota has increased significantly. We now appreciate that the composition or health of the gut microbiome can be linked to a wide range of disease states such as diabetes (Karlsson *et al.,* 2013) and obesity (Le Chatelier *et al.,* 2013). The microbial consortium associated with the soil is essential for productive terrestrial ecosystems, indeed many plants have evolved an absolute requirement for the presence of microbes surrounding their roots (Bonfante and Genre, 2010). Furthermore, the single-celled organisms, both eukaryotic and prokaryotic, that make up the communities of the ocean are fundamental to the global biogeochemical cycling and lay the foundations of every food-chain above them (Falkowski, Fenchel, and Delong 2008).

1.1.4 An ocean of microbes

The oceans cover approximately 70% of the Earth's surface, yet of the estimated 550 gigatons of carbon (Gt C) held as biomass on the planet only 1% is found there (Bar-On *et al.,* 2018), and this is heavily dominated by microbial taxa. Of the estimated 6 Gt C present in the ocean ~4 Gt C exists in the form of protists, algae, bacteria and archaea (Bar-On and Milo, 2019). These groups of organisms, along with the marine fungi, form microbial communities that are essential for the function of the ocean. One functional group of organisms within this are the primary producers (i.e. photosynthesing phytoplankton) of the ocean biome, reportedly responsible for around half of all global primary productivity (Field *et al.,* 1998). This statistic is made more remarkable by the relative proportions of producers to consumers in the marine environment, which is estimated to be 1:5. This demonstrates the astonishing rate of turnover of these organisms and how rapid the cycling of carbon is in this environment as well as all other nutrients (Bar-On and Milo, 2019).

Primary producers fix inorganic carbon and subsequently make it available to enter the carbon cycle. The carbon cycle relies on carbon fixation by autotrophs, which then enter the grazing food chain, or the microbial loop (Azam *et al.,* 1983). The microbial loop describes the cycling of nutrients through the organisms of the size fraction <200µm and built on the simplistic classical view of the food chain from phytoplankton upwards. It was described in 1983 but has since been

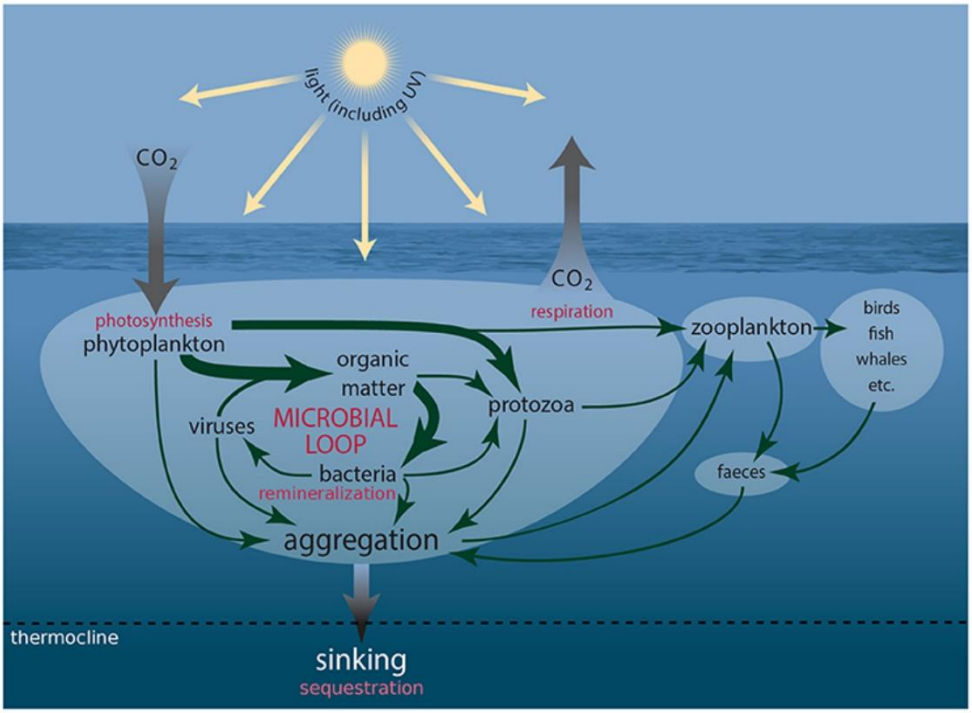


Figure 1.2 The microbial loop. *The microbial loop as proposed by Azam et al. (1983) and updated by Fenchel (2008). This schematic shows the connections amongst members of the microbial loop and the flux of carbon between trophic levels and tot he deep ocean. This figure is taken from Davidson et al. (2017).*

refined to include the varying trophic modes of many organisms such as mixotrophy demonstrated by many photosynthetic species (Fenchel, 2008). The originally proposed microbial loop, its updates and its link to the food chain is shown in Figure 1.2. Primary production in the oceans is primarily carried out by phytoplankton. Planktonic species in general are defined by their inability to move against the ocean currents. Phytoplankton are further categorised by their ability to photosynthesise and includes both cyanobacteria and eukaryotic algae. Non-photosynthetic protists also play a significant role in the cycling of carbon as well as other macro nutrients including nitrogen, phosphorus and sulphur. This is achieved by multiple processes including grazing of primary producers, other protists and bacteria, as well as osmotrophic absorption of nutrients from the "nutrient pool" created by decaying cells and nutrient upwelling from the deep ocean. The processing of major elements by microbial communities in the oceans maintains the balance of carbon:nitrogen:phosphorus:sulpher (C:N:P:S) (excluding anthropogenic alteration) and serves the planet as key biogeochemical engineers (Arrigo 2005; Falkowski, Fenchel, and Delong 2008).

1.1.4.1 Communities of the ocean – competing and sharing

When measured on geological timescales and averaged over the large volume of the oceans the ratios of C:N:P:S appear relatively stable. However, in local temporospatial scales the resources that microbial life relies on are often at a concentration that is limiting to both their activity and abundance. This has been demonstrated, for example, by enrichment of environmental samples with nitrogen (nitrate and ammonia) resulting in increases in photochemical efficiency and growth rates of the community (Graziano *et al.,* 1996). In addition, as a result of nutrient limitation, the composition of microbial communities exhibits seasonal changes. For example, during the summer and autumn months the Red Sea becomes nutrient deplete and is dominated by picoplankton species. However, during winter nutrient levels were seen to rise and the proportion of larger phytoplankton cells increased (Mackey *et al.,* 2007). Indeed, it is likely that there is a co-limitation effect on the community composition by nutrient availability and environmental factors as demonstrated by Anderson et al. (2022).

Nutrient limitation can be particularly acute in the photic zone, a layer of sub-surface water approximately 200m deep that is penetrated by the light of the sun, which hosts the photoautotrophic assemblage of oceanic microbes along with many of their heterotrophic contemporaries. Two of the most significant nutrient limitation patterns that have been studied

over the recent past are a limitation of nitrogen in the southern oceans and also of iron in areas that are enhanced with subsurface nutrient supply (Moore *et al.,* 2013). However, nutrient limitation is a complex state that depends not only on the concentration of a nutrient but also on the composition of the local microbial community and the relative bioavailability of the nutrient. Indeed, in many cases two or more nutrients will be co-limiting to certain members of the community (Browning *et al.,* 2017). As well as the predominant limiting nutrients (N, P and Fe) studies have demonstrated that many other micronutrients can be limiting in local environments, such as zinc, cobalt and vitamin B_{12} (Koch and Trimborn, 2019). The competition for these limited resources drives the turnover of microbial communities.

Such oceanic microbial communities can be highly diverse and complex in structure and function. This is well demonstrated by Lima-Mendez et al. (2015) in the interactome of the TARA oceans expedition data from the ocean photic zone. They demonstrate that the structure of the microbial community is formed not only by simple environmental factors but also all types of biotic interaction and function: grazing, primary production, parasitism and symbiosis. Although photosynthetic primary producers are often seen as, in a reductionist sense, the bottom of the food chain forming the base of these communities in reality many species actually have a mixotrophic lifestyle combining carbon fixation with bactorivory or osmotrophy. This further complicates community dynamics at the micro-scale.

That our understanding of the complexities of aquatic microbial interactions is limited is perhaps not surprising given the aforementioned research bias, the difficulty in culturing many marine species and our traditional terrestrial and anthropogenic viewpoint (Kazamia *et al.,* 2016). Microbial interaction is a vital component for these communities to function at full capacity. Interactions can be broadly described as active or passive (Kazamia *et al.,* 2016). A passive interaction could constitute the uptake of nutrients from the dissolved pool of nutrients, which reached this pool by "leaking" from living cells or as a result of cell lysis. Active interactions include trophic interactions as well as symbiotic interaction where both parties actively engage in the interaction.

1.1.4.2 Evidence of microbial interactions

Evidence of cell-cell interactions is hard to decipher in the environment. Various methods have been employed in attempt to identify, understand and categorise interactions. The advent of environmental sequencing allowed the collection of sequencing data for an unprecedented number of planktonic microbes (Falkowski and De Vargas 2004). Coupled with accurate location

metadata this could be transformed into co-occurrence matrices (e.g. Venter *et al.,* 2004; Rusch *et al.,* 2007). The aforementioned study of the TARA Oceans data by Lima-Mendez et al. (2015) took this a step further using machine learning and network analysis. This allowed them to predict both positive and negative interactions totalling 81590 interactions. The majority of these were characterised as positive (~78%). These predicted interactions also corroborated a number of previously reported interactions (Jolley and Jones 1977; Martinez-Garcia *et al.,* 2012). Although invaluable in furthering our appreciation of the diversity, abundance and co-occurrence of microbes in the ocean can only reveal so much. It does not account for any metabolic function underlying the predicted interactions.

Controlled co-culturing of individual species allows greater interrogation of interactions between the parties involved. Leveraging a variety of techniques, such as metatranscriptomics and proteomics allows for the unveiling of the underlying mechanism of the interaction and how it may have evolved (Kazamia *et al.,* 2016). Various examples of microbial interactions have been demonstrated in this way. When grown in coculture with the bacterium *Ruegeria pomeroyi*, the diatom *Thalassiosira pseudonana* undergoes differential expression of genes required for the recognition of external stimuli (Durham *et al.,* 2017). It also alters its metabolism, downregulating lipid biosynthesis and upregulating chitin metabolism. Subsequent changes seen in the transporter system of *R. pomeroyi* suggest that it survives and grows in the co-culture on diatom-derived nutrients. Interactions between phytoplankton and bacteria in the ocean is of particular importance as this contributes to the regulation of the microbial loop (Figure 1.2).

Signalling is an essential interaction within microbial communities. Indeed, bacteria-bacteria signalling, known as quorum sensing, has been reported in the scientific literature for over 40 years (Miller and Bassler, 2001; Nealson and Hastings, 1979). It is a mechanism by which individual bacterial cells can communicate at the level of the whole-community in response to external stimuli, for example population size. Interactions between bacteria and algal species are known to induce quorum sensing (Johnson *et al.,* 2016). Signalling between diatoms and bacteria has been investigated in recent years due to the environmental significance of diatoms and the influence of these interaction on global geochemical cycling. Shibl et al. (2020) identify two unique secondary metabolites that the diatom *Asterionellopsis glacialis* uses to modulate bacteria. Another study demonstrated that more common metabolites such as the amino acids tryptophan and taurine as well as the sulphur containing compound dimethylsulfoniopropionate (DMSP) can be released by

the diatom *Pseudo-Nitzschia multiseries* and taken up by bacteria in exchange for ammonium and the hormone indole-3-acetic acid (IAA) (Cirri and Pohnert, 2019). This, and the relevance to this thesis is further introduced in the introduction of Chapter 4.

1.2 The stramenopiles and their significance

A significant proportion of the diversity discussed above can be attributed to the stramenopile group of the EToL. In the latest consensus tree of the eukaryotes (Burki *et al.,* 2020) (Figure 1.1), the Stramenopila are situated within a recently defined supergroup known as TSAR. This supergroup contains the subdivisions Telonemia which is sister to the SAR group, which itself encompasses the groups Rhizaria, Alveolata and the Stramenopila. Each of these groups comprise of mostly single-cell species and include both photosynthetic and heterotrophic lineages. The photosynthetic groups of the stramenopiles have red algal derived plastids (Dorrell and Smith, 2011). The Rhiazaria on the other hand have predominantly been identified to have green algal derived plastids. The exception to this is the species *Paulinella chromatophora* which is a rare instance of primary plastid acquisition from a bacterium (Marin, Nowack, and Melkonian 2005). Alveolates are a diverse group, which shows examples of species with plastids deriving from a green alga, red alga and tertiary endosymbiosis. For example, the dinotoms are a group that have evolved as a result of a heterotrophic dinoflagellate engulfing a diatom (stramenopile) and subsequently maintaining its plastid (Imanian *et al.,* 2010).

1.2.1 Well characterised diversity within the stramenopiles

The stramenopile group contains thousands of species inhabiting terrestrial, freshwater and oceanic environments and with a cosmopolitan distribution ranging from the Arctic to the Antarctic circle. They display diversity in morphology with some members having evolved multicellularity (e.g., *Ectocarpus siliculosus*), others being single-cell bi-flagellates (e.g., *Cafeteria roenbergensis*) and some forming siliceous frustules (e.g., *Fragilariopsis cylindrus*). Figure 1.3 is a phylogeny of the group and Figure 1.4 displays some of the variety of the stramenopiles including those previously mentioned among others. As well as morphological and environmental multiplicity, stramenopiles also display a wide range of lifestyle. A large proportion of the group, and arguably the most well-studied, are the photosynthetic species. Collectively these are the Ochrophyta and include diatoms and kelps amongst many other groups. Other members of the stramenopiles perform heterotrophy such as the Bikosia (e.g., *C. ronebergensis*) and the Eogyra (e.g., *Pseudophyllomitus vesiculosus*). The Labyrinthulae are likely to perform osmotrophy and saprotrophy, turning over and remineralising decaying matter in the ocean (Martin *et al.,* 2016; Raghukumar, 2002). Finally,

parasitic species are spread across the group including within the Oomycota (e.g., *Phytophthora sojae*) and the Placidozoa (e.g., *Blastocystis hominis*).

Diatoms, formally known as the Bacillariophyceae (Figure 1.4 d-e), are the most well studied group of stramenopiles. These photosynthetic protists are ubiquitous in the marine environment found in every major body of water including at both poles. They are responsible for a significant proportion of global primary productivity (frequently reported as contributing 20%) (Nelson *et al.,* 1995; Falkowski and Knoll 2007). This organic carbon production is a key component to sustaining marine food webs and is consumed at a high rate. This makes the diatoms a key contributor to the carbon cycle (Armbrust, 2009). As well as their acute ecological importance diatoms are beginning to emerge as potential industrial players with various biotechnological applications being explored for future "green industry" (Sharma *et al.,* 2021; Jamali *et al.,* 2012). In the assessment of the TARA Oceans sequencing data by de Vargas et al. (2015), the Bacillariophyta were the 5th most abundant group of eukaryotes by number of rDNA reads and 6th most diverse as defined by number of OTUs identified.

The second most well-studied group of stramenopiles is the oomycetes. Also referred to as the pseudo-fungi, these soil-living parasites were initially considered fungi due to the similarities in morphological features. The genera *Phytophthora* has significant impact on terrestrial and agricultural landscapes. *P. sojae* causes soybean root and stem rot resulting in major crop loss every year with an estimated cost of $1-2 billion (Tyler, 2007). *Phytophthora infestans* is the parasite responsible for triggering the Irish potato famine and remains a destructive pathogen today (Goss *et al.,* 2014).

Various other stramenopile species play significant ecological roles or have potentially exploitable traits, which has led to significant research on these organisms. Examples of this include the aforementioned Labyrinthulae which play key roles in recycling nutrients in the ocean, the Eustigmatophyceae species *Nannochloropsis gaditana* which possesses promising lipid production capabilities (Ajjawi *et al.,* 2017) and the Raphidophyceae genera *Chattonella spp.,* which form harmful-algal-blooms (HABs) (Viana *et al.,* 2019).

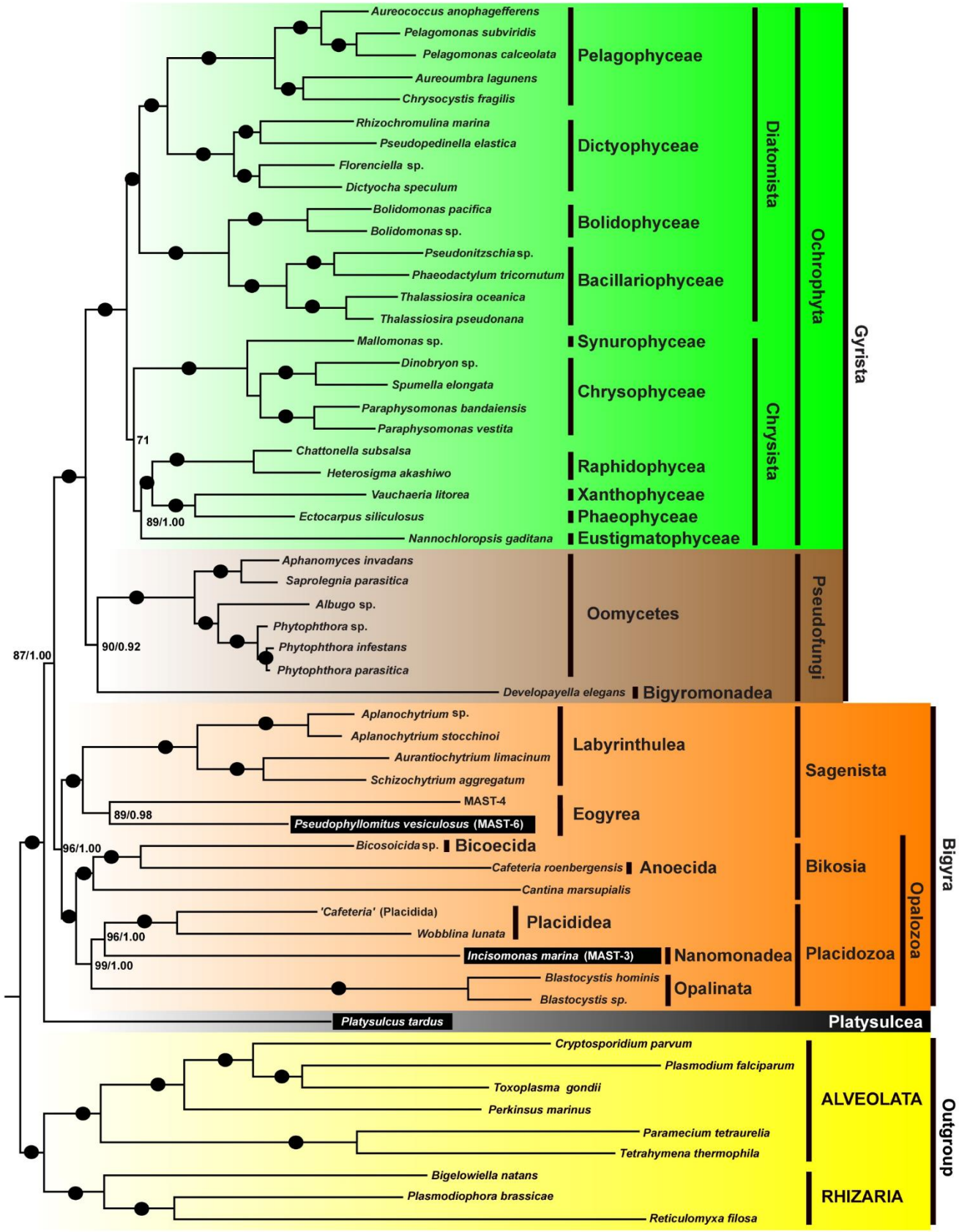


Figure 1.3 Phylogeny of the stramenopiles. *Phylogenomic analysis presented by Thakur et al. (2019). Maximum-likelihood tree based on the model LG + C60+F + G. ML bootstrap percentage (BP) values and Bayesian posterior probabilities (PP) where BP < 70% and PP < 0.92 are omitted from the figure and black dots correspond to a BP of 100% and a PP of 1.00. The tree is rooted on an outgroup consisting of members of the Alveolata and Rhizaria. Species highlighted in black boxes are three transcriptomes presented in the original paper.*

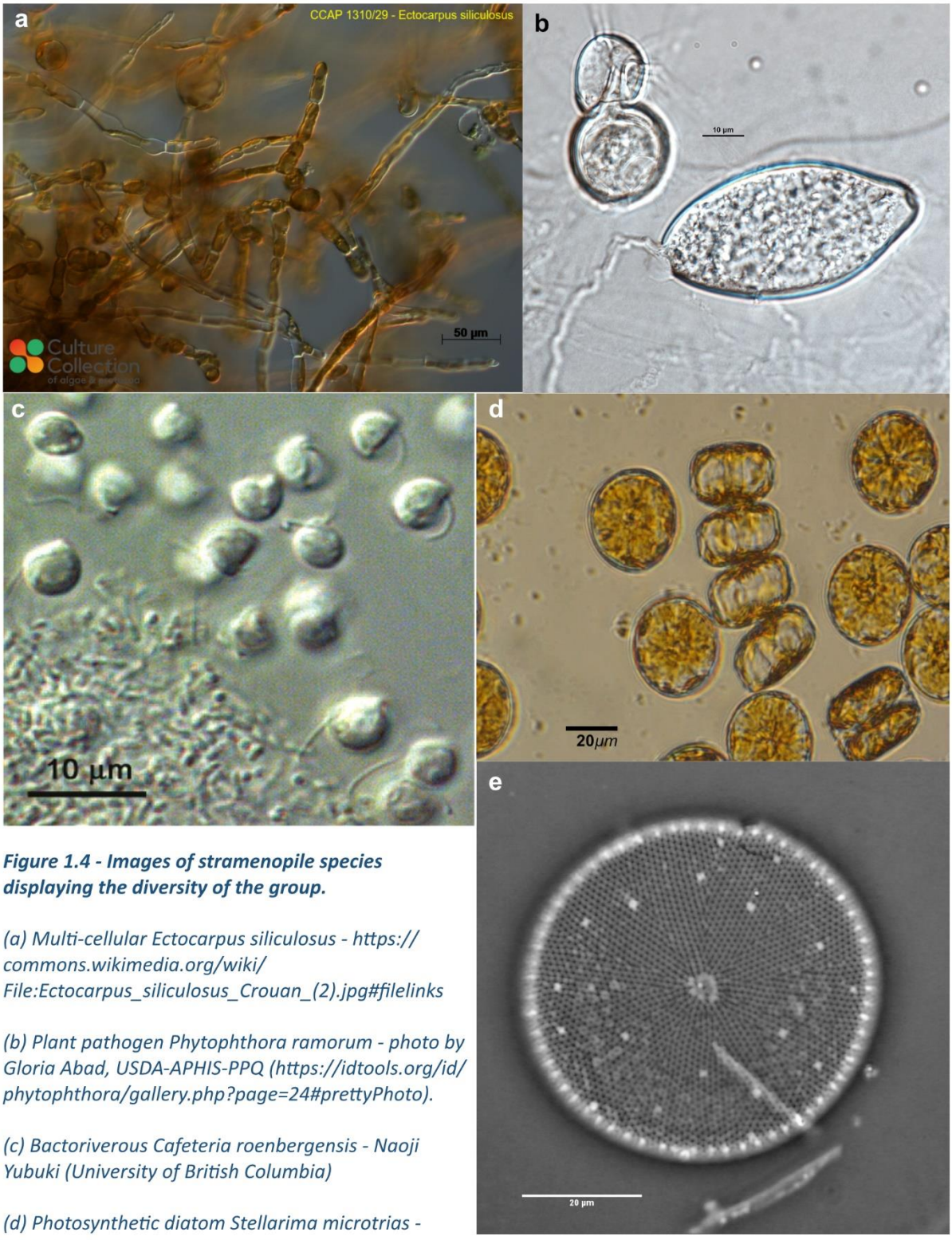


Figure 1.4 - Images of stramenopile species displaying the diversity of the group.

(a) Multi-cellular Ectocarpus siliculosus - https://commons.wikimedia.org/wiki/File:Ectocarpus_siliculosus_Crouan_(2).jpg#filelinks

(b) Plant pathogen Phytophthora ramorum - photo by Gloria Abad, USDA-APHIS-PPQ (https://idtools.org/id/phytophthora/gallery.php?page=24#prettyPhoto).

(c) Bactoriverous Cafeteria roenbergensis - Naoji Yubuki (University of British Columbia)

(d) Photosynthetic diatom Stellarima microtrias - image provided by Dr Sam Coffin

(e) Siliceous frustule of the diatom Thalassiosira tumida - image provided by Dr Sam Coffin

1.2.2 Known unknowns in the stramenopiles

As well as the vast "known" diversity described above the stramenopile group harbours many species that constitute "known-unknowns" particularly in the basal lineages of the group (Bigyra – see Figure 1.3). These species can be considered as known-unknowns by their identification via molecular environmental sampling (e.g., via amplicon sequencing experiments) but their lack of sustainable culture. Massana et al. (2004, 2014) used the 18S rRNA gene to identify a number of "ribogroups" from environmental sampling. These are previously unidentified OTUs that group with other stramenopile 18S sequences but are themselves distinct. Using this data 18 groups termed MASTs (MArine STramenopiles) were defined and represent a large proportion of species yet to be fully characterised. On top of these MAST groups, 5 ribogoups were defined within the photosynthetic Ochrophyta group, named MOCHs (Marine OCHrophytes). Indeed, the TARA Oceans expedition yielded millions of sequencing reads relating to stramenopile species (de Vargas *et al.,* 2015). In addition to those discussed in the previous section MAST groups 3 and 12 collectively had over 2.5 million rDNA reads and groups 4, 6, 7, 8, 9, 10 and 11 also collectively had over 2.5 million rDNA reads. The groups MAST3 and MAST12 also displayed some of the highest diversity by OTU count (>500). Delmont et al. (2022) recovered 174 stramenopile metagenome assembled genomes (MAGs) from TARA oceans metagenomic sequencing data. Of these MAGs, 35% were putative heterotrophs. This is in addition to a number of single cell amplified genomes derived from cell samples from the same expedition (Seeleuthner *et al.,* 2018). These data will be explored in Chapter 3.

MASTs and other as-yet uncultured stramenopiles, including those in generally well-studied groups, warrant further exploration for a number of reasons. In a fundamental sense these organisms frequently represent a large proportion of sequencing reads in environmental studies, which indicates a high level of abundance in the ocean (e.g., de Vargas *et al.,* 2015; Taylor and Cunliffe 2014). Our lack of understanding of these organisms limits our fundamental understanding of how this ecosystem functions in detail. Further, characterising a greater number of these species by generating physiological and genomic data will provide insights into the evolution of this group. Finally, given the diversity in lifestyle and metabolism displayed by their relatives, there is reason to expect to uncover more novel and potentially useful traits for future industrial purposes. A further introduction to the stramenopiles is given in Chapter 3.

1.3 Metabolism and B vitamins

Metabolism is the manipulation of organic molecules by a cell to transform basic molecules into those required for energy production, growth and cell replication (Cooper 2000). Over the course of evolution, metabolism has become more complex, in turn allowing life to evolve greater complexity with this process repeating in a cyclical manner. In extant organisms, metabolism is aided significantly by enzymatic co-factors. Approximately one third of all enzymes require a cofactor for function, either a metal cofactor such as iron, or a coenzyme, defined as cofactor that has an organic component (e.g., vitamins) (Monteverde *et al.,* 2017). Arguably, the most important group of cofactors is the B vitamins. This group make up 31% of all cofactors in use in extant organisms. Some have suggested that they may have been present before life emerged and indeed aided the formation of life c. 3.8 billion years ago (Monteverde *et al.,* 2017). In those organisms present on Earth today the ability to synthesise certain B vitamins *de novo* is common although not ubiquitous. The biosynthesis pathways for the B vitamins have been demonstrated as present in some organisms but not others whilst the requirement for these cofactors remains omnipresent. This results in the definition of an auxotroph; an organism that has an obligate requirement for a nutrient but lacks the ability to biosynthesise the nutrient *de novo* and subsequently must acquire the nutrient from an external source. Auxotrophy for B vitamins is widespread in extant organisms, both prokaryotic and eukaryotic, although the prevalence of auxotrophy for each B vitamin is variable (e.g. B_{12} auxotrophy is more common than B_7) (Croft, Warren, and Smith 2006; Paerl *et al.,* 2018).

1.3.1 The discovery of B vitamins

The B vitamins are a group of eight water-soluble molecules. They were initially termed “vitamin B” as a place holder for the water-soluble organic material found in yeast, wheat germ and other nutritionally rich sources, as distinct from the major nutritional groups (proteins, lipids, etc.). Upon identification of specific molecules, vitamin B was classified into separate entities (Robinson 1951; Emmett and Luros 1920). The current B vitamin nomenclature is as follows:

Vitamin B_1 – Thiamine – The first B vitamin to be elucidated, it plays an integral role in central metabolism and energy production (Fattal-Valevski, 2011).

Vitamin B_2 – Riboflavin – Biologically active in the forms flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), important in a variety of redox reactions in the cell (Powers, 2003).

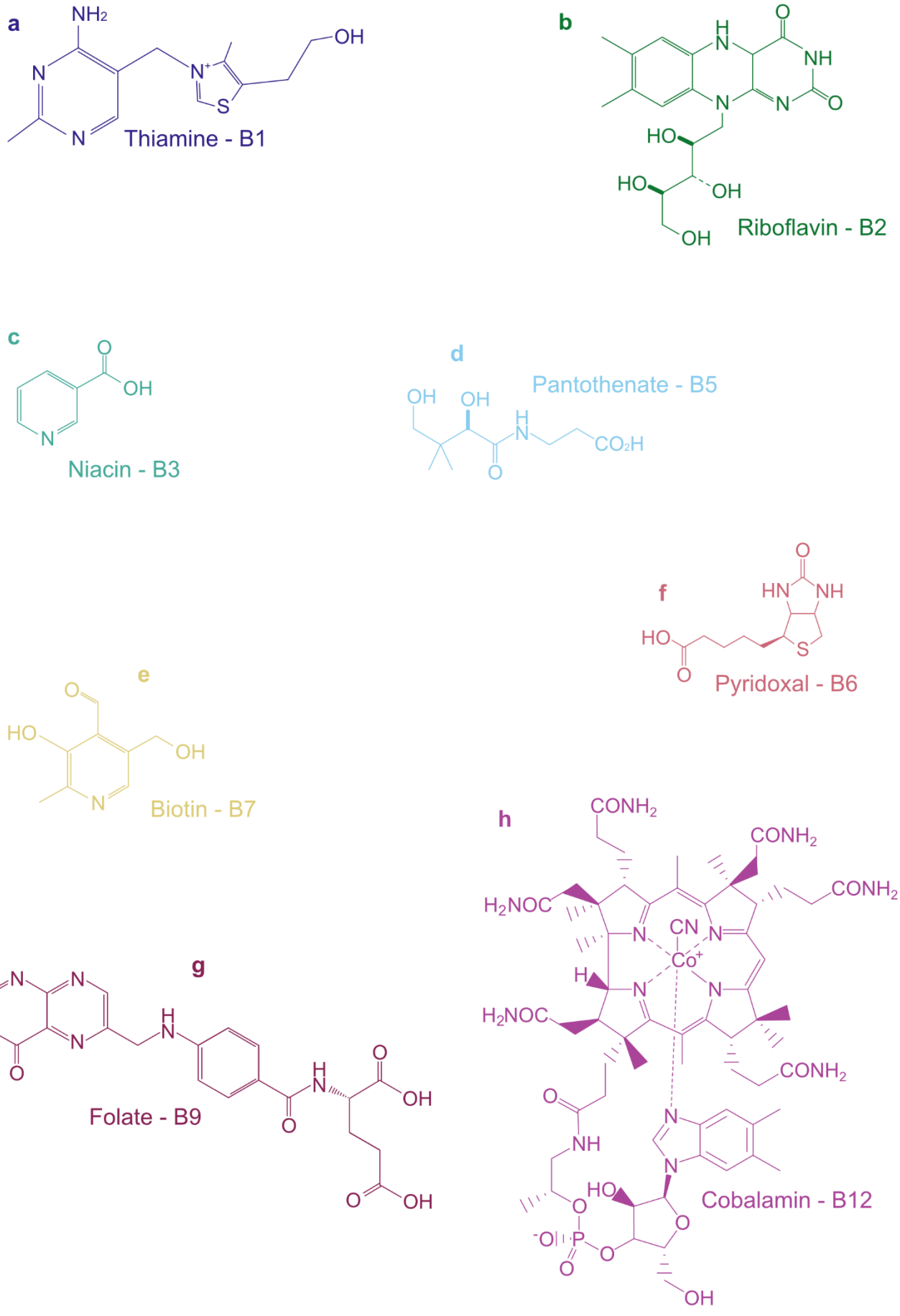


Figure 1.5 - The chemical structures of the B vitamins. *(a) Thiamine - B1. (b) Riboflavin B2. (c) Niacin - B3. (d) Pantothenate - B5. (e) Pyridoxal - B6. (f) Biotin - B7. (g) Folate - B9. (h) Cobalamin - B12.*

utilised in over 400 metabolic reactions (Meyer-Ficca and Kirkland, 2016)

Vitamin B_5 – Pantothenate – this B vitamin is essential in the synthesis of Coenzyme A (CoA) and acyl-carrier protein. These are essential in energy metabolism and fatty acid synthesis (Coxon *et al.,* 2005).

Vitamin B_6 – Pyridoxal – the collective term for six interconvertible forms, pyridoxamine, pyridoxal, pyridoxine and their phosphorylated counterparts. They are involved in many reactions most significantly in amino acid metabolism (Havaux *et al.,* 2009).

Vitamin B_7 – Biotin – this coenzyme is used by a variety of carboxylase enzymes but also plays roles in epigenetic control (Zempleni *et al.,* 2009).

Vitamin B_9 – Folate – involved in C1 transfer reactions for the synthesis of purines, pyrimidines, methionine and amino acid metabolism (Revuelta *et al.,* 2018).

Vitamin B_{12} – Cobalamin – Finally, vitamin B_{12} is the most complex of the B vitamins in a structural sense. It is likely the most well-studied of the B vitamins and is an essential cofactor for the enzymes methionine synthase and methylmalonyl CoA mutase amongst others (Martens *et al.,* 2002).

The structures of each of these molecules can be seen in Figure 1.5.

1.3.2 B vitamins in protist research

1.3.2.1 Establishing the importance of B vitamins

The first attempts to culture protists from the oceanic environment took place with limited success as early researchers found that environmental growth medium were not sufficient to sustain growth of the organism concerned (Beardall, 1988). Subsequent success was achieved with the advent of artificial sea water media, which included the addition of soil extract to provide a complex of inorganic and organic nutrients to the medium. Eventually the crucial components to success for many strains was the addition of the vitamins B_1, B_7 and B12 (Preisig and Andersen, 2005), although it was noted by Provasoli and Carlucci (1974) that very few species required all three of these vitamins for growth. More recent assessments of the requirements of algal species for exogenous B vitamins (auxotrophy) have been performed by Croft et al. (2006) and Tang et al. (2010). The former determined that ~50% of algae tested required exogenous B_{12} for growth, ~22% required B_1 and only ~5% B_7. This assessment, however, whilst including 306 algal species from across the EToL, was significantly biased towards green algae, which made up nearly half of the

species in the assessment. Tang et al. (2010) in a much smaller but more focused study of ~40 HAB species found an even higher prevalence of auxotrophy of ~90%. B vitamin auxotrophy is also widespread in bacterial lineages. When common oceanic bacterial clades (Cyanobacteria, Alphaproteobacteria, Gammaproteobacteria, and Bacteroidetes) were analysed they were found to have 76% of species lacking the *de novo* biosynthesis pathway for B_1, 78% for B_7 and 37% for B_{12} (Sañudo-Wilhelmy *et al.,* 2014).

Since the establishment of the importance of B vitamins for planktonic microbes there has been the hypothesis that some species influence the growth of other species by determining the availability of different B vitamins. Identifying and measuring the mechanisms by which B vitamins regulate microbial growth is challenging. Moreover, just measuring the ambient concentrations of B vitamins in the water column has proven challenging for a number of reasons such as the variability of results at different depths, seasonal variability and the vastness of the area to measure. However, most indications are that concentrations of these micronutrients are very low in large areas of the ocean (Sañudo-Wilhelmy *et al.,* 2014; 2012). The low levels of available vitamins in the water column might suggest that these micronutrients are a growth limiting factor for many species. Indeed, environmental amendment studies indicate that vitamin availability can influence microbial productivity (Koch *et al.,* 2012; Bertrand *et al.,* 2007).

1.3.2.2 The forgotten B vitamins

Although there are many studies demonstrating the importance of B_1, B_7 and B_{12} (the traditional three) in the environment and in culturing (e.g., Cruz-López and Maske 2016; Gobler *et al.,* 2007; Koch *et al.,* 2011), limited research has been done on the role of the vitamins B_2, B_3, B_5, B_6 and B_9 in microbial communities. In an effort to identify major contributors and consumers of B vitamins in the ocean Gómez-Consarnau et al. (2018) calculated ratios of transcripts for biosynthetic enzymes and vitamin dependent enzymes. In this study they included B_6 along with the traditional three and found that transcripts associated with B_6 synthesis were the highest of all the B vitamin synthesis transcripts, whilst transcripts associated with requirement of B_6 were the second highest. The levels of transcripts assigned to eukaryotes was generally low as they only measured those in picoeukaryotes (<3 µm) and did not discriminate taxonomically. They reason that while B_6 auxotrophy in marine microbes (prokaryotic and eukaryotic) is likely to be rare due to the importance of the vitamin for metabolism, the ratios of biosynthesis transcripts to dependent enzyme transcripts were rarely 1:1. This indicates that some species were consuming more than

they make, either utilising excess B_6 in the nutrient pool or evolving dependence. As mentioned above, Sañudo-Wilhelmy et al. (2012) measured the ambient concentrations of the traditional three B vitamins in areas of the ocean, however, they also measured B_2 and B_6, both of which were also very low in concentration in large areas.

This lack of research on the forgotten B vitamins leads us to two points of limitation. Firstly, the research thus far has been limited predominantly to photosynthetic protists (the bias in culturable species is even more pronounced than the bias in available sequencing data, as discussed above). Therefore, our knowledge on the culturing of heterotrophic species falls far behind that of the phototrophs. Secondly, until this point, we have limited data on the role that the other five B vitamins play in protist metabolism, auxotrophy and subsequent culture.

1.3.3 The role of B vitamins in microbial communities

The importance of B vitamins in metabolism, the frequency of auxotrophy, and their relative scarcity in the environment, makes them important currency in microbial communities (Zengler and Zaramela 2018; Sharma *et al.,* 2019). One of the most well studied models for this is B_{12}. Cobalamin is unique in the B vitamins as its large molecular structure makes it highly metabolically expensive to biosynthesise. No known eukaryote has retained or acquired the biosynthetic pathway, which is only present in certain members of the prokaryotes and requires over 20 enzymatic steps (Shelton *et al.,* 2019).

An example of the role that B_{12} can play in microbial interaction is demonstrated by the exchange of this vitamin for organic carbon by the red alga *Porphyridium purpureum* and the bacterial species *Halomonas sp.* (Croft *et al.,* 2005). Another example of vitamin cross-exchange is demonstrated in cultures of *Ostreococcus tauri* and *Dinoroseobacter shibae*. In this system the bacterial species provides the alga with vitamins B_1 and B_{12}, and in return the alga supplements the bacterial metabolism with vitamins B_3, B_7 and a precursor to B_9 (p-aminobenzoic acid) (Cooper *et al.,* 2018).

Direct interactions like this may be essential in the environment given the auxotrophic nature of many species and the relative concentrations of these vitamins (Croft, Warren, and Smith 2006; Sañudo-Wilhelmy *et al.,* 2012). Correlation and interaction analysis also demonstrate the close coupling of eukaryotic and prokaryotic microbes (Lima-Mendez *et al.,* 2015). Environmental analysis by Bertrand et al. (2015) suggests that interactions between bacteria and diatoms in the Southern Ocean plays a role in the limitation of phytoplankton growth by the vitamin B_{12}. Several other experiments have demonstrated B vitamins limit the growth of phytoplankton in the

environment. For example, when iron and B_{12} were added to environmental samples from the Ross Sea chlorophyll concentrations were observed to increase over time. This suggests that these nutrients are colimiting growth of phytoplankton in this area of the ocean (Bertrand *et al.,* 2007). Koch et al. (2012) showed that primary production correlates significantly with the rate of B_{12} uptake and that bacteria may be a sink for the concentrations of B vitamins in the ocean, limiting growth of eukaryotic plankton. It has also been shown that the availability of B_{12} in the ocean may affect different size categories of plankton differently. In vitamin amendment studies large phytoplankton have displayed a strong response to the addition of B_{12} whereas little to no effect was seen in the smaller size fractions (Sañudo-Wilhelmy *et al.,* 2006).

1.4 Structure and aims of this thesis

This thesis aims to explore the potential role of the full spectrum of B vitamins as growth factors for "known-unknown" marine microbial species, in particular focusing on taxa belonging to the stramenopile lineage. To do this, three approaches were taken. The results of these approaches form the three proceeding results chapters.

(1) Environmental sequencing and vitamin amendments of the forgotten B vitamins to probe the possibility that these vitamins are limiting for the growth of the microbial community, in particular for MASTs. The hypothesis for the experiment in this chapter is that given an addition of the forgotten B vitamins a change will occur in the microbial community structure.

(2) Analysis of B vitamin biosynthesis pathways in genomic data of stramenopile species to establish the capabilities of these organisms to perform *de novo* synthesise of each B vitamin or conversely identify auxotrophies. Utilising data from a large number of organisms will also allow the identification of patterns in auxotrophy across the stramenopile group and thus identify tracks in B vitamin biosynthesis pathway evolution. The hypothesis for the analysis in this chapter is that instances of auxotrophy can be identified by incomplete biosynthesis pathways and that these instances will be common to related organisms resulting in identifiable patterns of auxotrophy across the stramenopile group.

(3) Finally, utilising the only MAST3 species in culture as a model for a local microbial community. *Incisomonas marina* is a heterotrophic stramenopile that is maintained with a consortium of bacteria. Metagenomic sequencing and analysis will be utilised to investigate the potential interactions of this closed community, in particular with regards to B vitamin exchange. The

hypothesis for the work in this chapter is that the individuals in the *I. marina* community provide essential support to each other by active or passive interaction.

Following these lines of inquiry, I hope to further our understanding of the role B vitamins play in stramenopile communities and inform future culturing efforts of these understudied lineages.

2 – Assessing the effect of B vitamin amendments to natural assemblages of oceanic microbes

2.1 Introduction

As discussed in Chapter 1 Section 1.1.4 the oceans play hosts some of the most important species on earth. The microbial community of the global oceans drives biogeochemical cycles and is responsible for around 50% of primary productivity, the majority of which is performed by single-celled phytoplankton (Falkowski *et al.,* 2008). However, the ocean is a dilute environment. The resources that this microbial life relies on, including both macro and micronutrients, are often at a concentration that is limiting to both the activity and abundance of the phytoplankton (Moore *et al.,* 2013; Moore 2016). This is particularly true in the photic zone, a ~200 m layer of sub-surface water that is penetrated by the light of the sun, which hosts the photoautotrophic assemblage of oceanic microbes along with many of their heterotrophic counterparts. There are two significant nutrient limitation patterns that have been revealed over the recent past. One is limitation of iron in areas that are enhanced with subsurface nutrient supply, known as high nutrient low chlorophyll regions (Moore et al., 2001; Moore et al., 2013) such as the Southern Ocean, subarctic Northeast Pacific, and equatorial Pacific, where despite high levels of nutrients the phytoplankton community is not abundant (determined by chlorophyll concentration as measured by satellite imagery e.g. https://oceancolor.gsfc.nasa.gov). Outside of these areas, productivity is likely limited by the availability of inorganic nitrogen (Bristow et al., 2017). However, nutrient limitation is a complex state that depends not only on the concentration of a nutrient but also on the composition of the local microbial community and the relative bioavailability of the nutrient. Indeed, in many cases two or more nutrients will be co-limiting to certain members of the community (Browning *et al.,* 2017; Moore *et al.,* 2013). As well as the predominant limiting nutrients (N, P and Fe) studies have demonstrated that many micronutrients can be limiting in local environments (Koch and Trimborn, 2019).

Despite their importance, it is challenging to conduct research on oceanic microbial communities *in situ.* The vastness of the environment and the continuous advection and mixing of the water in which the community resides makes direct measurements of the same organisms very difficult, particularly over a period of time. Mesocosms and controlled laboratory studies of samples taken directly from the environment have been employed to overcome these issues (Steele, 2013). These types of experiment have been used to investigate a wide variety of oceanic processes and their

impact on the consortia of microbes in the sample, including the response to salinity changes, rising CO_2 concentrations and temperature changes (e.g., Allen *et al.,* 2020; Kaartokallio , Laamanen, and Sivonen 2005). Indeed, this approach has been employed in previous efforts to characterise the state of B vitamin limitation on the growth of microbial communities, in particular for B_{12}. In bottle incubations of samples from the Ross Sea, where it had previously been noted that iron was limiting primary production, the addition of B_{12} and iron saw greater increases in chlorophyll fluorescence as a proxy for phytoplankton abundance than with the addition of iron alone (Bertrand *et al.,* 2007). Similar results have been demonstrated for other areas of the ocean, including the South Atlantic gyre, the Long Island Sound and the Southern Ocean (e.g., Browning *et al.,* 2017; Koch *et al.,* 2012; Bertrand *et al.,* 2015). Through this type of experimental set up, the seasonal effects of nutrient limitation have been explored, demonstrating that the limit on growth by the concentration of B_{12} is only present at certain times of the year (Bertrand *et al.,* 2011). In addition, the co-limitation effect has also been demonstrated. Koch et al. (2011) showed that in the Gulf of Alaska amendments of B_{12} and nitrogen or iron to samples resulted in enhanced algal biomass and growth rates compared to samples with no amendments in nutrients, particularly among larger phytoplankton groups.

There is an ever-growing appreciation for the "microbial dark-matter" of the ocean, both prokaryotic and eukaryotic (del Campo *et al.,* 2014; Sunagawa *et al.,* 2015; Whitman, Coleman, and Wiebe 1998; Massana *et al.,* 2014). From advancements in sequencing technology and global sampling efforts we now have a greater understanding of the number of species that are likely to be living in the ocean which we currently have no way of growing in the conditions of a laboratory (de Vargas *et al.,* 2015; Massana *et al.,* 2014). The importance or vitamins B_1, B_7 and B_{12} to many cultured algal groups is well established (Croft *et al.,* 2006). However, little is known about the role the other five B vitamin might play in the growth of the currently unculturable species. To explore the potential role of those B vitamins not traditionally used in microbial culture (vitamins B_2, B_3, B_5, B_6 and B_9; see Section 1.3.2.2), in the growth of the currently uncultured diversity of microbes from the marine environment, an experiment was conducted that applied these micronutrients to natural samples and assessed the subsequent changes in the community.

2.1.1 Objectives

The aims of this experiment were to establish whether the concentration of the vitamins B_2, B_3, B_5, B_6 and B_9 were limiting the growth of any species in the environment, in particular those currently unculturable. This could then lead to a greater understanding of why these organisms are not yet in

stable culture. In addition, to assess whether these B vitamins are co-limited with macro nutrients, N, P and S were also added. These were used due to their inclusion in F/2 media typically used for growth of many stramenopile species (Guillard and Ryther, 1962), as this is the group of particular interest for this thesis. The sampling station E1, located in the English Channel was used. A wide diversity of microbes have been demonstrated to occur in the English Channel, including stramenopile species (Taylor and Cunliffe, 2014).

This experiment was intended to act as a pilot study and inform later repeats of the same concept. However, as a result of the global COVID-19 pandemic it was not possible to achieve what was planned in full for this section of the project. Had I been able to repeat this experiment various changes would have been made. For example, a longer incubation time might have allowed greater changes in the community composition to occur, in particular in the eukaryotic population. Another possible change might be the concentrations of the vitamin amendments, as it is possible that not enough was added to see the nutrient limitation overcome. Additionally, rather than N, P and S the macronutrient Fe could be a co-limiter. A different sampling time could be tested as it may be that this season (mid-late summer) is not one in which the microbial community experiences growth limitation due to nutrient concentration. Additional experiments would also have provided an opportunity to assess the effects of amendments made with the traditional three B vitamins (B_1, B_3 and B_{12}). With the exception of B_{12}, which was used as a comparison condition, these B vitamins were omitted from the pilot study due to the limitations in the size of the initial experiment (limited number of incubation bottles). Given greater capacity or multiple follow up experiments, additional controls would also have been used such as all B vitamins and all B vitamins + nutrients. Despite this, below I present the full results from this initial study.

This experiment was made possible by a studentship with the Marine Biological Association (MBA). Founded in 1884 the MBA has a long history of pioneering oceanic research. The partnership with this institute enabled this experiment to take place by providing access to the sampling site, the facilities to carry out an experiment of this scale and provided a wealth of experience and knowledge in the handling of environmental samples and mesocosm experiments.

2.1.2 Experimental design

A sample of sea water from the sampling station E1 in the English Channel was collected in August 2019 from the surface layer of water. Samples were passed through a 200 µm mesh to remove multicellular organisms and large protists. The water was subsequently divided in to thirty

individual one litre bottles. These bottles were split into six different conditions for a 24-hour incubation. The first condition was a control where no nutrient amendment was made to the bottles. The other conditions each had a different vitamin and/or nutrient amendment added to the bottle. The B vitamin mix was made up of the 'forgotten' B vitamins (B_2, B_3, B_5, B_6 and B_9 at concentrations: 0.26 µg/L, 98.5 µg/L, 101.3 µg/L, 102.8 µg/L and 1.77 µg/L respectively (Henson *et al.,* 2016)) and the nutrients added were the macronutrients N, P and S at concentrations above the measured levels typical of the time of year at E1 (S – 0.5 g/L (Na_2SiO_3), N – 1.25 g/L ($NaNO_3 9H_2O$), P – 0.08 g/L ($NaH_2PO_4 H_2O$)) (Woodward and Harris 2014; Henson *et al.,* 2016). These macronutrients were included to investigate any co-limitations that may be occurring. The conditions used were: + vitamin B_{12}, + B_{12} + nutrients, + B vitamin mix, + B vitamin mix + nutrients and + nutrients. After the 24-hour incubation these bottles were processed by filtering onto a 0.2 µm nitrocellulose filter which catches all microbial life from the sample. These filters were then snap-frozen before

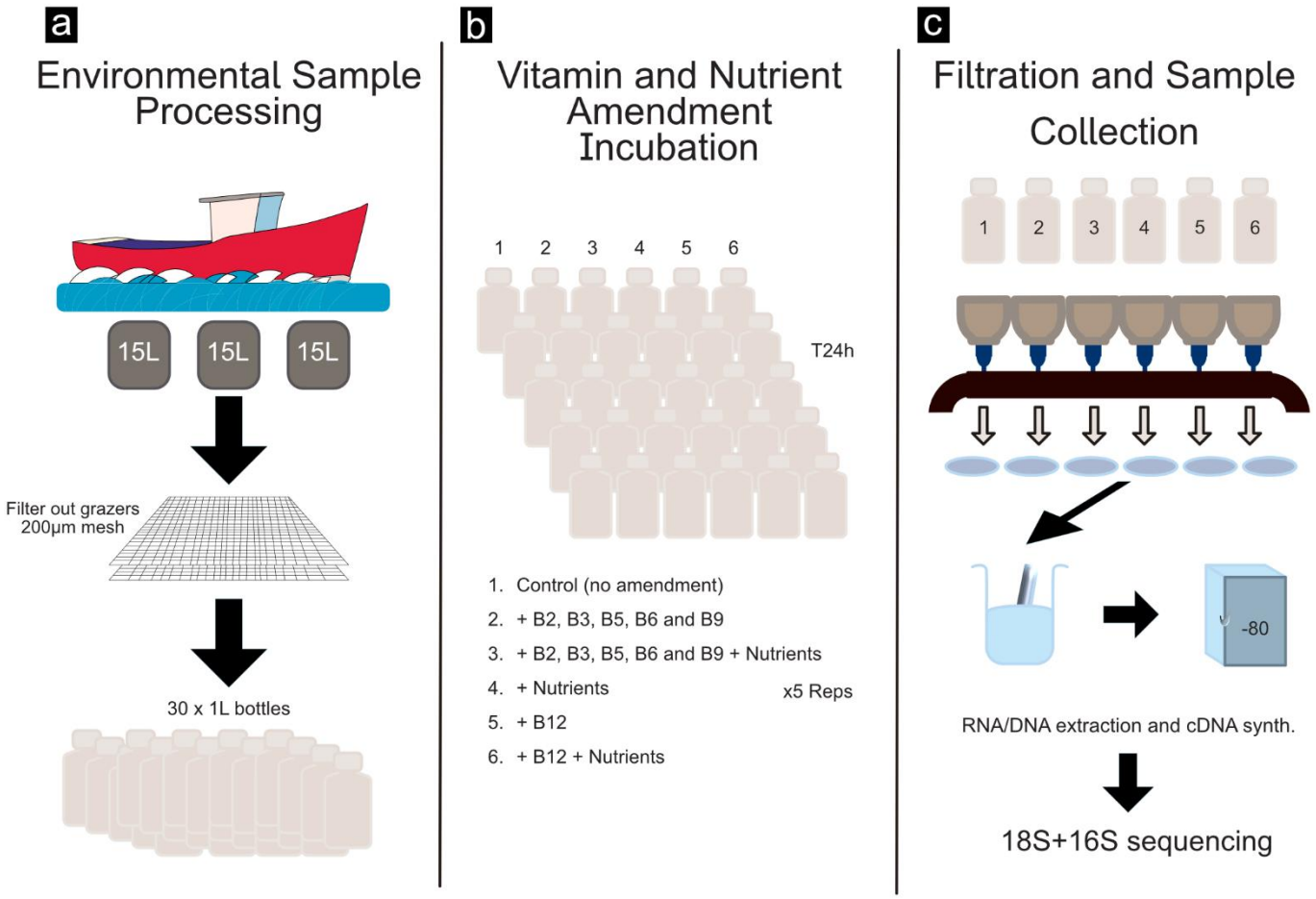


Figure 2.1 *Schematic of the experimental design of the nutrient amendment mesocosm experiment presented in this chapter. (a) Environmental samples were taken from the E1 sampling site (carried out by the MBA) in August 2019 from the surface. Samples were filtered for larger organisms (>200µm) and siphoned into 1L bottles. (b) Samples were then amended with varying nutrient regimes and incubated for 24 hours before (c) being filtered for microbial life on a 0.2µm nitrocellulose filter. From the filter, total DNA and RNA was extracted. Subsequently 16S and 18S amplicon sequencing was performed.*

processing for DNA and RNA extraction. A schematic of the experimental set up can be seen in Figure 2.1.

The effect these amendments had on the composition of the microbial community was measured using an amplicon sequencing approach. From the filter from each bottle DNA and RNA was extracted, this was carried out using a joint DNA and RNA extraction method (see Section 2.2.3). RNA was subsequently converted to cDNA using reverse transcription and these samples were sequenced along with the DNA samples. To detect members of the bacterial community that were present, primers were used to sequence the 16S ribosomal RNA genes (region V6-V8) present in the samples. For the eukaryotic contingent, primers for the 18S rRNA gene were used for sequencing. The sequenced samples were then put through an amplicon analysis pipeline to determine the presence and quantity of amplicon sequence variants (ASVs are analogous to the commonly used Operational Taxonomic Unit – OTU but overcome many of the limitations of OTUs) (Callahan *et al.,* 2016; Callahan *et al.,* 2016). rRNA has a short half-life (Blazewicz *et al.,* 2013), therefore the presence of transcripts indicates recent transcriptional activity. Levels of rRNA coupled with those for rDNA allows for a ratio of these two molecules to be calculated, providing an indication of the transcriptional activity relative to the abundances of the taxa, and thus a measure of relative activity (Bowsher *et al.,* 2019).

More general measures of the effect of the treatment conditions on the microbial community were carried out by measuring changes in alpha diversity and beta diversity (Whittaker, 1972). Alpha diversity refers to the number of species, i.e., species richness, and the distribution of the abundancies of the species in a community of a single site or sample. Shannon's index (Ortiz-Burgos 2016) can be used as a measure of alpha diversity. Beta diversity on the other hand allows comparisons of community composition across sites or samples in an experiment. More accurately it is a measurement of the dissimilarity in community composition and here it is measured using Bray-Curtis dissimilarity (Bray and Curtis, 1957).

2.2 Material and Methods

2.2.1 – Environmental sampling and processing

Environmental samples were taken by the Sepia Research Vessel of the Marine Biological Association at sampling station E1 in the English Channel. A sample of ~45 L was taken and processed by passing through a 200 µm mesh to remove large ciliates and multicellular organisms.

Samples were then decanted into 30 individual 1 L bottles for the experiment (i.e., for 6 treatments, 5 replicates).

2.2.2 – Vitamin amendments and incubation

Each bottle was amended with varying amendment conditions. Six different conditions were used:

1. Control (no amendment)
2. + Vitamin B_{12}
3. + B_{12} + nutrients
4. + B vitamin mix
5. + B vitamin mix + nutrients
6. + Nutrients

The B vitamin mix consisted of the following B vitamins at corresponding concentrations: B_2 (0.26 µg/L), B_3 (98.5 µg/L), B_5 (101.3 µg/L), B_6 (102.8 µg/L) and B_9 (1.77 µg/L). Vitamin B_{12} was at 0.95 µg/L. Nutrient additions were achieved by adding 1/60th of the concentration of f/2 stock nutrients (Bertrand *et al.,* 2007). The concentrations of the nutrients were as follows: S – 0.5 g/L (Na_2SiO_3), N – 1.25 g/L ($NaNO_3$ $9H_2O$), P – 0.08 g/L (NaH_2PO_4 H_2O). Bottles were incubated for 24 hours at 15 degrees Celsius in a light cycle of 16h light 8h dark.

2.2.3 – Sample processing, nucleic acid extraction and cDNA synthesis

Incubated bottles were each filtered onto individual 0.2 µm nitrocellulose filters and snap frozen on liquid nitrogen before being stored at -80 degrees prior to nucleic acid extraction. DNA and RNA were extracted using the ZymoBIOMICS DNA/RNA Miniprep kit (Cat. No.: R2002) following the joint extraction method in the manufacturer's instructions. RNA was subsequently converted into cDNA using Superscript III First-strand Synthesis System (Invitrogen cat. No. 18080051) according to manufacturer's instructions following the instructions for use with random hexamers.

2.2.4 – DNA and cDNA quality control and sequencing

Concentrations and quality of DNA extractions and cDNA samples were checked by nanodrop. Samples were then sent to the Integrated Microbiome Resource (IMR) for Miseq amplicon sequencing. The 18S rRNA gene was targeted to identify the eukaryotic community and the 16S rRNA gene for the bacterial community. The following primers were used:

18S (V4): Forward – CYGCGGTAATTCCAGCTC, Reverse – AYGGTATCTRATCRTCTTYG

16S (V6-V8): Forward – ACGCGHNRAACCTTACC, Reverse – ACGGGCRGTGWGTRCAA

Full protocol for sequencing and library prep can be found at

https://imr.bio/protocols.html#libAmp

2.2.5 – Sequence processing and statistical analysis

18S and 16S sequences were processed independently of each other following a pipeline initially developed at the MBA and subsequently customised for this work. The code for both 18S and 16S analysis is written in R (R Core Team, 2021) and can be found in Appendix Section 1.1. The pipeline uses DADA2 (v1.16.0) and phyloseq (v1.36.0) for the sequence processing and analysis respectively (Callahan *et al.,* 2016; McMurdie and Holmes 2013). Downstream statistical analysis uses the following packages: alpha diversity – phyloseq (estimate_richness), beta diversity –vegan (v2.5-7 - vegdist, adonis, betadisper), abundance plots – phyloseq (plot_bar), graphical plotting – ggplot2 and base R for general statistics such as aov and TukeyHSD.

Briefly, sequences are assessed for read quality before being trimmed to 270 and 220 for forward and reverse reads respectively. Subsequently primer sequences were also removed. DADA2 relies on error rate estimation to resolve amplicon sequence variants (ASVs). Next the error rates are calculated. Sequences are then temporarily dereplicated to reduce computational demand whilst resolving ASVs. The DADA2 algorithm is then applied to resolve ASVs. Following this, reads are merged to achieve full length sequences for the target amplicon region. A count is then performed to assess the number of instances of each ASV in the data set. Next chimeric reads are identified and removed. Taxonomic assignment is performed on the ASVs and metadata is aligned with the reads. Finally, a rarefaction step is performed to normalise the data by subsampling to an even coverage depth. Downstream analysis is then carried out using the phyloseq package (Appendix Section 1.2).

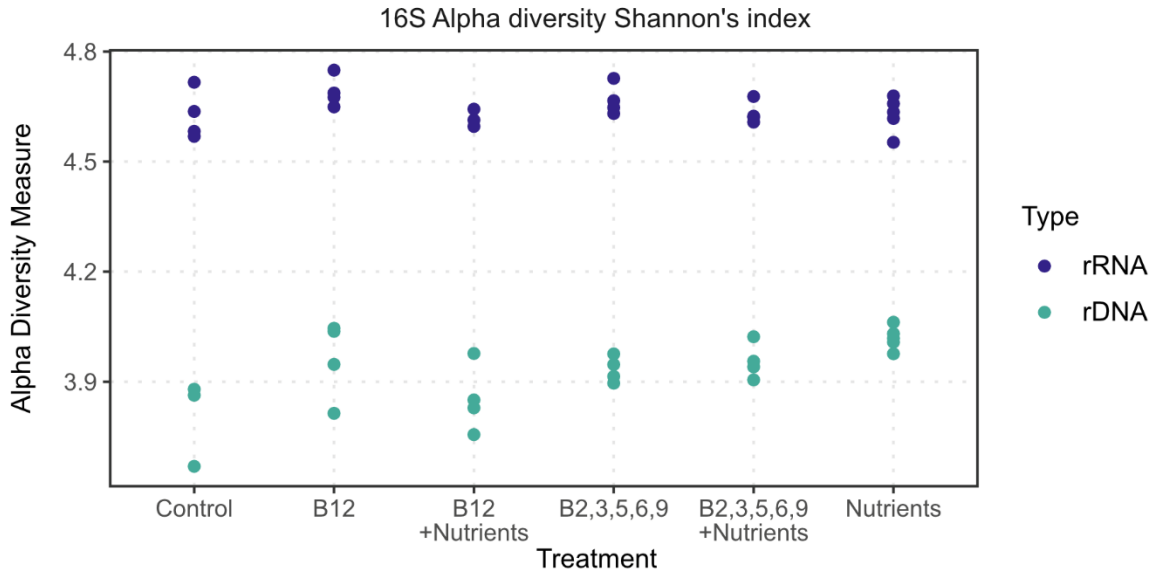
2.3 Results

2.3.1 Bacterial community composition

The composition of the community in each sample was measured in a variety of ways to build a picture that reflects as closely as possible the true nature of the sample. The measures used include assessing alpha diversity using Shannon's index, beta diversity visualised using principal coordinates analysis (PCoA – this is a method to measure and analyse dissimilarity in data based on a distance matrix, this is opposed to a PCA which is used to identify correlations in data – see https://www.sequentix.de/gelquest/help/principal_coordinates_analysis.htm)(Gower, 1967), relative abundancies of species present in each sample and an rRNA/rDNA (truly cDNA/DNA) ratio as a proxy measurement for species transcriptional activity.

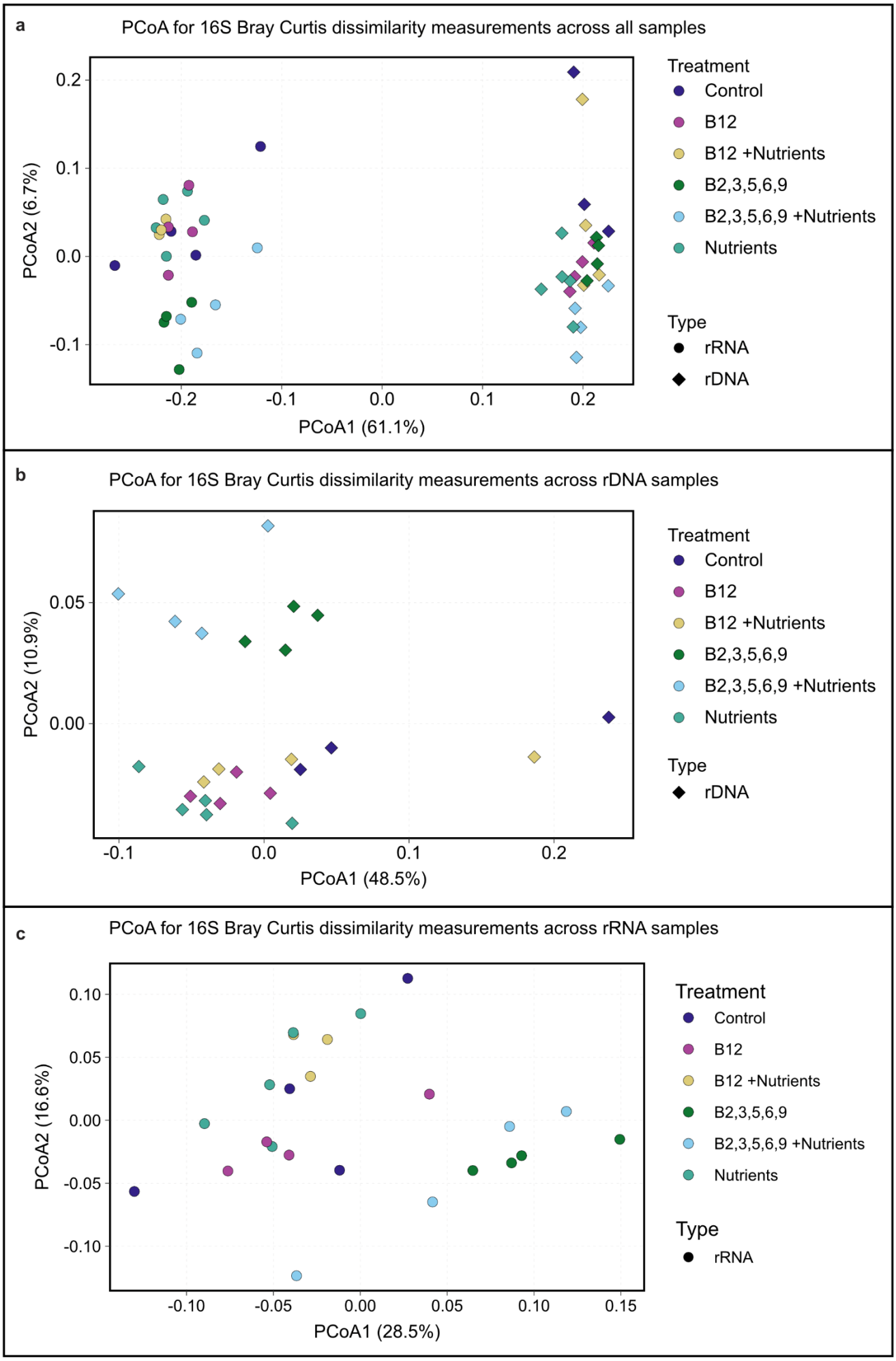
2.3.1.1 – 16S Alpha diversity

The Shannon's index measurement for each sample is plotted in Figure 2.2. The measurements of alpha diversity reveal a significant difference in the Shannon's index between the type of sample measured ($p < 2 \times 10^{-16}$ rRNA vs rDNA). Significance was tested for using a two-way ANOVA, which also measured the effect of treatment. The significant relationship between type (rRNA and rDNA) and alpha diversity is expected and is regularly seen in this type of study (e.g., Klein et al., 2016). There was also found to be a significant interaction between the Shannon's index and the treatment of the samples, albeit to a lesser extent than for the type of sample ($p = 0.0145$). To further analyse the significance of the results a Tukey's test was implemented. This tests for significant relationships in a pairwise fashion, in this case between treatments and between type. The Tukey's test revealed a significant relationship between type and Shannon's index, as previously identified by the two-way ANOVA, but when each treatment was compared to each other with the Tukey's test, no significant relationships were found, in contrast to the ANOVA.



▲ ***Figure 2.2. Alpha diversity across 16S (prokaryotic) rRNA and rDNA samples measured by Shannon's index.*** *Alpha (α) diversity measures the community diversity at individual sites. Here, α diversity was measured using Shannon's index which accounts for the total number of species in a sample and the distribution of abundance of the species at that site. rRNA samples are represented in purple and rDNA samples are represented in teal. Each point represents an individual sample. n=5 for each condition. N.B. some samples failed and are therefore not represented.*

► ***Figure 2.3 a-c. Beta diversity across 16S (prokaryotic) rRNA and rDNA samples measured by Bray-Curtis Dissimilarity and visualised by PCoA.*** *Beta diversity measures the differences in community composition between samples. Here it has been calculated using Bray Curtis Dissimilarity and visualised using Principal Coordinates Analysis (PCoA).* ***(a)*** *PCoA of all 16S samples. rDNA samples are represented as diamonds and rRNA as circles. Both are coloured according to treatment.* ***(b)*** *PCoA of 16S rDNA samples only.* ***(c)*** *PCoA of 16S rRNA samples only. Each point represents an individual sample. n=5 for each condition. N.B. some samples failed and are therefore not represented.*


a
PCoA for 16S Bray Curtis dissimilarity measurements across all samples
PCoA2 (6.7%)
PCoA1 (61.1%)
0.2
0.1
0.0
-0.1
-0.2
-0.1
0.0
0.1
0.2
Treatment
Control
B12
B12 +Nutrients
B2,3,5,6,9
B2,3,5,6,9 +Nutrients
Nutrients
Type
rRNA
rDNA
b
PCoA for 16S Bray Curtis dissimilarity measurements across rDNA samples
PCoA2 (10.9%)
PCoA1 (48.5%)
0.05
0.00
-0.1
0.0
0.1
0.2
Treatment
Control
B12
B12 +Nutrients
B2,3,5,6,9
B2,3,5,6,9 +Nutrients
Nutrients
Type
rDNA
c
PCoA for 16S Bray Curtis dissimilarity measurements across rRNA samples
PCoA2 (16.6%)
PCoA1 (28.5%)
0.10
0.05
0.00
-0.05
-0.10
-0.10
-0.05
0.00
0.05
0.10
0.15
Treatment
Control
B12
B12 +Nutrients
B2,3,5,6,9
B2,3,5,6,9 +Nutrients
Nutrients
Type
rRNA

2.3.1.2 – 16S Beta diversity

The prokaryotic community composition was further evaluated using a measure of beta diversity. Bray-Curtis dissimilarity index was used to measure the differences in community composition across treatments in the experiment. A PCoA was used to visualise the beta diversity measurements (Figure 2.3 a). The plot largely forms two distinct groups, one consisting of the rDNA samples and the other of the rRNA samples. The split can be observed across PCoA1 which accounts for 61.1% of the variance in the experiment, with 6.7% of the variance in beta diversity accounted for by PCoA2. Two statistical tests were used to evaluate the differences between beta diversity and the samples. These were a PERMANOVA and homogeneity of multivariate dispersion analysis. The first of these tests for significant overall difference in beta diversity between conditions. It tests for differences in both centroid position (the middle position of a circle encompassing the points of a group) of points in a condition and dispersion of points in a condition. The second tests only for significant differences in the dispersion of points in a condition.

As would be expected, the PERMANOVA revealed that the community composition was significantly affected by type. The community was also affected significantly by treatment but there was no significant interaction between these factors. The homogeneity of multivariate dispersion test was not significant, which shows that the significant p-value calculated during the PERMANOVA for treatment was not caused by differences in dispersion. The significant result is therefore likely to be explained by the centroid position of each sample. Indeed, when looking at the groups on the PCoA plots in Figures 2.3 a-c these differences can be observed.

To better understand variation in community composition between treatments, a PCoA plot of samples from individual types was made (Figure 2.3 b-c) and a pairwise PERMANOVA was set up. In this analysis each treatment-type combination was treated as individual, for example, rDNA+B_{12} and rRNA+B_{12} were handled separately. In each test where a rDNA type was tested against a rRNA type a significant difference was expected to be indicated by the p-value of the test. Indeed, this was the case for all except one test (rRNA B_{12} + nutrients Vs rDNA Control $p > 0.05$). Of the 30 tests between like types of samples (i.e., rRNA vs rRNA or rDNA vs rDNA) 16 were found to have a p-value below 0.05. Full pairwise PERMANOVA p-value results can be found in Appendix Section 1.3. Of the test assessing differences between the control treatment and each other treatment, both the rDNA and rRNA comparisons for the B vitamin mix had significant differences in community composition (*p= 0.034* and *p= 0.03* respectively). Other cases where a significant p-value was achieved in the

same treatment vs treatment test and in both data types were B_{12} Vs B vitamin mix, B_{12} Vs B vitamin mix + nutrients, B_{12} + nutrients Vs B vitamin mix + nutrients, B vitamin mix Vs nutrients and B vitamin mix + nutrients Vs nutrients.

2.3.2 Bacterial Taxa abundances and activities

2.3.2.1 – 16S Relative abundance measures

To gain a broad appreciation of the composition of the sample, abundances were evaluated by grouping taxonomically assigned reads at the rank of Phylum (further classified by Class for Proteobacteria). In both rRNA and rDNA types and across all treatments, samples were dominated by Bacteroidetes (rDNA: 41%, rRNA: 23%), Alphaproteobacteria (rDNA: 42%, rRNA: 35%) and Gammaproteobacteria (rDNA: 14%, rRNA: 35%) (Figure 2.4). The abundances of Gammaproteobacteria, Cyanobacteria and Verrucomicrobia were higher in the rRNA samples than in the rDNA samples, implying that they were more transcriptionally active than other species present. Conversely, Bacteroidetes had a greater relative abundance in the rDNA samples. The average relative abundances of the dominant taxa in each type of sample can be seen in Table 2.1.

Table 2.1. Relative abundances of dominant prokaryotic taxa in the two different types of samples. Standard Deviation in brackets.

	rDNA (%)	rRNA (%)
Bacteroidetes	40.8 (4.3)	22.5 (2.2)
Alphaproteobacteria	41.9 (5.9)	34.6 (4.1)
Gammaproteobacteria	14.4 (1.8)	34.9 (3.9)
Cyanobacteria	0.9 (0.2)	4.0 (1.5)
Verrucomicrobia	1.7 (0.3)	3.2 (0.5)
Lentisphaerae	0.1 (0.0)	0.3 (0.1)
Planctomycetes	0.0 (0.0)	0.2 (0.1)
Deltaproteobacteria	0.0 (0.0)	0.2 (0.0)

2.3.2.2 – 16S rRNA/rDNA ratios

To get an understanding of the community at a higher resolution, the data was evaluated at the rank of genus. The most abundant genus was that of *Aurantivirga,* a flavobacterium of the Bacteroidetes commonly found in the marine environment (Song *et al.,* 2015; Khan, Nakagawa, and Harayama 2008). The second most abundant was a genus of the SAR11 clade (Clade Ia), often reported as the most abundant microbes in the ocean (Giovannoni, 2017). The third was another cosmopolitan costal bacterial genus: the OM60(NOR5) clade of the Gammaproteobacteria (Yan *et al.,* 2009). The taxonomic classifications of these top genera demonstrate that the community is predominantly made up of species that are consistent with other investigations (Gilbert *et al.,* 2009; Shaw *et al.,* 2008).

To evaluate the effect of the treatments on the community a measure of transcriptional activity was calculated by taking a ratio of 16S rRNA to rDNA reads (16S ratio). This measure gives a proxy for growth. The total number of genera detected was 184 and the full list of these can be found in Appendix Section 1.4. The 16S ratios were plotted for the top 20 most abundant genera to gain further understanding of the community composition as well as activity. These plots can be seen in Figures 2.5 and 2.6 and the full taxonomic classifications of these genera can be found in Table 2.2.

The 16S ratios (Figure 2.5) show that in the majority of these most abundant genera there was a consistent level of activity across each of the treatments, suggesting that the treatment was not a factor in the ratio level. One group which is an exception to this is the *Litorivivens* who display a high 16S ratio in the non-treated control group (in two of the three samples) and high ratio in the B vitamin mix and nutrient treatments but a lower ratio in the two treatments that include a B_{12}

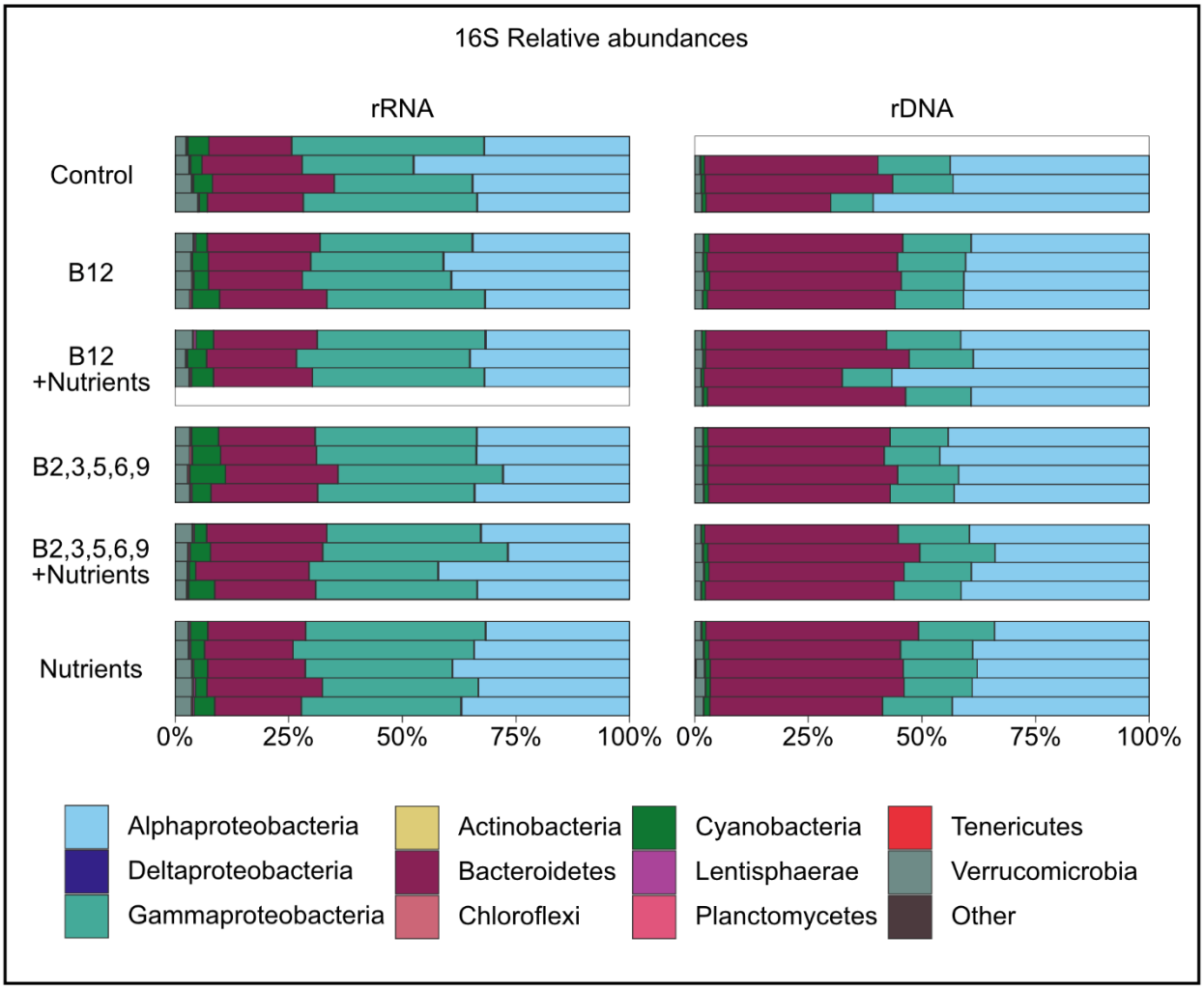


Figure 2.4 Relative abundances of 16S samples (rRNA and rDNA) at the family rank. *Percentage of reads assigned to each Family rank (further classified to the class level for Proteobacteria) in each sample. The top 11 most abundant families/classes are represented with the following families grouped into "other": Epsilonbacteraeota, Euryarchaeota, Firmicutes, Acidobacteria, Hydrogenedentes, Chlamydiae and Fibrobacteres.*

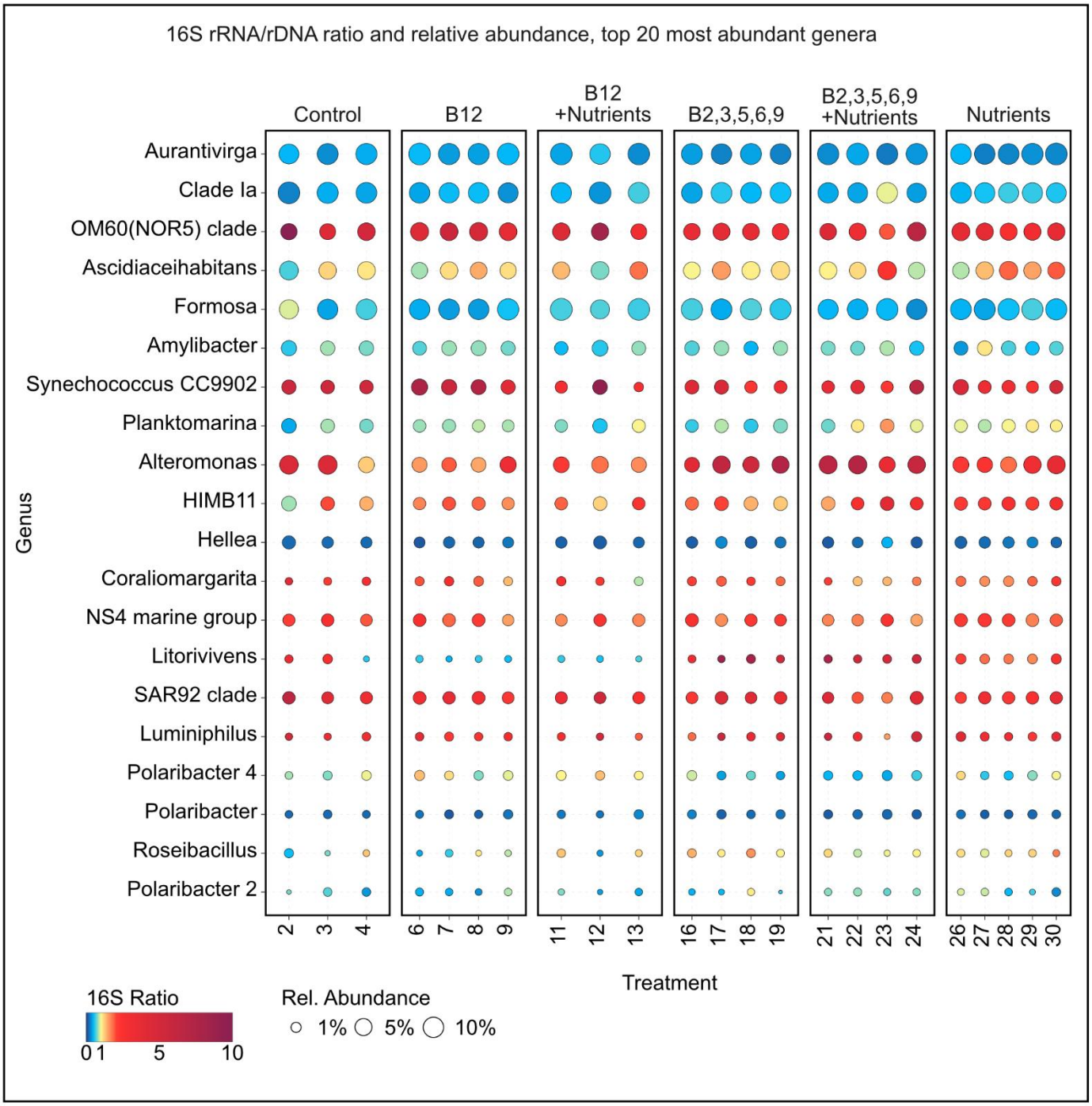
Table 2.2. Full taxonomic classifications of the top 20 genera in the 16S amplicon data. Taxonomic assignment performed using the SILVA database.

Abundance rank #	ASV #	Phylum	Class	Order	Family	Genus
1	**ASV 1**	*Bacteroidetes*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Aurantivirga*
2	**ASV 2**	*Proteobacteria*	*Alphaproteobacteria*	*SAR11_clade*	*Clade_I*	*Clade_Ia*
3	**ASV 3**	*Proteobacteria*	*Gammaproteobacteria*	*Cellvibrionales*	*Halieaceae*	*OM60(NOR5)_clade*
4	**ASV 4**	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Ascidiaceihabitans*
5	**ASV 5**	*Bacteroidetes*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Formosa*
6	**ASV 6**	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Amylibacter*
7	**ASV 11**	*Cyanobacteria*	*Oxyphotobacteria*	*Synechococcales*	*Cyanobiaceae*	*Synechococcus_CC9902*
8	**ASV 12**	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Planktomarina*
9	**ASV 13**	*Proteobacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas*
10	**ASV 15**	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*HIMB11*
11	**ASV 22**	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Hyphomonadaceae*	*Hellea*
12	**ASV 24**	*Verrucomicrobia*	*Verrucomicrobiae*	*Opitutales*	*Puniceicoccaceae*	*Coraliomargarita*
13	**ASV 25**	*Bacteroidetes*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*NS4_marine_group*
14	**ASV 30**	*Proteobacteria*	*Gammaproteobacteria*	*Cellvibrionales*	*Spongiibacteraceae*	*Litorivivens*
15	**ASV 32**	*Proteobacteria*	*Gammaproteobacteria*	*Cellvibrionales*	*Porticoccaceae*	*SAR92_clade*
16	**ASV 34**	*Proteobacteria*	*Gammaproteobacteria*	*Cellvibrionales*	*Halieaceae*	*Luminiphilus*
17	**ASV 38**	*Bacteroidetes*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Polaribacter_4*
18	**ASV 43**	*Bacteroidetes*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Polaribacter*
19	**ASV 44**	*Verrucomicrobia*	*Verrucomicrobiae*	*Verrucomicrobiales*	*Rubritaleaceae*	*Roseibacillus*
20	**ASV 46**	*Bacteroidetes*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Polaribacter_2*

amendment. *Roseibacillus* also shows a slight increase in the 16S ratio in the treatments compared to the control although not significantly.

Figure 2.6 plots the raw ratios as a box plot for each treatment and genus, clearly demonstrating the differences in activity of *Litorivivens* in the control and the treatments, in particular in the B vitamin mix treatments (significant p-value for control vs +B vitamin mix +Nutrients). A similar trend is observed for the genus *Alteromonas*. *Polaribacter* also has a significantly lower level of activity in the +B vitamin mix +Nutrient condition relative to the control. *Synechococcus* CC9902 had significantly lower activity in the +B vitamin mix condition and the +Nutrient conditions. However, the majority of these twenty most abundant genera have a consistent 16S ratio across treatments.

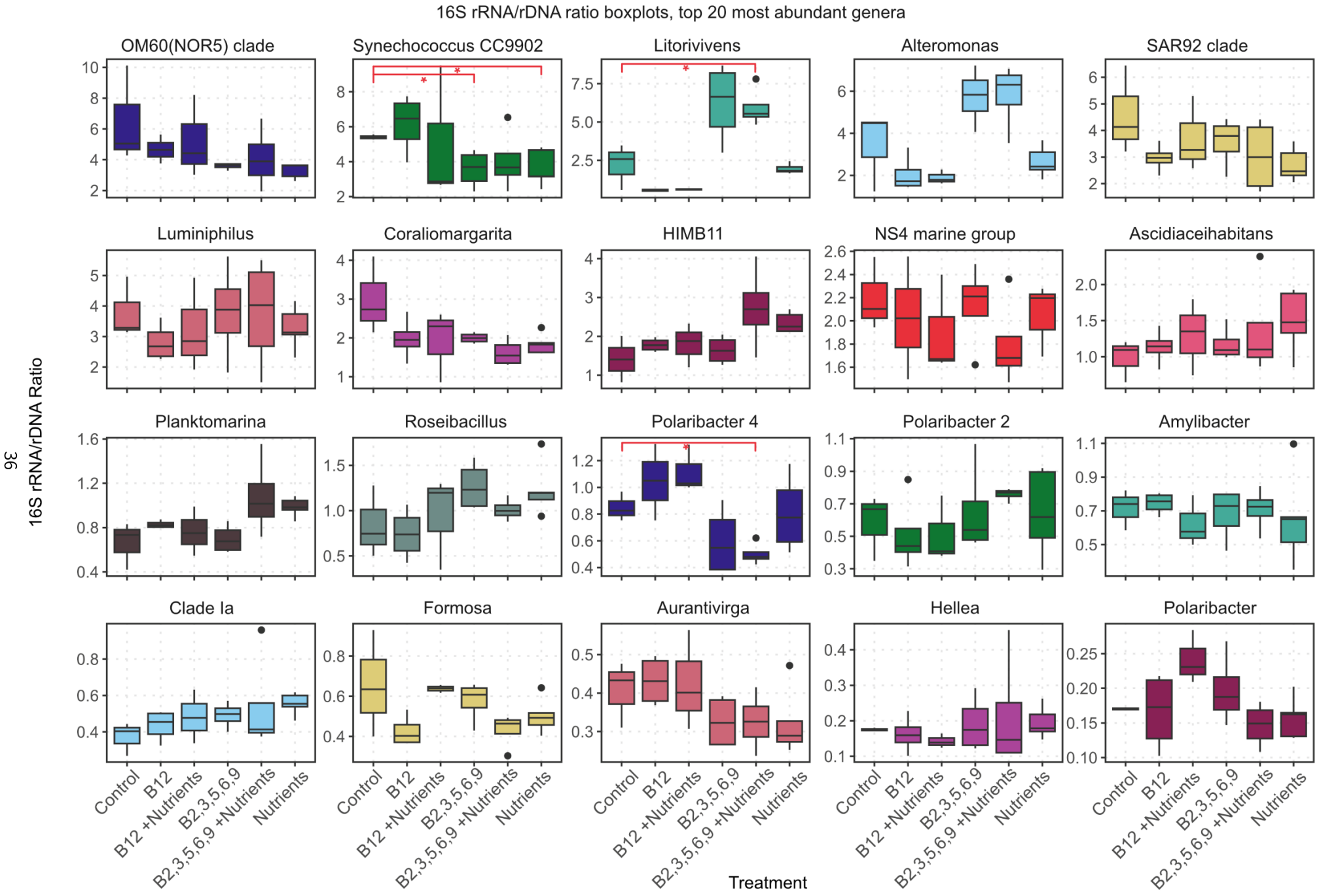
To get a more general appreciation of the activity of the bacterial groups in the experiment, the ratio data was re-plotted after grouping at the taxonomic rank of family. This plot can be seen in Figure 2.7. At this lower resolution a few trends are observed including the family of Cyanobiaceae being less active in Nutrient amended conditions and vitamin amended conditions compared to the control condition, this is supported by t-test significance values (<0.05). The family of Alteromonadaceae appears to be more active in the conditions +B vitamin mix and +B vitamin mix + nutrients when compared to the control. This corroborates the results seen at the genus level which showed *Alteromonas* as more active in the B vitamin amended conditions. Finally, the other apparent pattern can be seen in the family of Pseudoalteromonadaceae who appear to have significantly lower activity in conditions where the B vitamin mix and nutrients were applied, compared to no amendment. A full list of bacterial families can be found in Appendix Section 1.5.

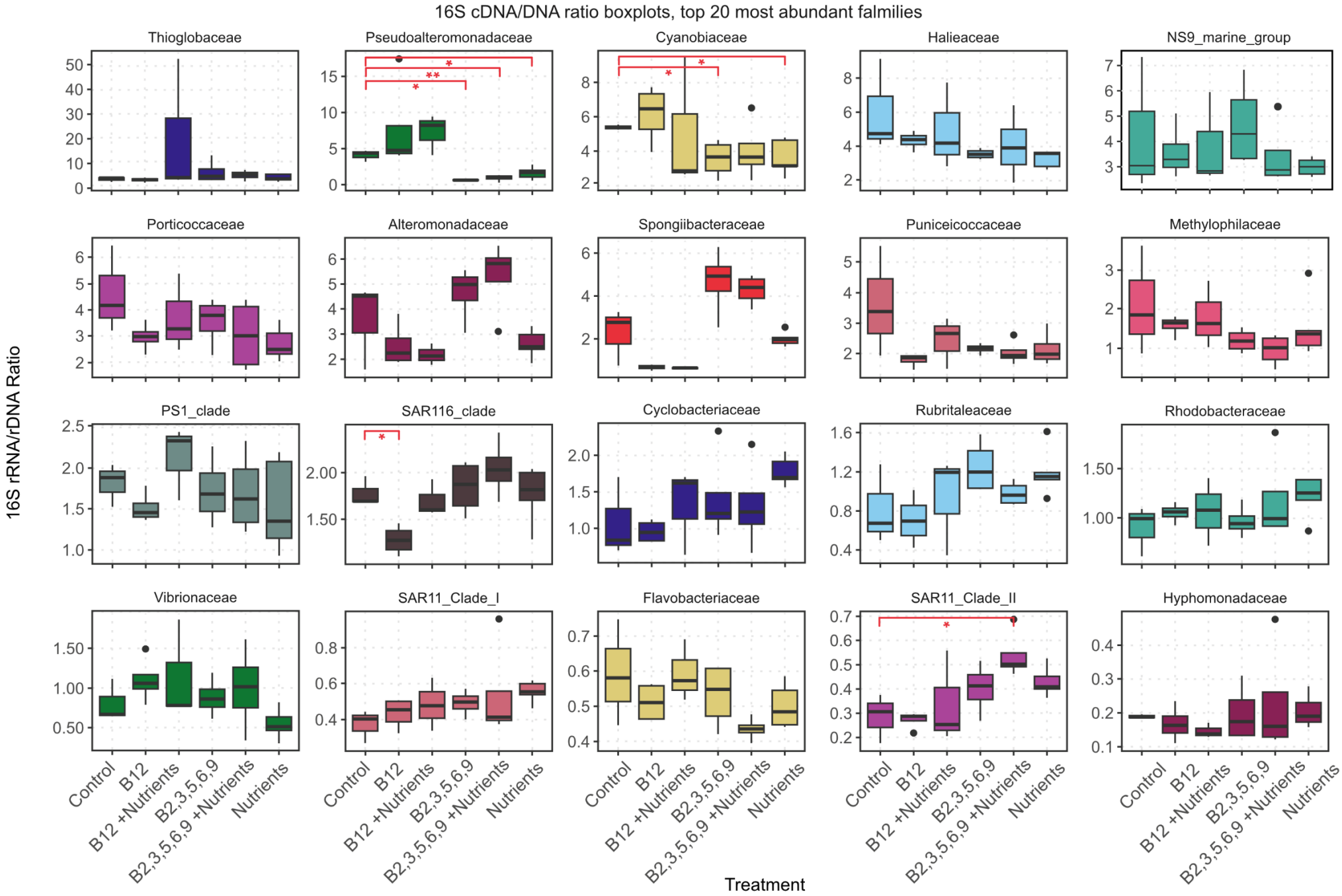


▲ ***Figure 2.5 16S Ratio and relative abundance of top 20 most abundant genera.*** *Circle plot displaying both the 16S rRNA/rDNA ratio and the relative abundance of the top 20 most abundant genera across the experiment. 16S ratio is displayed as a colour scale with blue representing a low ratio (inactive) and red representing a high 16S ratio (highly active). Relative abundance of each genus in each sample is represented by the size of the circle where a larger balloon represents a higher relative abundance.*

► ***Figure 2.6 16S ratios represented as box plots - genus level.*** *Top 20 most abundant genera in the 16S samples. This is an alternative display of the ratio data presented in Figure 2.5. Plots are ordered by max ratio. Significance values are indicated with red stars representing a significant p-value in a t-test against the control condition.*

► ***Figure 2.7 16S ratios represented as box plots - family level.*** *The 16S rRNA/rDNA ratios for each of the top 20 families across the experiment are represented as box plots for each treatment. These are ordered by max ratio. Significance values are indicated with red stars representing a significant p-value in a t-test against the control condition.*


16S rRNA/rDNA ratio boxplots, top 20 most abundant genera
OM60(NOR5) clade
Synechococcus CC9902
Litorivivens
Alteromonas
SAR92 clade
Luminiphilus
Coraliomargarita
HIMB11
NS4 marine group
Ascidiaceihabitans
Planktomarina
Roseibacillus
Polaribacter 4
Polaribacter 2
Amylibacter
Clade Ia
Formosa
Aurantivirga
Hellea
Polaribacter
16S rRNA/rDNA Ratio
Control
B12
B12 +Nutrients
B2,3,5,6,9
B2,3,5,6,9 +Nutrients
Nutrients
Treatment

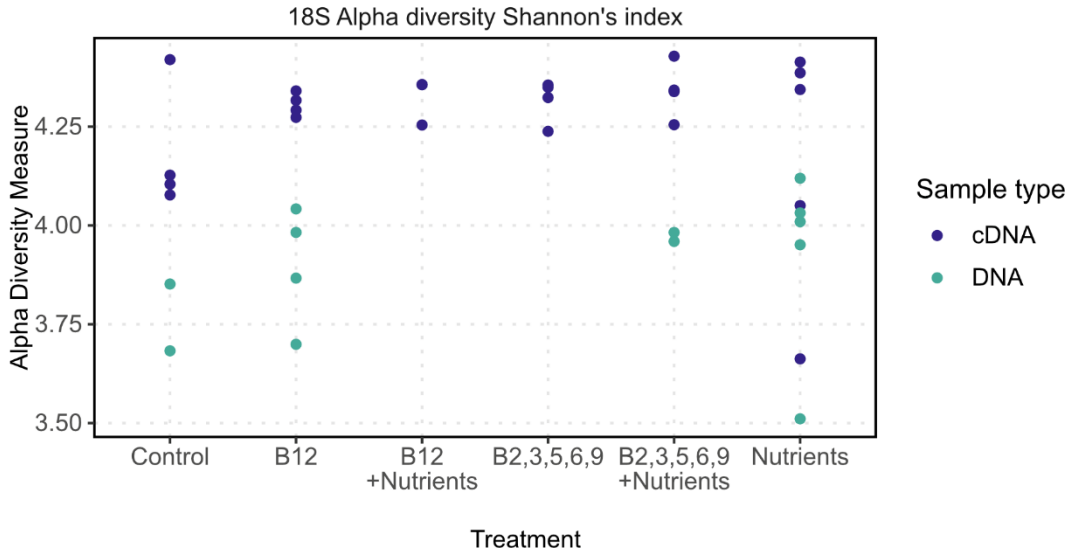

16S cDNA/DNA ratio boxplots, top 20 most abundant falmilies
Thioglobaceae
Pseudoalteromonadaceae
Cyanobiaceae
Halieaceae
NS9_marine_group
Porticoccaceae
Alteromonadaceae
Spongiibacteraceae
Puniceicoccaceae
Methylophilaceae
PS1_clade
SAR116_clade
Cyclobacteriaceae
Rubritaleaceae
Rhodobacteraceae
Vibrionaceae
SAR11_Clade_I
Flavobacteriaceae
SAR11_Clade_II
Hyphomonadaceae
16S rRNA/rDNA Ratio
Control
B12
B12 +Nutrients
B2,3,5,6,9
B2,3,5,6,9 +Nutrients
Nutrients
Treatment

2.3.3 Eukaryotic community composition

In a preliminary analysis run samples were heavily dominated by multicellular species and despite taking measures to remove such organisms before the experiment was conducted DNA from these species remained. To better assess the results of the experiment on the protist community, reads assigned as "Opisthokonta" at the taxonomic rank of "Supergroup", representing animals and fungi, were removed prior to re-running the analysis in full.

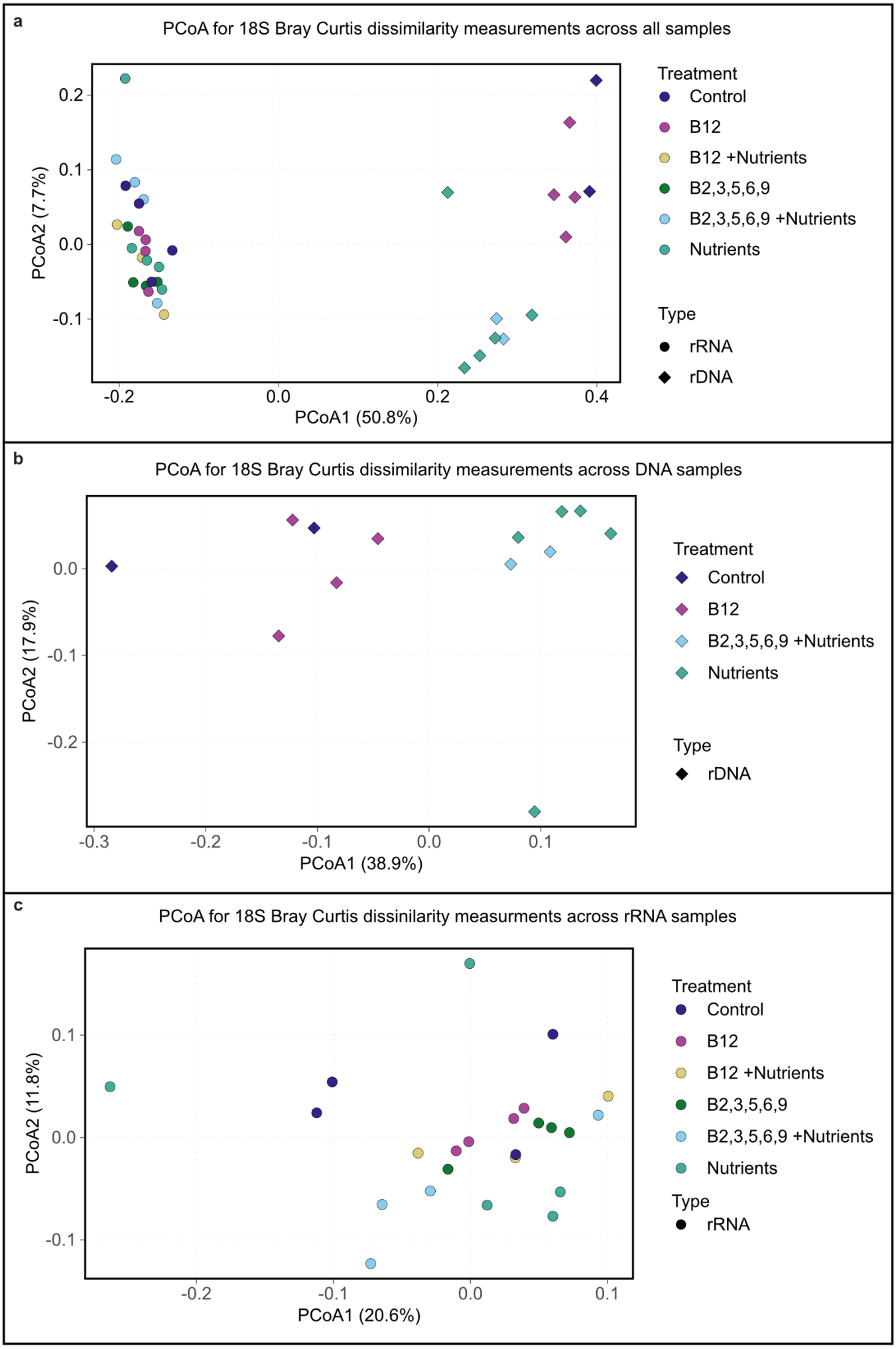
2.3.3.1 – 18S Alpha diversity

The community composition of the Eukaryotes was assessed using the same measures as for the prokaryotic community. A similar pattern was observed for the alpha diversity measures, with the Shannon's index being consistently higher for the rRNA type samples compared to the rDNA type (Figure 2.8). However, the quality of the 18S amplicon sequencing was lower than that of the 16S and with some samples failing the quality control steps, as a result some treatments did not have both rRNA and rDNA samples. Despite this, a two-way ANOVA of the type and treatment revealed a significant relationship between the Shannon's index of the samples and type ($p = 6.7 \times 10^{-7}$). In



▲ ***Figure 2.8. Alpha diversity across 18S (eukaryotic) rRNA and rDNA samples measured by Shannon's index.*** *Alpha (α) diversity measures the community diversity at individual sites. Here, α diversity was measured using Shannon's index which accounts for the total number of species in a sample and the distribution of abundance of the species at that site. rRNA samples are represented in purple and rDNA samples are represented in teal. Each point represents an individual sample. n=5 for each condition. N.B. some samples failed and are therefore not represented.*

► ***Figure 2.9 a-c. Beta diversity across 18S (eukaryotic) rRNA and rDNA samples measured by Bray-Curtis Dissimilarity and visualised by PCoA.*** *Beta diversity measures the differences in community composition between samples. Here it has been calculated using Bray Curtis Dissimilarity and visualised using Principal Components of Analysis (PCoA).* ***(a)*** *PCoA of all 18S samples. rDNA samples are represented as diamonds and rRNA as circles. Both are coloured according to treatment.* ***(b)*** *PCoA of 18S rDNA samples only.* ***(c)*** *PCoA of 18S rRNA samples only. Each point represents an individual sample. n=5 for each condition. N.B. some samples failed and are therefore not represented.*


a
PCoA for 18S Bray Curtis dissimilarity measurements across all samples
PCoA2 (7.7%)
PCoA1 (50.8%)
0.2
0.1
0.0
-0.1
-0.2
0.0
0.2
0.4
Treatment
Control
B12
B12 +Nutrients
B2,3,5,6,9
B2,3,5,6,9 +Nutrients
Nutrients
Type
rRNA
rDNA
b
PCoA for 18S Bray Curtis dissimilarity measurements across DNA samples
PCoA2 (17.9%)
PCoA1 (38.9%)
0.0
-0.1
-0.2
-0.3
-0.2
-0.1
0.0
0.1
Treatment
Control
B12
B2,3,5,6,9 +Nutrients
Nutrients
Type
rDNA
c
PCoA for 18S Bray Curtis dissinilarity measurments across rRNA samples
PCoA2 (11.8%)
PCoA1 (20.6%)
0.1
0.0
-0.1
-0.2
-0.1
0.0
0.1
Treatment
Control
B12
B12 +Nutrients
B2,3,5,6,9
B2,3,5,6,9 +Nutrients
Nutrients
Type
rRNA

contrast to the 16S analysis, there was no significant effect of treatment on the Shannon's index. A Tukey's test corroborated the result of the ANOVA.

2.3.3.2 – 18S Beta diversity

The beta diversity was again visualised using a PCoA plot and can be seen in Figure 2.9a. The same trend observed in the 16S analysis was observed in the 18S of a clear split between rRNA and rDNA samples, although the rDNA samples formed a less distinct group and there were fewer of them (as explained above). There was still a visual difference in the placings of the two types of samples with PCoA1 accounting for 50.8% of beta diversity variation across the experiment and PCoA2 accounting for 7.7%. The PERMANOVA test on this dataset revealed that the community composition of the samples was affected by both type and treatment ($p = 0.001$ and $p = 0.003$ respectively). It also suggested a significant interaction between these two factors ($p = 0.017$). Homogeneity of multivariate dispersions also revealed a significant p-value, which indicates that the significant PERMANOVA p-value for treatment can in part be explained by the dispersions of the groups. In this case, centroid positions appear to be less clearly separated which suggests the dispersions of the groups is the dominant factor in determining the significant interaction between treatment and beta diversity (see Section 2.3.1.2 for description of tests).

To further investigate this, a pairwise PERMANOVA was set up as described for the 16S analysis. A significant p-value ($p<0.05$) would suggest a difference in the community composition between the treatments being tested. Of the test between like types of samples the following were significant in the rRNA: control vs B vitamin mix + nutrients, B_{12} vs B vitamin mix and B_{12} vs B vitamin mix + nutrients. The following were significant in the rDNA samples: B_{12} vs nutrients. In no cases was there a significant difference in community composition between two treatments in both the corresponding rRNA and rDNA samples. Full pairwise PERMANOVA p-value results are in Appendix Section 1.6.

2.3.4 Eukaryotic taxa abundancies and activities

2.3.4.1 – 18S relative abundances

Abundance of eukaryotic taxa were evaluated at the taxonomic rank of Division for broad community analysis. Samples were dominated by a few distinct groups although to varying degrees in the two data types. The rRNA samples were dominated by *Dinoflagellata, Ochrophyta* and *Cercozoa* accounting for 31%, 25% and 15% of rRNA reads respectively. rDNA reads were dominated by the same organisms but in different proportions: *Dinoflagellata* (41%), *Ochrophyta*

(33%) and to a lesser extent *Cercozoa* (8%). Relative abundance of dominant Divisions can be seen in Figure 2.10 and the average abundance of these Divisions can be found in Table 2.3. Reads representing *Dinoflagellata* and *Ochrophyta* are more abundant in rDNA samples than in rRNA samples, whereas for the Divisions *Cercozoa*, *Chlorophyta*, *Ciliophora*, *Haptophyta* and *Telonemia* the reverse is true.

Table 2.3. Relative abundances of dominant eukaryotic taxa in the two different types of samples. Standard Deviation in brackets.

	rDNA	rRNA
Dinoflagellata	41.3 (6.6)	30.7 (6.6)
Cercozoa	8.1 (3.8)	15.1 (2.5)
Ochrophyta	33.0 (4.0)	24.6 (4.3)
Chlorophyta	5.6 (1.5)	7.2 (1.9)
Ciliophora	5.3 (0.9)	9.0 (1.4)
Haptophyta	2.9 (1.8)	8.2 (2.3)
Telonemia	0.4 (0.3)	1.5 (0.6)
Stramenopiles X	1.5 (0.9)	0.7 (0.4)
Pseudofungi	1.0 (0.4)	1.5 (0.4)
Cryptophyta	0.1 (0.1)	0.3 (0.1)
Picozoa	0.1 (0.1)	0.4 (0.2)

2.3.4.2 – 18S ratios

18S ratios of rRNA/rDNA were calculated as a proxy for species activity. There were 923 species detected in the experiment and 253 genera. To begin to analyse this data the top 20 most abundant genera are displayed in Figure 2.11 with their 18S ratio. The full taxonomic assignment for these top 20 genera and all 253 genera can be found in Appendix Section 1.7. Due to the loss of samples during the sequencing and QC steps, there are a limited number of samples with both rRNA and rDNA samples and therefore limited ratio data to plot (this unfortunately resulted in no ratio data for the conditions - B vitamin mix and +B_{12} +nutrients). One genus that appears to be affected by the B vitamin mix +nutrient treatment compared to the control is *Ostreococcus.* This genus has a relatively low ratio in the control and a higher ratio in the +B vitamin mix +nutrients treatment as well as the +nutrient only treatment. However, with only two control samples having ratio data it is difficult to draw concrete conclusions. Indeed, this is further demonstrated in Figure 2.12 which plots the ratios as a box plot for each genus. Very few differences can be seen and those that can (i.e., *Karenia* and Telonemia-Group-2) are likely due to outliers or to two very different data points. Ratio box plots were also plotted at the higher resolution of the ASV (species) level (Figure 2.13). A number of the top ASVs were the same as those plotted in the genera plots. The full taxonomic

classifications of these top twenty ASVs can be found in Table 2.4 and the classification of all species detected can be found in Appendix Section 1.8

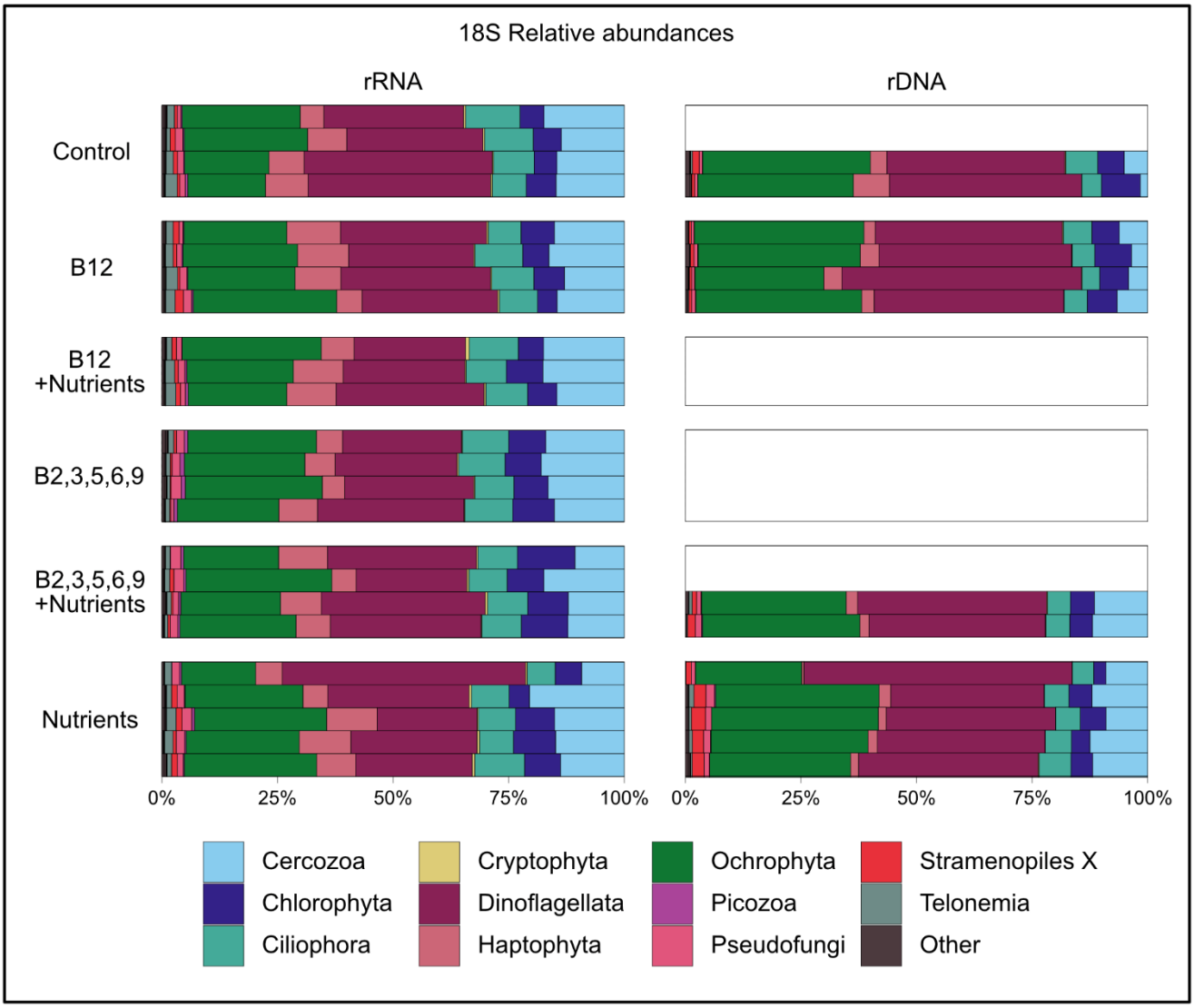
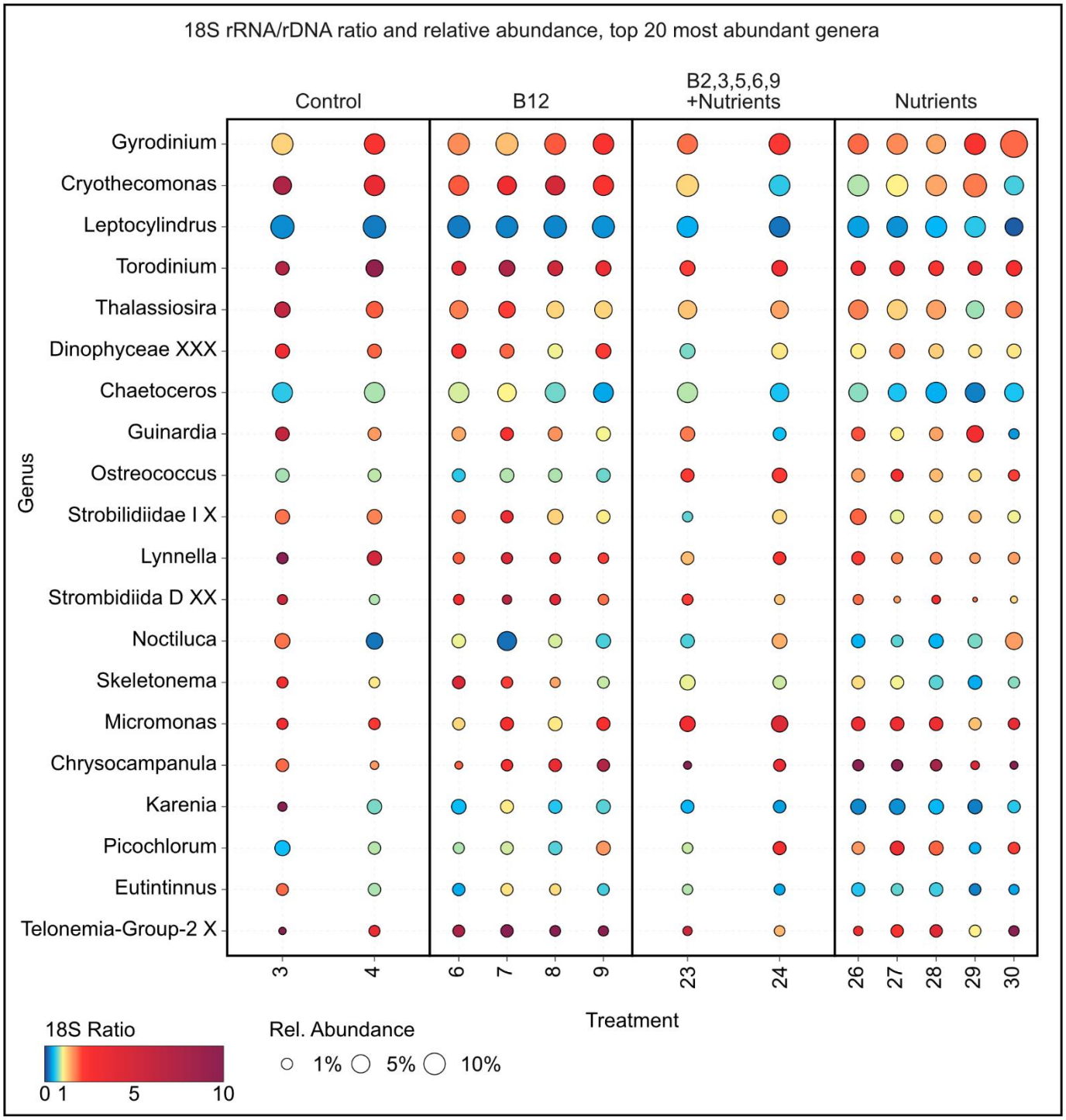


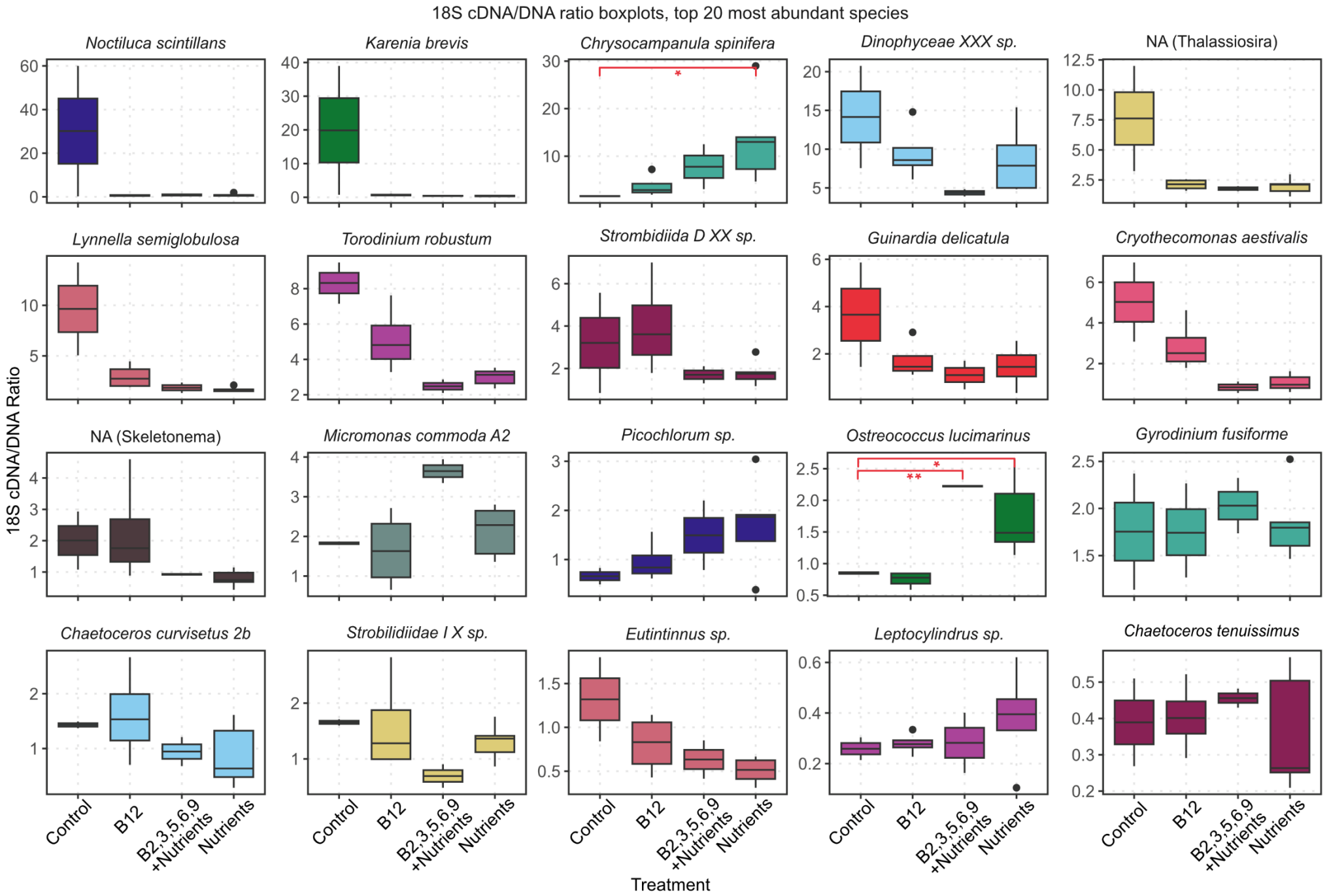
Figure 2.10 Relative abundances of 18S rRNA and rDNA samples taxonomically assigned at the rank of family*. Percentage of reads assigned to each division rank (as classified by the pr2 database) in each sample. The top 11 most abundant divisions are represented with the following families grouped into "other": Centroheliozoa, Prasinodermophyta, Katablepharidophyta, Conosa, Sagenista, Radiolaria, Opalozoa, Streptophyta, Apicomplexa, Apusomonadidae.*



▲ ***Figure 2.11 18S Ratio and relative abundance of top 20 most abundant genera.*** *Circle plot displaying both the 18S rRNA/rDNA ratio and the relative abundance of the top 20 most abundant genera across the experiment. 18S ratio is displayed as a colour scale with blue representing a low ratio (inactive) and red representing a high 18S ratio (highly active). Relative abundance of each genus in each sample is represented by the size of the circle where a larger balloon represents a higher relative abundance.*

► ***Figure 2.12 18S ratios represented as box plots (genus level).*** *Alternative display of the same ratio data as for figure 2.5. The 18S rRNA/rDNA ratios for each of the top 20 genera across the experiment are represented as box plots for each treatment. Significance values are indicated with red stars representing a significant p-value in a t-test against the control condition.*

► ***Figure 2.13 18S ratios represented as box plots (species level).*** *Top 20 most abundant ASVs representing the species level. Data as presented in Figure 2.11 with the exception that ASVs are represented at the species level rather than genus. Significance values are indicated with red stars representing a significant p-value in a t-test against the control condition.*


18S cDNA/DNA ratio boxplots, top 20 most abundant species
Noctiluca scintillans
Karenia brevis
Chrysocampanula spinifera
Dinophyceae XXX sp.
NA (Thalassiosira)
Lynnella semiglobulosa
Torodinium robustum
Strombidiida D XX sp.
Guinardia delicatula
Cryothecomonas aestivalis
NA (Skeletonema)
Micromonas commoda A2
Picochlorum sp.
Ostreococcus lucimarinus
Gyrodinium fusiforme
Chaetoceros curvisetus 2b
Strobilidiidae I X sp.
Eutintinnus sp.
Leptocylindrus sp.
Chaetoceros tenuissimus
18S cDNA/DNA Ratio
Control
B12
B2,3,5,6,9 +Nutrients
Nutrients
Treatment

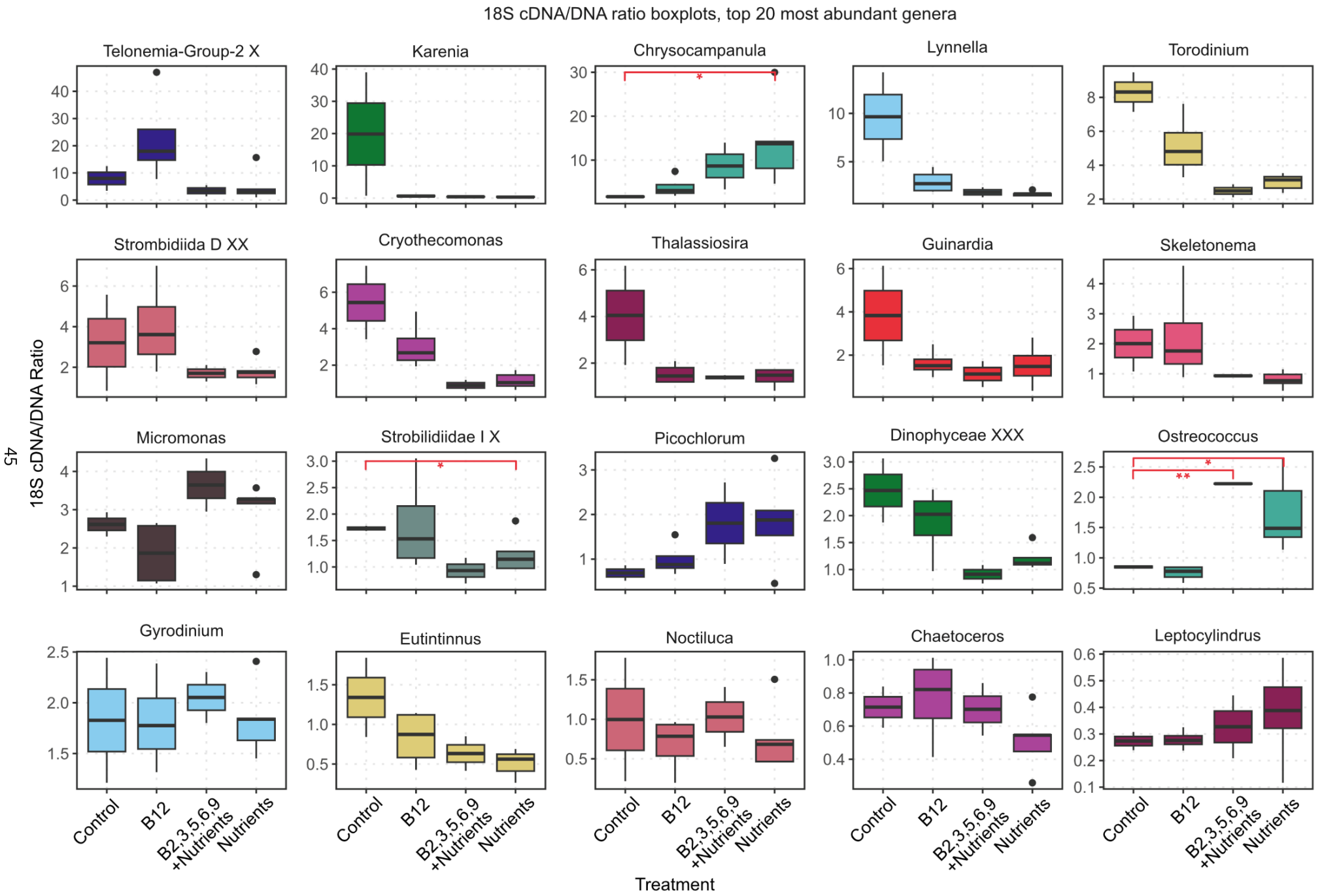

18S cDNA/DNA ratio boxplots, top 20 most abundant genera
Telonemia-Group-2 X
Karenia
Chrysocampanula
Lynnella
Torodinium
Strombidiida D XX
Cryothecomonas
Thalassiosira
Guinardia
Skeletonema
Micromonas
Strobilidiidae I X
Picochlorum
Dinophyceae XXX
Ostreococcus
Gyrodinium
Eutintinnus
Noctiluca
Chaetoceros
Leptocylindrus
18S cDNA/DNA Ratio
Control
B12
B2,3,5,6,9 +Nutrients
Nutrients
Treatment

Table 2.4. Full taxonomic classifications of top 20 ASVs from 18S amplicon data. Classifications were performed using the PR2 database

	Supergroup	Division	Class	Order	Family	Genus	Species
ASV 2	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Gymnodiniaceae*	*Gyrodinium*	*Gyrodinium_fusiforme*
ASV 3	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Cryomonadida*	*Cryothecomonas-lineage*	*Cryothecomonas*	*Cryothecomonas_aestivalis*
ASV 6	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Leptocylindrus*	*Leptocylindrus_sp.*
ASV 7	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Torodiniales*	*Torodiniaceae*	*Torodinium*	*Torodinium_robustum*
ASV 8	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Thalassiosira*	*NA*
ASV 11	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Chaetoceros*	*Chaetoceros_curvisetus_2b*
ASV 13	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Dinophyceae_X*	*Dinophyceae_XX*	*Dinophyceae_XXX*	*Dinophyceae_XXX_sp.*
ASV 14	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Chaetoceros*	*Chaetoceros_tenuissimus*
ASV 15	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Guinardia*	*Guinardia_delicatula*
ASV 16	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Bathycoccaceae*	*Ostreococcus*	*Ostreococcus_lucimarinus*
ASV 19	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Strobilidiidae_I*	*Strobilidiidae_I_X*	*Strobilidiidae_I_X_sp.*
ASV 20	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Lynnellidae*	*Lynnella*	*Lynnella_semiglobulosa*
ASV 21	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida_D*	*Strombidiida_D_X*	*Strombidiida_D_XX*	*Strombidiida_D_XX_sp.*
ASV 23	*Alveolata*	*Dinoflagellata*	*Noctilucophyceae*	*Noctilucales*	*Noctilucaceae*	*Noctiluca*	*Noctiluca_scintillans*
ASV 24	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Skeletonema*	*NA*
ASV 25	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Mamiellaceae*	*Micromonas*	*Micromonas_commoda_A2*
ASV 26	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiales_X*	*Chrysocampanula*	*Chrysocampanula_spinifera*
ASV 27	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Kareniaceae*	*Karenia*	*Karenia_brevis*
ASV 29	*Archaeplastida*	*Chlorophyta*	*Trebouxiophyceae*	*Chlorellales*	*Chlorellales_X*	*Picochlorum*	*Picochlorum_sp.*
ASV 30	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Tintinnida*	*Eutintinnidae*	*Eutintinnus*	*Eutintinnus_sp.*

2.4 Discussion

The results of this experiment show that amending natural microbial communities with mixtures of B vitamins and nutrients can change the composition of a community. However, although in both the prokaryotic and eukaryotic cohorts there was an indication that the treatment conditions caused a change in the community composition of the samples (significant global relationship between treatment and beta diversity in 16S and 18S) the organisms responsible for the changes were difficult to determine. In the case of the eukaryotic community this is perhaps not surprising as an incubation time of only 24 hours does not allow for many, if any rounds of cellular duplication in most species.

In the prokaryotic community a significant change in composition was detected between the control and the addition of the B vitamin mixture (B_2, B_3, B_5, B_6 and B_9) which indicates that these micronutrients play a role in the dynamics of this group. Indeed, this had been indicated by Gómez-Consarnau et al. (2018) where they demonstrated that different members of the community contribute to the biosynthesis of vitamins B_1, B_3, B_6 and B_{12} to different degrees making some species net consumers and others net producers. Two of the most highly abundant genera in the experiment appear to respond dramatically to the nutrient amendments: *Litorivivens* and *Alteromonas.* These genera both responded in a similar pattern across the different treatments: a ratio of above 0 but below 5 in the control treatment, a much higher ratio in the two treatments that included the B vitamin mix and ~0 ratio in the treatments that included B_{12}. In addition, in both cases the Nutrient treatment resulted in a ratio similar to the control ratio. The increase in activity in the B vitamin mix treatments might suggest that these vitamins are limited for the growth of these genera in this environment or that increases in the amount of these vitamins (or one of the vitamins in the mix) is used as a signal for proliferation. The fact that activity is lower in the B_{12} treatments compared to the control is intriguing. Although the genus of *Litorivivens* is poorly studied with no available genome, the genus of *Altermonas* has a number of species with genomic data available. None of these species are predicted to biosynthesise B_{12} (Kudahl, 2018) and the majority of them are predicted to have both METE and METH methionine synthase isoforms (B_{12} independent and B_{12} dependant respectively), making them independent of B_{12} which makes the reduction of activity upon addition of B_{12} more curious.

The 18S ratio data revealed very few changes in activity across the experiment. Although the box plots suggest that some change due to the treatments this is likely observed because of a lack of data (only two samples for each control were available) skewing the boxplot. Indeed, the overall

sequencing effort for the 18S data was relatively poor. Many of the sequencing runs failed and many others had very low read depth. The consequence of this was a low level of rarefaction, which limits the data used from the samples with much better read coverage. This is however necessary to maintain enough samples for a worthwhile comparison between treatments. Lower read depth for 18S samples compared to 16S samples is not surprising given the likely difference in abundance of cells in a given sample (Keeling and Campo, 2017). However, internal anecdotal evidence has found that in eukaryote-bacteria co-cultures it is often more difficult to extract the bacterial DNA or RNA than it is the eukaryote.

Despite the limitations of the sequencing quality for the 18S samples, some insights can be made. Of the top twenty most abundant species, fifteen were phytoplanktonic species belonging to the groups Chlorophyta, Dinoflagellata, Haptophyta or the diatoms (Bacillariophyceae). These are the type of species expected to be seen at this sampling station from this depth of water (photosynthetic species in the photic zone) validating the experimental method to an extent. The second most abundant species was *Cryothecomanas aestivalis* of the phylum Cercozoa. *C. aestivalis* is a heterotrophic species that performs phagotrophy on diatom cells, reportedly favouring the diatom *Guinardia delicatula* (Drebes *et al.,* 1996) which was the ninth most abundant species in this study. The other four species were members of the ciliate division whose trophic mode is typically heterotrophic, feeding on other protists and bacteria (Weisse, 2017).

One ASV that showed increased cDNA/DNA ratio levels in the treatment conditions was *Chrysocampanula spinifera,* which had higher ratios in the B vitamin mix treatment and in the nutrient condition than in the control. *C. spinifera* belongs to an early diverging branch of the supergroup Haptophyta (family: Prymnesiales) and is likely to be mixotrophic (De Vargas *et al.,* 2007). This trophic mode might explain their response upon nutrient addition if other more dominant autotrophic taxa are proliferating using the available light.

The initial aim of the experiment was to assess the results in the context of the stramenopiles, in particular MASTs. Ochrophyta sequences were dominant in the experiment and displayed reduced activities in most treatment conditions compared to the control (e.g., *Thalassiosira* and *Skeletonema* Figure 2.12 and 2.13). MASTs were detected and are represented in Figure 2.10 by the group “Stramenopiles X”. However, individual ASVs were not detected at a high enough abundance to assess their response to the experiment. This could be achieved in future adaptations of this

experiment by either using greater sample volumes or by grouping the data in a different way (e.g., at the MAST group number).

Of course, there are many other ways the data presented here could be analysed to identify trends. This could include analysing the average ratios of all taxa in each condition, which might demonstrate whether any of the conditions in general are beneficial for the community as a whole, rather than individual taxa or genera. Also, plotting raw abundance data might have value for understanding the dominance of different taxa. This would however not take into account the copy-number variation in the 18S and 16S genes in different species which is a common limitation in this type of experimental set up, not least because the copy numbers for most species remain unknown (Bowsher *et al.,* 2019; Lee *et al.,* 2009; Louca *et al.,* 2018). Indeed, in eukaryotes this issue is further complicated by intraspecific rDNA copy number variation (varying copy numbers from strain to strain of a species) and by the plasticity of rDNA copy number in response to environmental conditions (Lavrinienko *et al.,* 2021). In the design of this study the focus was to produce the RNA/DNA ratio data to give a proxy to relative activity in different conditions to assess response to the nutrient treatments. However, further iterations of this pilot study could employ a modelling technique as employed by utilising the Ribosomal RNA Operon Copy Number Database (rrnDB) (Martin et al. 2021). A further consideration may be given to the extraction methods used as it is known that different methods can yield different results and therefore change the perceived composition of the community (Catlett et al. 2019).

In conclusion the results of this experiment hint at nutrient limitation in this environmental sample from the English Channel. However, further experimental work would be required to determine which nutrients are most limiting and to which taxa specifically. The cases of *Litorivivans* and *Altermonas* in the 16S data provide evidence that bacterial taxa exist in a nutrient limited state and this warrants further investigation to understand drivers on the coastal population of prokaryotes. Further iterations of this experimental design could explore More specifically the result of amending the samples with each individual B vitamin as well as those not included in this pilot study (B_3 and B_7). That in turn could lead to amendments with different combinations of the vitamins to analyse co-limitations to growth and or activity. One could also envision a more verbose design by integrating full RNA sequencing approaches to assess the transcriptional response of the community to the vitamin amendments. The fact the alpha diversity was measured as being higher in the cDNA than in the DNA samples is surprising. Intuitively one would expect the

opposite result as the sequencing of DNA amplicons should pick up both active and inactive cells whereas sequencing of transcripts for cDNA production would only pick up transcriptionally active cells. Indeed this has been reported many times (Salgar-Chaparro and Machuca 2019 and contained references). The result reported in this pilot study may be the result of a methodological error such as a suboptimal DNA extraction or an artefact of the PCR step of the sequencing protocol or the sequencing runs themselves, indeed a few of the sequencing runs failed – evidenced by the fact that there was not ratio data for every replicate and every condition. This could potentially be mitigated to an extent by randomising the sample plate prior to submission for sequencing.

3 – Comparative genomics of B vitamin biosynthesis pathways in the Stramenopila

3.1 – Introduction

As discussed in Chapter 1 Section 1.2, members of the stramenopile group of eukaryotes operate in fresh-water environments, brackish and estuarine waters as well as the ocean. As well as displaying diversity in habitat they also display a variety of lifestyle modes. The most well characterised of the stramenopiles are the photosynthetic lineages, however the group also has members who live heterotrophically. The oomycetes are parasitic osmotrophs (Lamour *et al.,* 2007; Latijnhouwers *et al.,* 2003), the Labyrinthuleae live a saprotrophic lifestyle (Raghukumar 2002; Raghukumar and Damare, 2011; Martin *et al.,* 2016) and members of the MAST groups are believed to be bacterivorous (Cavalier-Smith and Scoble, 2013; Massana *et al.,* 2014; Seeleuthner *et al.,* 2018).

Since the development of cheap readily available sequencing runs, genomic analysis has become a mainstream technique to understand the underlying biology of phenotypic variation. Comparative genomics allows researchers to identify common genomic elements between species with similar traits, for example in genome wide-association studies or simply identify common genes in related species. Additionally, the increase in sequencing data has expanded the power of phylogenetic analysis into phylogenomics, the use of multiple genetic sequences from each species of interest to generate a species tree based on a range of genes, not just one, allowing more robust relationships to be identified (Delsuc *et al.,* 2005). Combining comparative genomics with phylogenomics results in a powerful tool for understanding the evolution of genomes, genes and phenotypes.

Indeed, these analysis techniques have been employed to identify functional characteristics of newly sequenced genomes by characterising similar sequences with sequences of known function. Various tools and databases have been developed to allow this type of genome annotation to be carried out with relative ease and without the requirement of deep understanding of the underlying statistical functions taking place. Examples of such tools include the Gene Ontology (GO) database (Ashburner *et al.,* 2000), gene annotation tools like eggNOG (Huerta-Cepas *et al.,* 2019) and protein domain annotation tools such as Pfam and Interpro (Mistry *et al.,* 2021; Mitchell *et al.,* 2019). These tools allow the user to apply meaning to their sequence data *in silico* and either confirm or inform *in vivo* and *in vitro* experiments.

Another application of predicted gene or protein function is the detection of intact (or otherwise) metabolic pathways. The KEGG database (Kanehisa *et al.,* 2016a) is a collection of gene and protein

sequences characterised by function and arranged into known, experimentally validated, metabolic pathways. This allows the characterisation of functional biosynthetic pathways in newly sequenced genomes prior to physiological validation. Homologues to known pathway enzymes can be identified in the novel genome and demonstrate that all the necessary enzymes required to carry out a function are present (Helliwell *et al.,* 2013). This process can also be carried out with expression data to further validate that this pathway is being actively used by an organism. To a certain extent the logical reverse can be deduced where if no homologues are detected the metabolic pathway may be absent. This however has limitations due to the limited variation of pathway enzyme characterised and included in these databases which may not detect a distantly related sequence in a distantly related organism.

Using this methodology, it is possible to detect broken or fragmented pathways. It can also be used to study the evolution of the enzymes within the pathways. For example comparing structure and active site motifs to identify groups of homologous enzymes that have drifted in function to form a pathway (such as the case in tryptophan metabolism) (Horowitz, 1945; Lobley *et al.,* 2003). Conversely, pathways may evolve in a patchwork manner, recruiting enzymes of varied structure and function to the common biosynthesis goal, as has been demonstrated for B_5 biosynthesis in *Escherichia coli* (Lobley *et al.,* 2003). Increasing the availability of enzyme sequences and structures will only further our understanding of these evolutionary procedures.

As we are aware of the acute requirement of B vitamins for the growth of many Bacillariophyceae species in culture (Croft *et al.,* 2006; Tang *et al.,* 2010), it is reasonable to expect the trait of auxotrophy to be present in other members of the stramenopiles. However, this is not something that has been examined extensively to our knowledge. Given the multiplicity in lifestyles of the stramenopiles one could imagine the evolution of dependence on different B vitamins based on varying growth conditions and community structures. In this chapter I employ such comparative genomics methodologies to investigate the diversity of the stramenopiles in the context of B vitamin biosynthesis capability and assess whether the ability to biosynthesise the B vitamins *de novo* (with the exception of B_{12}) is as variable across the group as other traits are.

3.2 Materials and Methods

3.2.1 – Data sources

Data for this work was compiled from a range of sources. The genomes used in the in-depth analysis section were acquired from the Joint Genome Institute Genome portal

(https://mycocosm.jgi.doe.gov/heterokonta/heterokonta.info.html), the TARA oceans Single-cell amplified genomes (SAGs) were downloaded from https://www.genoscope.cns.fr/tara/ in the "Single-cell genomes" section (Seeleuthner *et al.,* 2018), and the transcriptomes were downloaded from the supplementary data of Thakur et. al (2019). A summary of the species used and the relative data sources can be found in Table 3.1.

Data for the large-scale analysis was also compiled from multiple sources. SMAGs (SAGs and MAGs) were acquired from https://www.genoscope.cns.fr/tara/ from the "Tara Oceans Eukaryotic Genomes (the "SMAGs")" section. Data from The Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) (re-assembled) was accessed from https://zenodo.org/record/257410#.YiocckDP1PY (Johnson *et al.,* 2019; Keeling *et al.,* 2014). Annotations for each file were acquired from supplementary sources in respective papers and used to select stramenopile accessions. The data from the in-depth analysis, plus any new JGI genomes, was also used in the large-scale analysis to compile a total data set of 512 stramenopile accessions. For each of these accessions, nucleotide sequences were acquired and used to make Anvi'o contig databases (contig-db)(Eren *et al.,* 2015).

3.2.2 – Data processing and analysis – In-depth analysis

3.2.2.1 – Phylogenomic tree of the stramenopiles

A phylogenomic tree of the stramenopiles was created to contextualise the species used in the in-depth analysis. This was achieved by the following process: Firstly, the individual alignments of the 120 genes used by Thakur et al. (2019) were acquired from https://doi.org/10.6084/m9.figshare.9252968 and imported into Geneious prime (Geneious Prime 2019.1.3). For the species used in the in-depth analysis that were not included in the tree presented by Thakur et al. (2019), the consensus sequence of each of the 120 genes were used a query sequence in BLAST searches to find homologues. New alignments were subsequently made with the additional data from these additional species (MAFFT Geneious prime plug-in)(Katoh *et al.,* 2002). Sequences causing spurious alignments caused by incorrectly identified homologues were removed before realignment of the data. Geneious prime was then used to create a concatenated alignment of these 120 individual gene alignments. The concatenated alignment was then stripped of sites with at least 80% gaps. Trees were then built with IQtree (Chernomor *et al.,* 2016; Hoang *et al.,* 2018) with the following parameters: iqtree -s infile.txt -bb 10000 -safe -bnni -alrt 10000 -st AA -seed 1000 -msub nuclear -t RANDOM -nt AUTO -pre out -m TEST

3.2.2.2 – Assessment of completeness

Benchmarking Single Copy Orthologues (BUSCO)(Manni *et al.,* 2021) was used as a metric to measure the completeness of each organism's genomic information. BUSCO version 4.1.3 was used with the lineage dataset stramenopiles_odb10 (Creation date: 2020-08-05, number of species: 27, number of BUSCOs: 100). In the case of amino acid (AA) sequences BUSCO was performed in "protein" mode and in the case of transcriptomic sequences BUSCO was performed in "transcriptome" mode.

3.2.2.3 – Sequence similarity searches for homologous biosynthesis enzyme sequences

Predicted AA sequences of each of the 20 species with genomic information (JGI genomes and TARA SAGs) were acquired and imported into Geneious prime. The transcriptomes from Thakur et al. (2019) were also imported into Geneious prime. A BLAST (Altschul *et al.,* 1997) database was built for all of the AA sequences and a separate one for the transcriptomic sequences.

Next, the biosynthetic pathways for each vitamin were defined and the compliment of enzymes required were identified (Table 3.2). This was achieved by cross referencing literary sources and the KEGG pathways database (Kanehisa *et al.,* 2016a, 2017). After defining the relevant enzymes, a representative sequence was found via the NCBI. This was used as an initial BLAST query sequence against the diatom species in the study and where possible a homologous diatom sequence was identified and was carried forward as the subsequent query sequence for that enzyme. In cases where no diatom sequence was found the initial query sequence was carried forward for the full BLAST analysis.

BLASTp was used for AA vs AA searches and tblastn (Altschul *et al.,* 1990) was used to search the AA queries against the transcriptomes (matrix used was BLOSUM62 with word size of 3 and a gap cost of 11). In order to reduce the likelihood of false negatives in the search, a maximum e-value threshold of 0.1 was used. The top hit for each query in each organism was then scrutinised to validate the nature of the sequence. The InterPro Scan (Jones *et al.,* 2014; Mitchell *et al.,* 2019) plug-in for Geneious was used to predict the protein domains of each of the query sequences used and for each of the top hits and then cross-referenced back to ensure only true homologues were recorded. In cases where no clear protein domain signature was found other techniques were used to check the likelihood of true homology including multiple-sequence alignment (MAFFT plug-in for Geneious, default parameters) (Katoh *et al.,* 2002) and phylogenetic (neighbour-joining) tree building (Fasttree plug-in for Geneious, default parameters)(Price *et al.,* 2009).

3.2.2.4 – Hypergeometric testing

In order to provide a statistical estimation of biosynthetic pathway completeness, a hypergeometric test was employed. This accounted for the level of missing data for each organism's genomic data. The hypergeometric test follows the formula:

$$px(k) = \Pr(X = k) = \frac{(\frac{K}{k})(\frac{N-K}{n-k})}{(\frac{N}{n})}$$

The probability of a complete pathway was calculated for each organism for each B vitamin. *N* is the size of full genome for the organism (estimated by extrapolation of BUSCO completeness estimation), *K* is the minimal number of enzymes required for biosynthesis of the given B vitamin, *n* is the number of number of entries in the fasta file for the organism in question and *k* is equal to the number of confirmed pathway enzymes found in that organisms' genomic data for that B vitamin pathway. Hypergeometric tests were performed in R using the *phyper* package. The null hypothesis for each test was that the full biosynthetic pathway is present, therefore the lower the p-value the greater the probability that the null is rejected and the greater the confidence in categorising the organism as having an incomplete biosynthetic pathway and therefore auxotrophic for the B vitamin involved in the test.

3.2.3 – Data processing and Anvi'o analysis workflow – Large scale analysis

3.2.3.1 – Anvi'o processing

Having created a contig-db from the nucleotide fasta file of each accession, an HMM profile search is conducted to identify single-copy orthologues and rRNA genes. This is achieved with the Anvi'o command anvi-run-hmms and uses hmmer search (hmmer.org). The output of this search is stored in the respective contig-db for each accession. These results were then used to estimate the completeness of each accession using the Anvi'o program anvi-estimate-genome-completeness. The level of completeness was used to filter out any accession with a completeness of below 40%. The data set was also filtered to leave a maximum of two accessions per species/strain. Leaving a final dataset of 321 accessions.

3.2.3.2 – Metabolism estimation

Next, the program anvi-estimate-metabolism was run on each contig-db. This program uses the KEGG database to annotate functions and metabolic pathways. More specifically, it annotates a contig-db with HMM hits from KOfam, a database of KEGG Orthologs (KOs). This was run with default parameters. Once each contig-db was annotated with KOs the program anvi-estimate-

metabolism could be run. This analyses the presence and absence of KOs in each KEGG pathway module and assesses the completeness of these modules to make a prediction of metabolic functions. However, presence and absence of a custom list if KOs was used to assess B vitamin pathways. This was done to match the defined biosynthetic pathways used in the in-depth analysis.

3.2.3.3 – Phylogeny

The phylogenomic tree used for this analysis was achieved with the following workflow. Sequences from the HMM hits previously identified by anvi-run-hmms were extracted from each contig-db and used to create a concatenated alignment of these single-copy genes for all accessions. This concatenated alignment was then trimmed to remove columns of the alignment where 20% of sites were a gap. The trimmed alignment was used to make a maximum-likelihood tree using IQ-tree using the following flags: -bb 10000 -safe -bnni -alrt 10000 -st AA -seed 1000 -msub nuclear -t RANDOM -nt AUTO.

3.2.3.4 – Figure creation

The Anvi'o interactive interface was used to generate figures. This combined the KO presence and absence annotations for each accession and the phylogenomic tree. Figures were configured as a linear phylogram. The binning function was used to create manual annotations of Class groups.

Table 3.1. List of selected species used to analyse B vitamin biosynthesis pathways in stramenopiles. ** = Genome, **= Single-Cell Amplified Genome, *** = Transcriptome*

Species	Source	Stramenopila Group	Lifestyle
Aureococcus anophagefferens	JGI*	Pelagophyceae	Phototrophic
Pelagophyceae sp.	JGI*	Pelagophyceae	Phototrophic
Fragilariopsis cylindrus	JGI*	Bacillariophyceae	Phototrophic
Pseudo-nitzschia multiseries	JGI*	Bacillariophyceae	Phototrophic
Phaeodactylum tricornutum	JGI*	Bacillariophyceae	Phototrophic
Thalassiosira pseudonana	JGI*	Bacillariophyceae	Phototrophic
Ochromonadaceae sp.	JGI*	Chrysophyceae	Phototrophic
CHRYH2	TARA Oceans**	Chrysophyceae	Heterotrophic
CHRYH1	TARA Oceans**	Chrysophyceae	Heterotrophic
Nannochloropsis oceanica	JGI*	Eustigmatophyceae	Phototrophic
Phytophthora cinnamomi var cinnamomi	JGI*	Oomycetes	Hetero(osmo)trophic
Phytophthora soja	JGI*	Oomycetes	Hetero(osmo)trophic
Aplanochytrium kerguelense	JGI*	Labyrinthuleae	Hetero(sapro)trophic
Aurantiochytrium limacinum	JGI*	Labyrinthuleae	Hetero(sapro)trophic
Schizochytrium aggregatum	JGI*	Labyrinthuleae	Hetero(sapro)trophic
MAST4E	TARA Oceans**	Eogyrea	Heterotrophic
MAST4C	TARA Oceans**	Eogyrea	Heterotrophic
MAST4A2	TARA Oceans**	Eogyrea	Heterotrophic
MAST4A1	TARA Oceans**	Eogyrea	Heterotrophic
Pseudophyllomitus vesiculosus	Thakur *et al.,* (2019)***	Eogyrea	Hetero(phago)trophic
Incisomonas marina	Thakur *et al.,* (2019)***	Nanomonadea	Heterotrophic
MAST3F	TARA Oceans**	Nanomonadea	Heterotrophic
MAST3A	TARA Oceans**	Nanomonadea	Heterotrophic
Platysulcus tardus	Thakur *et al.,* (2019)***	Platysulcea	Heterotrophic

Table 3.2 List of B vitamin pathway enzymes used as queries for homologue searches. Where appropriate the JGI protein ID of the stramenopile query used for the corresponding enzyme is also listed. Associated Interpro domains for each enzyme is recorded as determined by Interpro Scan.

Vitamin	Pathway enzyme	Functional enzyme name	Name	EC No.	KO - number	Initial Query	Secondary (stramenopile) query JGI Protein ID	Query Interpro domains
B_1								
	iscS	-	Cystine desulfurase	2.8.1.7	K04487	AT5G65720	jgi|Phatr2|34119	IPR010240, F005572, MF_00331, IPR016454, IPR000192, PF00266
	ThiF	-		2.7.7.73	K03148	AAB95618.1	jgi|Phatr2|34373	IPR000594, IPR012731
	ThiI	-	tRNA uracil 4-sulfurtransferase	2.8.1.4	K03151	NP_414957	-	IPR003720, MF_00021, TIGR00342, IPR020536, cd01712, PF02568
	dxs	-		2.2.1.7	K01662	AT3G21500	jgi|Phatr2_bd|1689	IPR005477, MF_00315, PTHR43322:SF8
	ThiH	-	2-iminoacetate synthase	4.1.99.19	K03150	NP_418417	-	IPR012726, TIGR02351
	ThiO	-	glycine oxidase	1.4.3.19	K03153	CBA18928	jgi|Phatr2|31544	IPR023209, PTHR11530, IPR006076, PF01266
	ThiM	-	hydroxyethylthiazole kinase	2.7.1.50	K00878	AT3G24030	-	IPR000417
	ThiG	-	thiazole synthase	2.8.1.10	K03149	AIF62680	jgi|Thaps3_bd|1620	IPR033983, IPR008867

	THI4	-		2.4.2.59	K03146	AT5G54770.1	-	IPR036188, IPR027495
	THI20	HMPK	HMP kinase	2.7.1.49	K00941	NP_014586	jgi|Phatr2|47293	IPR027574, IPR004399
	THI5/NMT1	HMPPS	phosphomethylpyrimidine synthase	(listed as THI5 in KEGG)	K18278	jgi|Phatr2|33535	-	IPR027939, PTHR31528, IPR015168, PTHR31528:SF1
	ThiC	HMPPS	phosphomethylpyrimidine synthase	4.1.99.17	K03147	AT2G29630.2	jgi|Phatr2|38085	IPR037509, IPR038521, IPR002817, IPR002817
	ThiDN	HMPPK	hydroxymethylpyrimidine kinase	2.7.4.7	K14153	AT1G22940	jgi|Phatr2|47293	IPR013749, IPR004399
	ThiE	TMP-Ppase	thiamine-phosphate pyrophosphorylase	2.5.1.3	K00788	EGI08266.1	jgi|Thaps3|262963	IPR034291, IPR022998, IPR036206
	ThiL		thiamine-monophosphate kinase	2.7.4.16	K00946	WP_089627679.1	-	IPR006283
	TPK	-	thiamine pyrophosphokinase	2.7.6.2	K00949	AT1G02880	jgi|Phatr2|5423	IPR007371, IPR036759, IPR006282
	TH2	-	thiamine phosphate phosphatase	3.1.3.100	K06949, K22911	AT5G32470	-	IPR004305, IPR016084
B_2								
	RibA	GTPCH2	GTP cyclohydrolase II	3.5.4.25	K01497	AT2G22450	jgi|Phatr2|10882	IPR032677, IPR000926, IPR036144
	RibD1	RIBD1		3.5.4.26	K11752	AT4G20960	jgi|Phatr2|33427	IPR002125
	RibD2	RIBD2		1.1.1.193	K11752	AT3G47390	-	IPR024072, IPR004794, IPR002734
	-	PYRP		3.1.3.104	K20862	AT4G11570.1	-	IPR041492
	RibB	DHBPS	DHBP synthase	4.1.99.12	K02858	AT2G22450	jgi|Phatr2|6454	IPR000422, IPR017945

	RibH	DMRLS		2.5.1.78	K00794	AT2G44050.1	jgi|Phatr2|7173	IPR002180, IPR034964
	RibE	riboflavin synthase	Riboflavin synthase	2.5.1.9	K02523	AT2G20690	jgi|Phatr2|38165	IPR017938, IPR001783, IPR026017
	RibK	FHY1	FMN hydrolase	3.1.3.102	K20860	AT1G79790.1	-	IPR023214
	rfk	RK	Riboflavin kinase	2.7.1.26	K20884	AT4G21470.1	jgi|Phatr2|35659	IPR015865
	-	NUDT23	FAD nucleotidohydrolase	3.6.1.18	K18453	AT2G42070.1	-	IPR020476, IPR000086, IPR029401
B_3								
	NadB	AO	Aspartate oxidase	1.4.3.16	K00278	34395939	jgi|Phatr2|30807	IPR003953
	NadA	QS	Quinolinate synthase	2.5.1.72	K03517	AT5G50210	jgi|Phatr2|14735	IPR003473, IPR036094
	NadX	AD	Aspartate dehydrogenase	1.4.1.21	K06989			
	NadC	QPRT	quinolinic acid phosphoribosyltransferase	2.4.2.19	K00767	AT2G01350	jgi|Phatr2|16599	IPR004393, IPR022412, IPR002638,
	NadD	NMNAT	nicotinate-nucleotide adenylyltransferase	2.7.7.18, 2.7.7.1	K00969, K06210, K06211	AT5G55810.2	-	IPR014729, IPR004821, PTHR12039, PTHR12039:SF0, SSF52374
	NadE	NADS	NAD+ synthase	6.3.1.5	K01916	AT1G55090	jgi|Phatr2|45509	IPR014445, IPR022310, IPR003010
	ppnK	NADK	NAD+ kinase	2.7.1.23	K00858	AT3G21070	jgi|Phatr2|12501	IPR002504
	pncB	NPRT	nicotinate phosphoribosyltransferase	6.3.4.21	K00763	AT4G36940	jgi|Phatr2|29887	IPR040727, IPR041525, IPR041619, IPR007229, IPR006405
	HAAO	-	3-hydroxyanthranilic acid oxygenase	1.13.11.6	K00452	110278810	jgi|Schag1|79768	IPR010329
	KMO	-	kynurenine 3-monooxygenase	1.14.13.9	K00486	75446355	jgi|Phatr2|49644	-

	KYNU	-	kynureninase	3.7.1.3	K01556	75345309	jgi|Schag1|94483	IPR010111, IPR000192
	HPAB	-		1.14.14.8	-	426020938	-	IPR004925, IPR024674, IPR024719
B_5								
	ilvE	BCAT	branched-chain amino acid aminotransferase	2.6.1.42	K00826	AT1G10060	jgi|Phatr2|32849	-
	PanB	KPHMT	ketopantoate hydroxymethyltransferase	2.1.2.11	K00606	AT2G46110	jgi|Phatr2|8765	IPR003700
	PanE	KPR	2-ketopantoate reductase	1.1.1.169	K00077	XP_001415462.1	jgi|Phatr2|16595	IPR013752
	PanC	PS	Pantothenate synthetase	6.3.2.1	K01918	AT5G48840	jgi|Phatr2|22955	IPR003721
	PanD	ADC	aspartate 1-decarboxylase	4.1.1.11	K01579	NP_414673	-	IPR003190, IPR009010
	FMS1	POA	polyamine oxidase	1.5.3.17	K17839	1Z6L_A	-	IPR002937, IPR036188
	PAO4	ALDH	aldehyde dehydrogenase (NAD+)	1.2.1.3	K00128	AT1G23800	-	IPR015590, IPR016161, IPR016162, IPR016163
B_6								
	pdxS	PDX1		4.3.3.6	K06215	AT2G38230	jgi|Phatr2|29885	IPR001852
	pdxT	PDX2		4.3.3.6	K08681	AT5G60540	jgi|Schag1|86456	IPR002161
	pyrO	PPOX	pyridoxamine phosphate oxidase	1.4.3.5	K00275	AT5G49970	jgi|Phatr2|5762	IPR004443
	-	PLR	pyridoxol dehydrogenase	1.1.1.65	K05275	AT5G53580	jgi|Phatr2|47510	IPR023210, IPR036812
	pyrK	PLK	pyridoxal kinase	2.7.1.35	K00868	AT5G37850	jgi|Phatr2|8828	IPR004625, IPR013749
	-	PHOSPHO2	pyridoxal phosphatase	3.1.3.74	K07758	NP_064711	-	IPR016965, IPR006384
B_7								
	BioF	KAPAS	7-keto-8-aminopelargonic acid synthetase	2.3.1.47	K00652	AT5G04620	jgi|Phatr2|36165	IPR004839, IPR015424

	BioA	DAPAS	7,8-diaminonanoate transaminase	2.6.1.105	K00833	584840	jgi|Phatr2|19427	IPR005815, IPR005814
	BioD	DTBS	dethiobiotin synthetase	6.3.3.3	K01935	584841	-	IPR004472
	BIO3-1	DAPAS/DTBS	Bi-functional bioA-bioD	2.6.1.62/6.3.3.3	K19562	AT5G57590	-	IPR004472, IPR005814
	BioB	BS	Biotin synthase	2.8.1.6	K01012	AT2G43360	jgi|Phatr2|21296	IPR002684, IPR024177
B_9								
	FolE	GCH1	GTP cyclohydrolase	3.5.4.16	K01495	AT3G07270	jgi|Phatr2|21868	IPR001474, IPR020602
	phoD	-	alkaline phosphatase	3.1.3.1	K03841	AT5G42370	-	IPR018946, IPR038607
	FolB	DHNA	dihydroneopterin aldolase	4.1.2.25	K01633	AT3G11750	-	IPR006157, IPR006156, IPR043133
	FolK	HPPK	hydroxymethyldihydropteridine pyrophosphokinase	2.7.6.3	K13941	AT4G30000	jgi|Phatr2|12107	IPR000550, IPR035907
	-	DHPS	dihydropteroate synthase	2.5.1.15	K13941	AT4G30000	jgi|Phatr2|12107	IPR006390, IPR000489, IPR011005
	FolC	DHFS	dihydrofolate synthase	6.3.2.12	K11754	AT5G41480	jgi|Phatr2|47685	IPR001645, IPR018109
	-	DHFR	dihydrofolate reductase	1.5.1.3	K13998	AT2G16370	jgi|Phatr2|1690	IPR001796, IPR012262, IPR024072
	-	FPGS	folylpolyglutamate synthase	6.3.2.17	K01930	AT3G55630	jgi|Phatr2|48414	IPR001645
	-	GGH	gamma-glutamyl hydrolase	3.4.19.9	K01307	AT1G78660	jgi|Phatr2|13333	IPR015527
	pabA	ADCS	aminodeoxychorismate synthase	2.6.1.85	K01664	XP_003081683.1	-	IPR005801, IPR005256
	pabB	ADCL	aminodeoxychorismate lyase	4.1.3.38	K18482	XP_003080205.2	-	IPR001544
B_{12}								
	METH	-	B12-dependent methionine synthase	2.1.1.13	K00548	34395941	jgi|Phatr2|23399	IPR011822, IPR004223, IPR033706, IPR003726,

							IPR003759, IPR006158, IPR000489, IPR036594, IPR037010, IPR036724
METE	-	Methionine synthase (B12-independent)	2.1.1.14	K00549	2851502	jgi|Phatr2|28056	IPR006276, IPR013215, IPR002629, IPR038071
MCM	-	methylmalonyl-CoA mutase	5.4.99.2	K01847	XP_002289878.1	-	IPR006099, IPR006098, IPR016176

3.3 – Results

3.3.1 – Initial assessment of Stramenopila species and their ability to biosynthesise B vitamins

3.3.1.1 – Experimental design

As an initial screen to investigate how the biosynthetic capabilities of the stramenopiles differs across the group a number of species with genomic information available were selected for analysis. The list of species along with information about their lifestyle can be found in Table 3.1. The data was acquired from three sources, genomes from the JGI, SAGs from the TARA oceans expedition (Seeleuthner *et al.,* 2018) and transcriptomes of deep branching stramenopiles produced by Thakur et al. (2019). The species used in this initial assessment can be seen in the context of the phylogeny of the stramenopiles in Figure 3.1, with representative species from most of the major groups. They also include both heterotrophic and phototrophic lifestyles.

Before assessing these species for their ability to biosynthesise each B vitamin, the *de novo* biosynthesis pathways for each compound had to be defined. This was achieved using the KEGG pathways database (Kanehisa *et al.,* 2016a) and cross referencing with literary reports of the pathways (Roje, 2007; Webb *et al.,* 2007). Each enzyme required to form the B vitamins 1 to 9 has been identified and are listed in Table 3.2. The propensity of each species to biosynthesise these vitamins was assessed by the presence or absence of homologues to these pathway enzymes, determined by sequence similarity searches using BLAST (Altschul *et al.,* 1990).

3.3.1.2 – Genome completeness assessment

The quality of the data to be analysed was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologues)(Seppey *et al.,* 2019; Simão *et al.,* 2015). This counts the number of present single-copy orthologues for common genes in a genome to give a score for relative completeness. In this instance the reference database was the stramenopile_10odb database consisting of 100 genes (Manni *et al.,* 2021). Completeness of the data used for each species can be seen in Table 3.3.

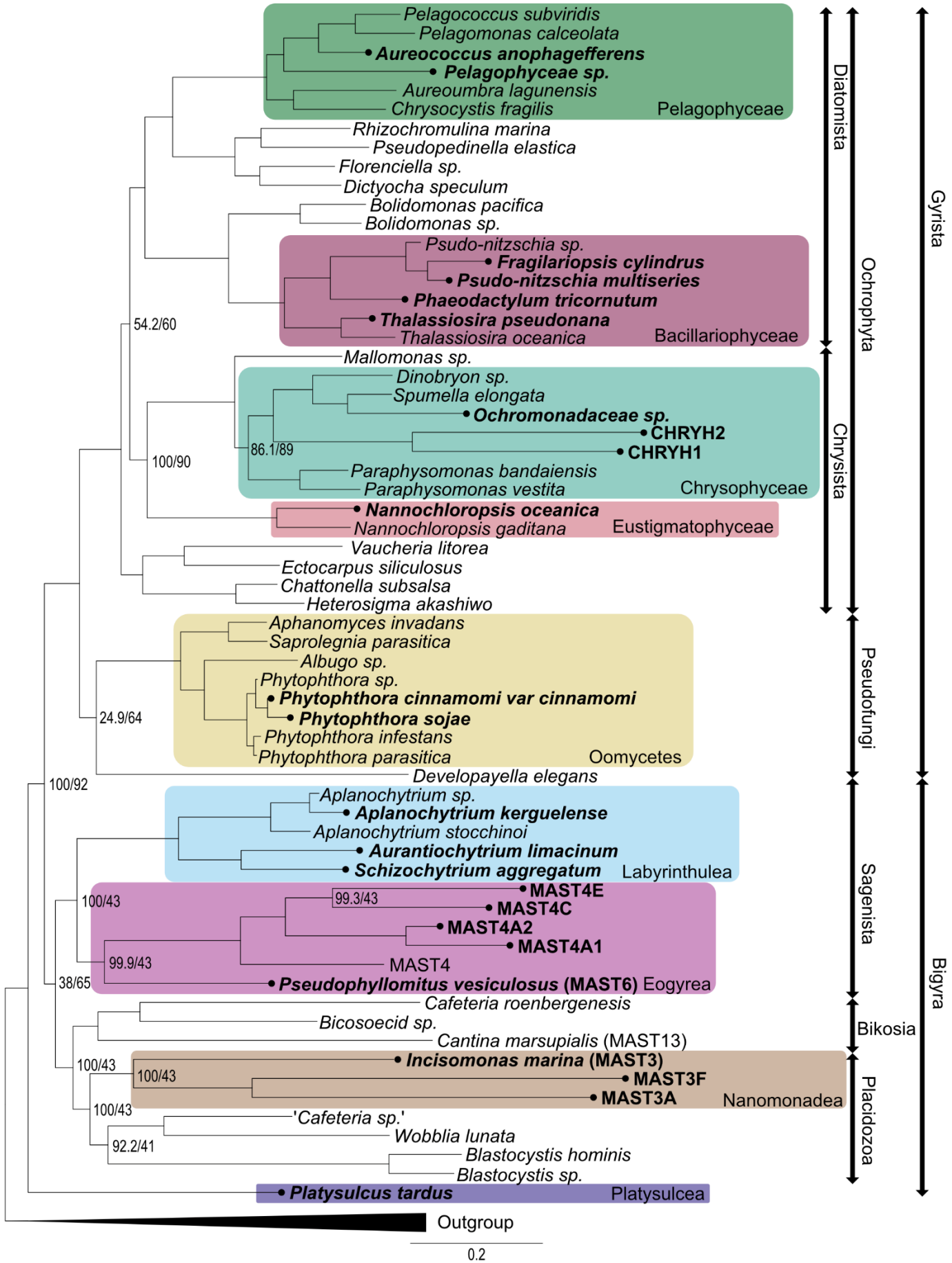


Figure 3.1. *Phylogenomic tree of the Stramenopila. Species involved in the initial screen for B vitamin biosynthesis pathways are indicated in bold. The tree was created using a concatenated alignment of 120 gene alignments with columns removed where 20% of sites were gaps. Gene alignments from Thakur et al. (2019) were used with additional data from the species of interest added. IQtree was used in model prediction mode to create a maximum-liklihood tree using ultrafast bootstrapping and the -altr flag, resulting in two values of node support. Those nodes with support values 100/100 are omitted. Taxonomic ranking on the right of the tree reflects those in Thakur et al., (2019). Outgroup includes species from the Rhizaria and Alveolata.*

Table 3.3 – BUSCO assessments – BUSCO Version 3.1 lineage dataset stramenopiles (eukaryota, 2020-08-05)

	Complete	Single copy	Duplicated	Fragmented	Missing	% Completion
Aureococcus anophagefferens	85	82	3	10	5	85
Pelagophyceae sp.	96	94	2	0	4	96
Fragilariopsis cylindrus	96	49	47	4	0	96
Pseudo-nitzschia multiseries	100	0	100	0	0	100
Phaeodactylum tricornutum	94	94	0	3	3	94
Thalassiosira pseudonana	96	94	2	3	1	96
Ochromonadaceae sp.	80	71	9	11	9	80
CHRYH2	23	22	1	7	70	23
CHRYH1	41	41	0	8	51	41
Nannochloropsis oceanica	98	2	96	0	2	98
Phytophthora cinnamomi var cinnamomi	96	94	2	4	0	96
Phytophthora soja	99	98	1	1	0	99
Aplanochytrium kerguelense	94	91	3	3	3	94
Aurantiochytrium limacinum	94	88	6	2	4	94
Schizochytrium aggregatum	91	86	5	3	6	91
MAST4E	51	51	0	3	46	51
MAST4C	49	36	13	3	48	49
MAST4A2	52	48	4	2	46	52
MAST4A1	41	39	2	7	52	41
Pseudophyllomitus vesiculosus	58	58	0	6	36	58
Incisomonas marina	74	73	1	4	22	74
MAST3F	32	30	2	5	63	32
MAST3A	43	30	13	5	52	43
Platysulcus tardus	67	66	1	5	28	67

3.3.1.3 – Thiamine – B_1

The biosynthesis pathway for thiamine is predominantly partitioned into two separate branches: the thiazol pathway (HET-P) and the pyrimidine pathway (HMP-P). At the point at which these two branches converge the thiamine phosphatase synthase enzyme (TMP-PPase) catalyses the reaction between HET-P and HMP-P to give the product thiamine phosphate (TMP). The product of the thiazole branch, HET-P, can be achieved through three separate branches depending on the organism.

1) In bacteria (Jurgenson *et al.,* 2009) HET-P can be the product of tyrosine/glycine metabolism, Pyruvate and Glyceraldehyde-3P, ThiS and IsC-SH and Cysteine, requiring the enzymes deoxylulose-5P synthase (encoded by dxs), ThiF, ThiI, ThiO and ThiH, together with cysteine desulphurase, IscS, the sulphur donor.

2) First characterised in yeast (Praekelt and Meacock, 1992), but also subsequently demonstrated in higher plants and the green alga *Chlamydomonas reinhardtii* (Croft *et al.,* 2007), HET-P is synthesised by THI4. The substrates for this enzyme were established by X-ray crystallography of both Arabidopsis enzymes (Godoi *et al.,* 2006) to be NAD+ and glycine, together with a S atom donated by a backbone cysteine residue.

3) There is also a scavenging pathway, where an external source of HET can be phosphorylated into HET-P by the enzyme encoded by *thiM* gene.

The presence and absence results for thiamine pathway enzymes in stramenopile genomes can be seen in Figure 3.2 as well as a schematic of the described pathway.

Each of the Bacillariophyceae species had homologues to each of the required enzymes for *de novo* B_1 biosynthesis. The same was true for *Nannochloropsis oceanica* (eustigmatophyte) and one of the Pelagophyceae species assessed – *Pelagophyceae sp.* The other pelagophyte, *Aureococcus anophagefferens*, had all but one of the required enzymes. The missing enzyme in this species' case was TMP-PPase indicating it is unlikely to perform the final reaction to form thiamine monophosphate. This result corroborates the report of Tang et al. (2010) who lists this species as B_1 auxotrophic.

Very few homologues to any of the defined B_1 pathway enzymes were identified in any of those species belonging to the Bigyra (labelled 13-24, Figure 3.2). This indicates a possible divergence between photosynthetic lineages and heterotrophic where the former has acquired the ability to

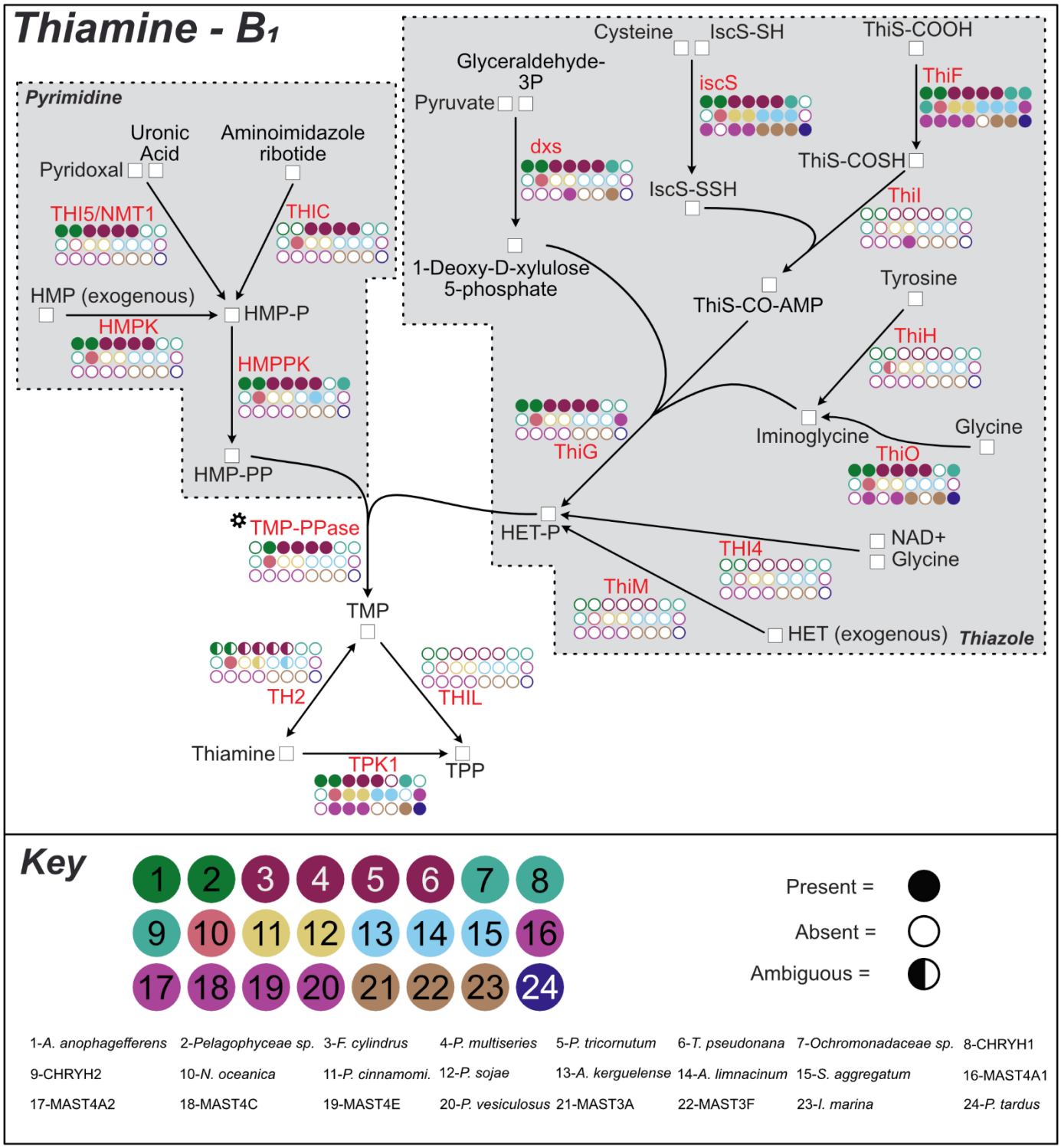


Figure 3.2. *Consensus biosynthesis pathway for thiamine from KEGG pathways and the literature. Overlayed with presence and absence results of each required enzyme for a selection of stramenopile species, determined by homology search with BLAST. Representative queries were used for each enzyme in the pathway to search for homologues in each species. Hits were manually verified using protein domain prediction (Interpro Scan) as well as multiple sequence alignments and phylogenetic tree building where necessary. The array associated with each enzyme name (in red) indicate whether that enzyme was identified in each species where the position in the array corresponds to the position in the Key. A present enzyme is indicated by a solid circle, where no homologue was identified an empty circle is displayed. Any instances where it could not be determined if the top hit was bona fide it is represented as a half circle. The array is colour coded by class and this corresponds to the phylogenomic tree of Figure 3.1.*

biosynthesise thiamine *de novo.* One group however is inconsistent with this. The chrysophytes are especially diverse with some members of this class being photosynthetic, some being heterotrophic and some being mixotrophic (del Campo and Massana, 2011; Graupner *et al.,* 2018; Lie *et al.,* 2018). Of the three species analysed in this experiment two are thought to be heterotrophic (CHRYH1 and CHRYH2) and one is thought to be phototrophic (*Ochromonadaceae sp.*). None of these three revealed homologues to enough of the thiamine pathway enzymes to predict them as *bona fide* biosynthesisers. That the Bigyra groups are unlikely to biosynthesise is supported by reports of thiamine auxotrophy for many of these species or close relatives (Robbins, 1938; Goldstein and Belsky, 1963; Goldstein, 1973; Tani, Yoneda and Suzuki, 2018).

Two different enzymes have been identified as producing the pathway intermediate HMP-P in the pyrimidine branch of the pathway. The HMPP synthase which forms HMPP from aminoimidazole ribotide is called ThiC and the other enzyme, forming HMPP from pyridoxal and uronic acid, is called THI5 (sometimes referred to as NMT1). In the four Bacillariophyceae homologues were identified to both of these HMPP synthases. However, in *N. oceanica* only a ThiC-like enzyme was identified and in the two pelagophytes a THI5-like sequence was detected. Llavero Pasquina (2020) identified the THI5 sequence in *P. tricornutum* previously. However, upon knocking this gene out, the organism did not become B_1 dependent, questioning the function of this gene as an HMPP synthase.

The ThiI enzyme, tRNA-uracil 4-sulfurtransferase (EC:2.8.1.4), proved to be widely elusive in this study. Only one positive hit was identified. That was in the transcriptome of *P. vesiculosus* (Eogyrea). As an outlier this result was scrutinised. ThiI protein sequences from the NCBI were downloaded and aligned with a translated protein sequence of the transcript from the tBLASTn hit. The translated protein from the transcript was 1,045 amino acids long. Around twice as long as those sequences acquired from the NCBI. Indeed, when analysed for protein domains, ThiI domains were only identified in the latter half of the sequence. Alignment of the sequences showed that large chunks of the sequence did not align at all with the other sequences despite the same ThiI domains being predicted (Pfam PF02568). In a phylogenetic tree of this alignment, the sequence was predicted to be more closely related to the bacterial sequences than eukaryotic sequences. This positive hit should therefore be treated with caution and could be contamination. However, as ThiI was not identified in those species known to biosynthesise B_1 it is not considered diagnostic of

auxotrophy. The same can be said for ThiH for which only one hit was detected in the species *N. oceanica.*

3.3.1.4 – Riboflavin – B_2

Riboflavin is the product of two converging pathway branches (Figure 3.3a). One starts with the molecule D-Ribulose 5-phosphate, and the other commences with GTP. The first enzyme of the GTP pathway is GTPCH2 which synthesises 2,5-Diamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one that is then converted 5-amino-6-(5'phospho-D-ribitylamino)uracil by the enzymes RIBD1 and RIBD2 via a two-step process. Few Stramenopile species surveyed had hits for both of these enzymes. However, this result may not confer auxotrophy in all cases. Many of the hits for one or the other of these enzymes were predicted to be bifunctional by protein domain prediction. A bifunctional enzyme capable of performing both the deamination and reduction functions of these two enzymes can still produce 5-amino-6-(5'-phospho-D-ribitylamino)uracil as a product. Nevertheless, *P. sojae* has no hit for either of these enzymes or riboflavin synthase suggesting a true inability to *de novo* synthesise B_2. This may be a result of this organism's parasitic lifestyle, indeed in culture it is grown on a medium including a yeast extract which contains exogenous B2 (Tyler, 2007).

The Bacillariophyceae again appear to have the ability to biosynthesise the vitamin as does *N. oceanica*, the two pelagophytes and the three Labyrinthuleae. Neither of the MAST4A SAGs have hits for riboflavin synthase. However, this may be an artefact of missing data as a result of the sequencing method rather than an indication of auxotrophy. This is especially likely to be the case as nearly all of the other pathway enzymes are present in the other Eogyra. *P. vesiculosus* and MAST4E, have the full pathway present. The three MAST3 species (MAST3A, MAST3F and *I. marina*) had very few positive hits for the enzymes in this pathway. This could genuinely be due to pathway incompleteness and the development of riboflavin auxotrophy. However, it is difficult to say this categorically. Both MAST3A and MAST3F had homology to the riboflavin synthase which may suggest that the other missing pathway enzymes could be false negatives. However, the fact that enzymes were missing consistently throughout the group would suggest this did not happen by chance, and that these three organisms truly are auxotrophic or exclusively convert a precursor into B_2. Only one of the oomycetes analysed had homologous sequences for all of the key enzymes (*Phytophthora cinnamomi var cinnamomi*).

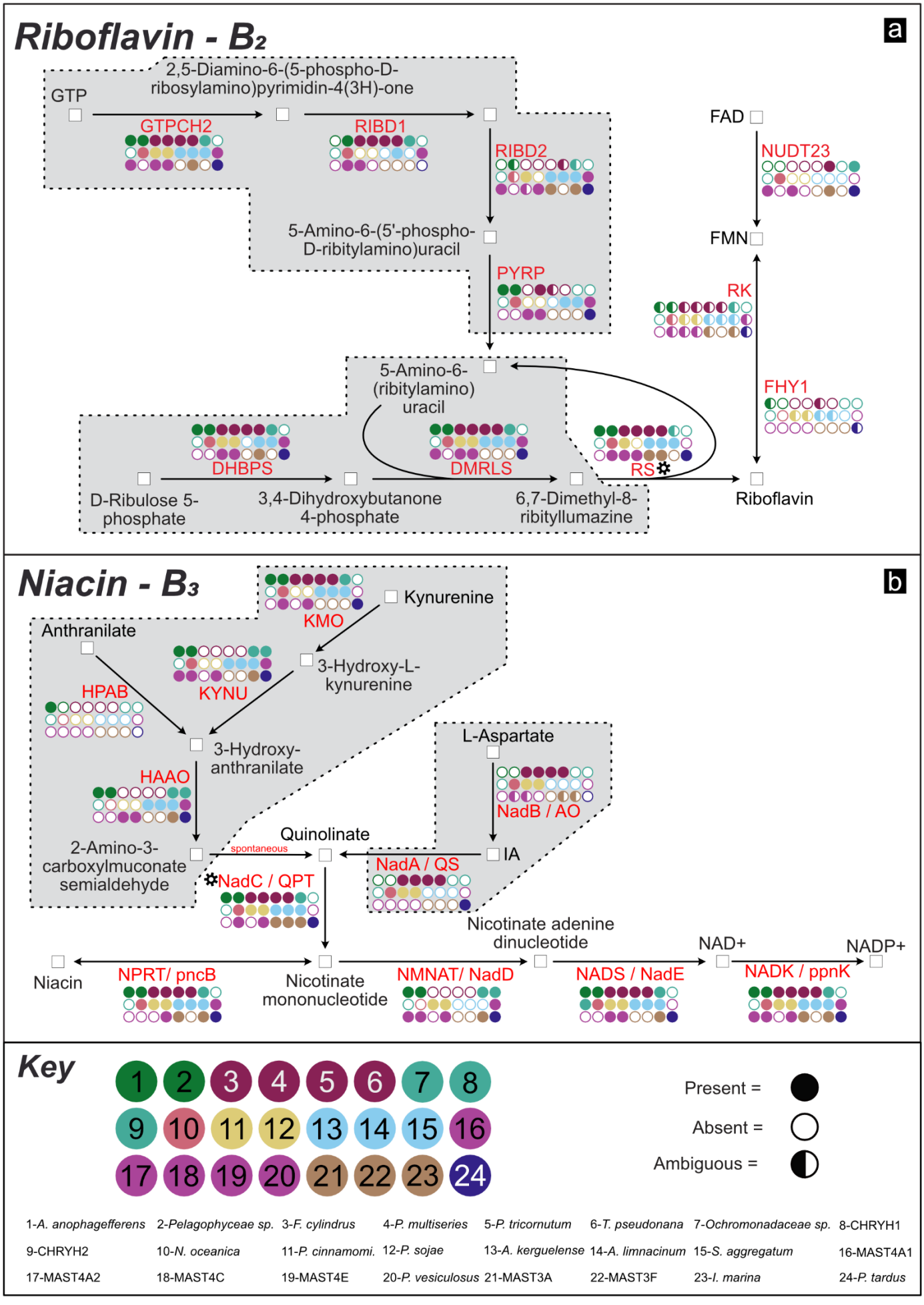


Figure 3.3 (a-b). *Pathway and presence and absence results for (a) riboflavin - B2 and (b) niacin- B3. See Figure 3.2 for full description.*

The enzyme 5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase (pyrP) was difficult to identify in many species. However, as it could not be identified in some known prototrophs it could not be considered as an auxotrophy marker in this analysis (although a true negative would result in auxotrophy due to the critical nature of this step in the pathway).

The hit for riboflavin synthase in *Ochromonadaceae sp.* is around 50% of the length of all other positive hits and when protein domain prediction is performed no signature riboflavin synthase domains are assigned. This could be a result of degradation, which would confer auxotrophy. However, this could also be a result of assembly or protein prediction errors. One other enzyme was missing in this organism, pyrP.

3.3.1.5 – Niacin – B_3

The biologically active forms of the vitamin B_3 are the metabolically essential cofactors NAD+ and NADP+. They can be biosynthesised *de novo* via two pathways with distinct early stages and a common latter stage. One of the two alternative starting pathways is rooted in tryptophan metabolism (Gaertner and Shetty, 1977) and starts with the metabolite kynurenine, converted to 3-hydroxy-L-kynurenine by kynurenine-monooxygenase (KMO). This is metabolised into 3-hydroxy-anthranilate by kynureninase (KYNU) which is subsequently converted to 2-amino-3-carboxylmuconate semialdehyde by 3-hydroxyanthranilic acid oxygenase (HAAO). A spontaneous reaction occurs to form the pathway intermediate quinolinate.

The alternative pathway is a two-enzyme process that starts with the metabolite L-aspartate which is converted to iminoaspartate (IA) via aspartate oxidase (Griffith *et al.,* 1975). IA is then converted to quinoliante by quinolinate synthase. From this point the biosynthesis branches converge to a common pathway. Quinolinate is processed by the enzyme quinolinic acid phosphoribosyltransferase into nicotinate mononucleotide (NM). Nicotinate-nucleotide adenylyltransferase converts NM to nicotinate adenine dinucleotide (NAD). NAD is converted to NAD+ by NAD+ synthase and this is phosphorylated to NADP+ by NAD+ kinase. This pathway is displayed in Figure 3.3b with the presence and absence results for each organism.

The Bacillariophyceae, *N. oceanica* and the two oomycetes in the analysis had homologous sequences to the enzymes in the L-aspartate starting pathway (i.e., aspartate oxidase/aspartate dehydrogenase and quinolinate synthase). Conversely, the two pelagophytes, *Ochromonadaceae sp.*, the three Labyrinthulae and *Platysulcus tardus* all had homologous sequences to the enzyme of the alternative pathway (i.e., the enzymes KMO, KYNU and HAAO).

MAST3A and MAST3F had no homologous sequences to any of the initial pathway enzymes but both had the enzyme to convert quinolinate to nicotinate mononucleotide. With the exception of *P. vesiculosus*, which had the full kynurenine pathway, the species in the Eogyra only had patchy matches to some of the enzymes in the kynurenine pathway but not the aspartate pathway.

NAD+ synthase and NAD+ kinase were widely identified across the species analysed. However, NMNAT was not identified in a number of species most notably the Bacillariophyceae, *N. oceanica* and the Labyrinthulae. This absence is curious for a number of reasons. Firstly, it is the only metabolic step capable of going from NM to d-NAD. Secondly, these species have the enzymatic steps in the proceeding pathway and finally simply due to the importance of NAD+ and NADP+ as cofactors. Indeed, if NMNAT was truly absent, it is likely to confer auxotrophy for this cofactor, yet we know exogenous sources of this vitamin are not required for growth in the Bacillariophyceae.

3.3.1.6 – Pantothenate – B_5

The vitamin pantothenate is the condensation of the intermediates pantoate and β-alanine by the enzyme pantothenate synthetase (PS). Pantoate is produced in a linear pathway from valine via the metabolites α-ketoisovalerate and ketopantoate. The enzymes involved in these reactions are branched-chain amino acid aminotransferase (BCAT), ketopantoate hydroxymethyltransferase (KPHMT) and 2-ketopantoate reductase (KPR) respectively. There are two routes to β-alanine production. A single step process sees L-aspatate converted to β-alanine by aspartate 1-decarboxylase (ADC). An Alternative pathway converts spermine to 3-aminopropanal by the enzyme polyamine oxidase (POA), then to β-alanine by aldehyde dehydrogenase (NAD+) (ALDH).

The ability to synthesise pantothenate *de novo* appears to be a universal trait within the sub-divisions of the stramenopiles (Figure 3.4a). Only four species have consistent missing enzymes throughout the pathway. *P. tardus*, MAST3F and the CHRYHs each have no homologous sequences for KPHMT, KPR or PS which, would suggest a lack of biosynthesis capacity. Although for some species like the *Ochromonadacea sp.* and MAST4C and E there may be only one missing enzyme in the series, the fact that pantothenate synthase is present would suggest that this species can produce pantothenate and the missing enzyme is likely to be a false negative caused by missing data.

Despite some evidence of presence in eukaryotes (Liu *et al.,* 2012), the enzyme ADC is found predominantly in prokaryotes. The source of β-alanine in eukaryotes is generally via the route described from the metabolite spermine. Despite this, one hit for ADC was found in

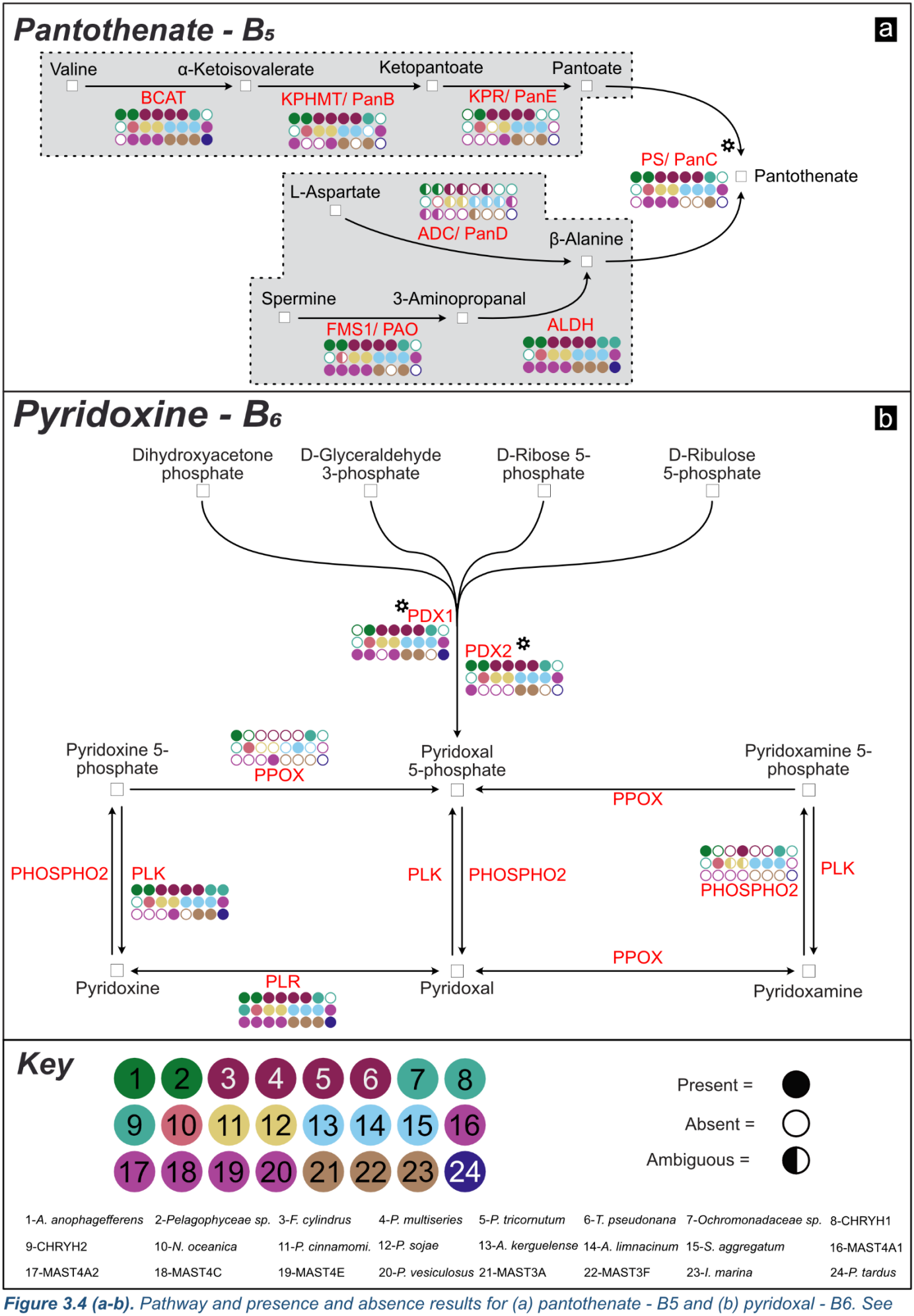


Figure 3.4 (a-b). *Pathway and presence and absence results for (a) pantothenate - B5 and (b) pyridoxal - B6. See Figure 3.2 for full description.*

pelagophyceaee sp. and was annotated by InterPro Scan to be an aspartate decarboxylase. It is indicated as ambiguous in Figure 3.4a as it was not possible from alignments or phylogenetic trees, to deduce whether this was a contaminant sequence. The alignment does, however, show that many conserved residues in other ADC sequences (obtained from NCBI) are missing in the sequence from *Pelagophyceaee sp.* suggesting it is not a *bona fide* ADC sequence.

3.2.1.7 – Pyridoxal – B_6

There are two known pathways for B_6 biosynthesis. The first discovered pathway was found in *Escherichia coli* and is known as the DXP-dependent pathway due to the derivation in deoxyxylose 5-phosphate (Tambasco-Studart *et al.,* 2005). The other pathway found initially in fungi involves two essential enzymes that produce pyridoxal 5-phosphate. These two enzymes are known as PDX1 and PDX2 and this pathway is referred to as the DXP-independent pathway. There is no evidence, in any of the species involved in this study, of homologues to the DXP-dependent biosynthetic pathway. Both PDX1 and PDX2 are required for the first step of the pathway. The majority of species analysed have both of these enzymes – the pathway and results are in Figure 3.4b. Exceptions to this are *I. marina*, *A. anophagefferens*, *P. tardus*, *P. visiculosus* and the MAST4 SAGs C and E. The phosphatase enzymes responsible for converting between the phosphorylated and non-phosphorylated forms of the B_6 molecule are not identified in many of the species. In particular pyridoxamine phosphate oxidase (PPOX) and pyridoxal phosphatase (PHOSPHO2). If these are true negatives, there would be problems for these organisms in converting between the different forms necessary for metabolism. However, these types of enzymes are notoriously difficult to identify by sequence similarity due to the relative abundance of similar enzyme types in many genomes.

3.3.1.8 – Biotin – B7

The pathway for biotin production is linear (Figure 3.5a). It goes through a four-step process from pimeloyl-CoA to 7-keto-8amino-pelargonic acid to 7,8-diaminopelargonic acid to dethiobiotin and finally to biotin. These metabolic steps are catalysed by the following enzymes: 7-keto-8-aminopelargonic acid synthetase (KAPAS), 7,8-diaminonanoate transaminase (BioA), dethiobiotin synthetase (BioD) and biotin synthase (BioB) respectively. The two-step section from 7-keto-8-amino-pelargonic acid to dethiobiotin is in some cases catalysed by a single bifunctional enzyme BIO3-1.

It is evident that the Gyrista group can biosynthesise the vitamin as well as the Labyrinthulea. It is also evident that the MAST species, both groups three and four, lack the capacity to produce this

vitamin *de novo*. They have no homologous sequences for the enzymes BioA or BioD or the bifunctional BIO3-1 (Figure 3.5a). They also show no evidence of a biotin synthase enzyme and are therefore likely to be auxotrophic for biotin. The same appears to be true for CHRYH1 and CHRYH2 although as previously stated these SAGs are of low genome coverage and the prediction of no biosynthesis capability is tentative.

Notably, as seen in the dichotomy of pathway branches in the niacin pathway, a similar trend is seen in the split between the presence of either the plant-like BIO3-1 bifunctional enzyme or the bacterial-like BioA, BioD configuration. The bacillariophycea species *P. tricornutum*, *N. oceanica*, *Ochromonadacea sp.* and the oomycetes had homology to the bifunctional enzyme whereas, if identified, other species had homology to the split configuration.

3.3.1.9 – Folate – B_9

Similar to the thiamine pathway the folate biosynthesis pathway is initiated with two branches, which converge (Figure 3.5b). Both of the initial pathways are necessary for B_9 production. The first pathway is initiated with the conversion of the metabolite GTP to 7,8-dihydroneopterin triosephosphate by the enzyme GTP cyclohydrolase (GCH1). Alkaline phosphatase converts this to 7,8-dihydronenapterin which is then converted to 6-hydroxymethyl-7,8-dihydronenapterin (HMDHP) by dihydroneopterin aldolase (DHNA). HMDHP is converted to HMDHP pyrophosphate by hydroxymethyldihydropteridine pyrophosphokinase (HPPK). The other branch of the pathway sees chorismite converted to 4-amino-4-deoxychorismate by aminodeoxychorismate synthase (ADCS). This is then converted to pABA via aminodeoxychorismate lyase (ADCL).

At this point the pathways converge and the enzyme dihydropteroate synthase (DHPS) produces dihydropteroate from the products of the two branches, HMDHP pyrophosphate and pABA. Dihydropteroate is converted to dihydrofolate (DHF) by dihydrofolate synthase. DHF can then be cycled through biologically active forms of folate such as tetrahydrofolate and folate polyglutamates by the enzymes dihydrofolate reductase (DHFR) and folypolyglutamate synthase (FPGS).

The organisms belonging to the Gyrista group appear to possess the ability to synthesise B_9 *de novo*. All the organisms in this group have a homologous enzyme at each position in the pathway. The exceptions to this pattern are the two chrysophyte SAGs, GGH in *F. cylindrus* and *T. pseudonana*. Additionally, sequences for the enzyme DHNA were undetectable in all species.

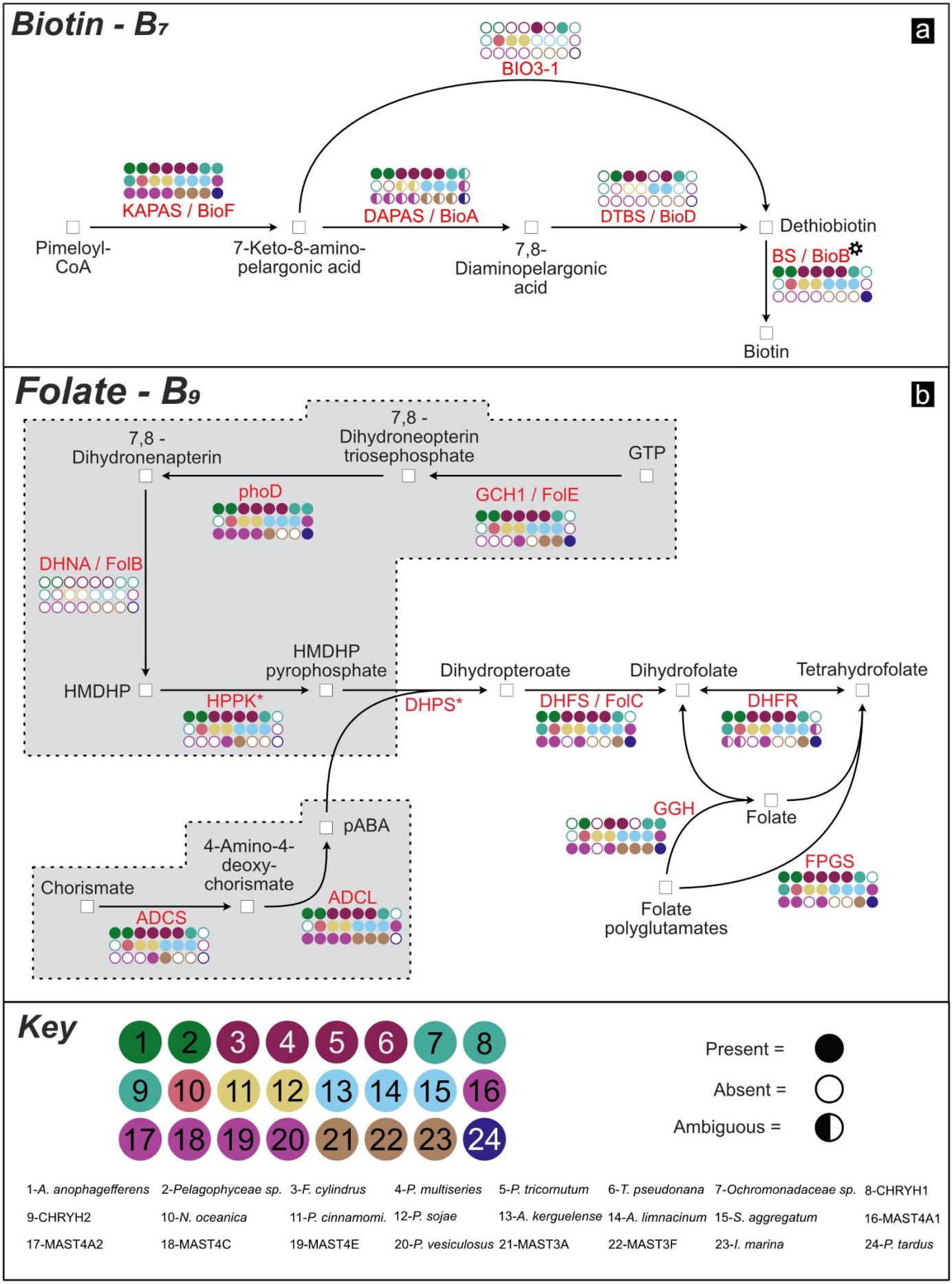


Figure 3.5 (a-b). *Pathway and presence and absence results for (a) biotin - B7 and (b) folate - B9. See Figure 3.2 for full description.*

The pattern is less clear for those organisms belonging to the Bigyra. The Labyrinthulaea show the same pattern of results as the Gyrista group. However, the MAST species have many missing enzymes in an apparently ununiform distribution. The relatively close relation *P. vesiculosus* appears to have biosynthesis capacity, which may suggest that the sporadic hits in the other Eogyrea may be due to genome completeness (the completeness average of the four MAST4 SAGs is ~45%). The story is similar for the MAST3 species. Although in the individuals there is no complete pathway, the mismatch nature of the presence and absences suggests that a fully complete genome may show a full pathway.

P. tardus has all necessary enzymes except HPPK/DHPS and those in the pABA pathway. This may be due to the nature of the transcriptome sequencing, an issue with genome coverage or a genuine missing link in the pathway. There may also be an alternative source of pABA rendering this biosynthesis branch redundant.

3.3.1.10 – Cobalamin – B_{12}

Vitamin B_{12} is exclusively biosynthesised by prokaryotes (Croft *et al.,* 2005; Martens *et al.,* 2002; Shelton *et al.,* 2019). For this reason, B_{12} auxotrophy in eukaryotes is not defined by missing biosynthesis enzymes, rather by the presence or absence of B_{12}-dependant enzymes. Methionine synthase is one of the most studied of these enzymes. There are two isoforms of methionine synthase, METH the B_{12}-dependent form and METE the B_{12}-independent isoform. An organism with only METH will be dependent on an exogenous source of B_{12} for growth (Helliwell *et al.,* 2011). We identified ten species that were predicted to have only the METH isoform of methionine synthase (Table 3.4), including both heterotrophic and phototrophic representatives.

The presence of the B_{12}-dependent enzyme methylmalonyl-CoA mutase (MCM) does not appear to cause absolute auxotrophy in algae (Helliwell *et al.,* 2011). However, it has been suggested that it may allow some photosynthetic organisms, such as *Euglena gracilis*, to survive heterotrophically by the breakdown of odd chain fatty acids (Croft *et al.,* 2006). Sequence similarity for MCM was detected in all major groups of the stramenopiles, including known B_{12} independent representatives (e.g. *P. tricornutum* and *F. cylindrus*) (Helliwell *et al.,* 2011). The only species for which no homologue was found were those with poor genome completeness.

Auxotrophy status is typically defined by the presence or absence of METE. The distribution of the presence of this enzyme did not appear to follow an evolutionary trend nor correlate with lifestyle. For example, two out of the four Bacillariophyceae were shown to be auxotrophs for B_{12} by the sole

presence of METH: *T. pseudonana* and *P. multiseries*. Whereas the other two analysed had both METH and METE genes, allowing them to live independently of exogenous B_{12}. Both oomycetes were also determined to be B_{12} independent along with *N. oceanica*, *P. tardus* and two out of three Labyrinthulea species (*Schizochytrium aggregatum* and *Aplanochytrium kerguelense*). Both MAST4A species, MAST3F, *I. marina*, *P. vesiculosus*, *Aurantiochytrium limacinum* and *Ochromonadacea sp.* were categorised as auxotrophs due to a lack of METE.

Table 3.4 Presence and absence of B_{12} associated enzymes.

Species	METH	METE	MCM
Aplanochytrium kerguelense	✓	✓	✓
Aurantiochytrium limacinum	✓	✗	✓
Aureococcus anophagefferens	✓*	✗	✓
Fragilariopsis cylindrus	✓	✓	✓
Nannochloropsis oceanica	✓	✓	✓
Ochromonadaceae sp.	✓	✗	✓
Pelagophyceae sp.	✓	✗	✓
Phaeodactylum tricornutum	✓	✓	✓
Phytophthora cinnamomi var cinnamomi	✓	✓	✓
Phytophthora soja	✓	✓	✓
Pseudo-nitzschia multiseries	✓	✗	✓
Schizochytrium aggregatum	✓	✓	✓
Thalassiosira pseudonana	✓	✗	✓
CHRYH1	✗	✗	✗
CHRYH2	✗	✗	✗
MAST3A	✗	✗	✗
MAST3F	✓	✗	✓
MAST4A1	✓	✗	✗
MAST4A2	✓	✗	✓
MAST4C	✗	✗	✓
MAST4E	✗	✗	✓
Incisomonas marina	✓	✗	✓
Platysulcus tardus	✓	✓	✓
Pseudophyllomitus vesiculosus	✓	✗	✓

3.3.1.11 – Predicting biosynthesis capabilities from presence and absence of pathway genes

From the results described above and displayed in Figures 3.2-5 the ability of each organism to biosynthesise each vitamin was predicted. This was achieved by two complementary methods. The first took an empirical approach, analysing the number of missing enzymes in a given vitamin pathway for a given species. It also considered the nature of these absentees, including in each pathway the key enzymes required (e.g., synthase enzymes, key enzymes are indicated in Figures

3.2-5 with an asterisk) and assessing the importance of that missing enzyme and its propensity to be identified by sequence similarity. For example, in Figure 3.5a, the organism *P. vesiculosus* has a homologue to the first pathway enzyme but not to the final enzyme biotin synthase, a key enzyme. The prediction would be that *P. vesiculosus* does not biosynthesise B_7. On the other hand, *P. tardus* does have a homologous sequence to biotin synthase. Although the intermediate enzymes DAPAS/DTBS are missing the prediction in this case would be that *P. tardus* is likely to biosynthesise B_7 as the key enzyme biotin synthase is present.

The second approach aimed to add a level of statistical reasoning to the predictions. Here, a hypergeometric test was employed which allowed the predictions to take into account the level of missing data for each organism on top of the presence and absence of each pathway enzyme. To achieve this a number of parameters were defined , N is the size of full genome for the organism (calculated by extrapolation of number of genes based on BUSCO completeness estimation), K is the minimal number of enzymes required for biosynthesis of the given B vitamin, n is the number of entries in the fasta file for the organism in question and k is equal to the number of confirmed pathway enzymes found in that organism's genomic data for that B vitamin. The following equation was used to calculate a p-value with the null hypothesis that the given B vitamin biosynthesis pathway is complete in the given organism:

$$px(k) = \Pr(X = k) = \frac{(\frac{K}{k})(\frac{N-K}{n-k})}{(\frac{N}{n})}$$

The resulting p-values can be used to help categorise non-producers as the lower the p-value the greater the confidence in rejecting the null-hypothesis that the pathway is complete. The limitation of this statistical approach is that each enzyme is treated equally, whereas we know that some enzymes are difficult to identify by their nature (common sequence motifs, high numbers of similar enzymes) and other enzymes might be more diagnostic, such as key enzymes like biotin synthase discussed above. For these reasons, the predictions from each method should be interpreted together. The full predictions from both approaches can be seen superimposed in Table 3.5.

Table 3.5. Predictions of biosynthesis capacity for each B vitamin (except B_{12}) in each species. The colour of the cell reflects the likelihood of biosynthesis capability as predicted from the presence and absence of biosynthesis pathway genes in the genomic data for each organism. The number is the p-value from a hypergeometric test and represents the likelihood that a pathway is complete given the enzymes found and the amount of missing genomic data predicted by BUSCO where the null-hypothesis is that the pathway is complete (therefore the lower the p-value the greater the likelihood the pathway is not functional).

Species	B_1	B_2	B_3	B_5	B_6	B_7	B_9
A. anophagefferens	0.4556	0.2833	1	**0.6228**	0.2774	0.4779	0.4556
Pelagophyceae sp.	0.3351	1	1	1	1	1	0.3351
F. cylindrus	0.3350	0.2484	0.2171	1	1	1	0.0581
P. multiseries	**0.0000**	1	**0.0000**	1	1	1	**0.0000**
P. tricornutum	0.4611	1	0.3098	1	1	1	0.4611
T. pseudonana	0.3349	0.2483	0.2171	1	1	1	0.0580
Ochromonadaceae sp.	**0.0009**	0.7903	1	0.7378	1	1	0.8926
CHRYH1	0.3574	**0.0248**	0.6907	0.2179	0.3481	0.4580	0.1515
CHRYH2	0.2919	0.1603	0.4959	0.2083	0.5929	0.7716	0.8207
N. oceanica	0.1829	1	0.1141	0.1141	1	1	0.1829
P. soja	**0.0000**	**0.0000**	1		1	1	0.0955
P. cinnamomi var c.	**0.0000**	0.2484	1	0.2171	1	1	0.3350
S. aggregatum	**0.0000**	1	0.4832	0.4320	1	1	0.6106
A. limacinum	**0.0000**	1	0.3514	1	1	1	0.4613
A. kerguelense	**0.0000**	0.3516	0.3516	1	1	1	0.4615
MAST4A1	0.1516	0.9981	0.8938	0.9953	1	0.4580	0.8166
MAST4A2	**0.0419**	0.7394	0.4563	0.6180	1	0.2831	0.5730
MAST4C	0.0620	0.2432	0.9438	0.8998	0.7599	0.3276	0.6474
MAST4E	**0.0479**	1	0.2103	0.8810	0.2400	0.2976	**0.0479**
P. vesiculosus (MAST6)	0.0712	0.9779	1	1	0.6636	0.2030	0.9957
MAST3A	0.1234	0.6501	0.1226	0.9937	1	0.4240	0.7793
MAST3F	0.5956	0.8466	0.2885	0.7064	1	0.6162	0.5956
I. marina (MAST3)	**0.0239**	**0.0017**	0.8785	1	0.0676	0.0566	0.2479
P. tardus	**0.0185**	0.9394	1	0.0969	0.5511	0.4015	0.4316

	All or almost all enzymes missing, confident in lack of biosynthesis
	Many enzymes missing including essential enzymes, likely lack of biosynthesis
	Almost all present, a few missing enzymes, likely to biosynthesise
	All essential enzymes or all present, confident that biosynthesis capacity is present

3.3.2 – Class scale analysis

In curating the pathways and analysing the select group of organisms detailed above some interesting patterns emerged hinting at more substantial patterns across the wider group. To investigate whether these patterns held true for a greater number of species, or if they were an artefact of a small sample size, a large-scale analysis was conducted. This was achieved by compiling a dataset of 321 accessions (with at least 40% completeness as measured by Anvi'o) from a combination of the MMETSP (Keeling *et al.,* 2014), TARA oceans SMAGs (Seeleuthner *et al.,* 2018), JGI genomes (Nordberg *et al.,* 2014) and Thakur et al. (2019) transcriptomes.

Analysing the presence and absence of pathway genes at this scale allows more robust patterns to be identified across the stramenopiles as a whole, especially when the organisms are grouped at different taxonomic ranks. However, the higher number of accessions involved in the analysis make it more difficult, if not impossible, to manually verify every hit raising the likelihood of false negative and positives. That said, the following analysis should not be scrutinised at the level of an individual but at the group level discussed. For continuity the same enzymes defined and searched for in the previous section were used for this analysis although presence and absence were determined by hmmer searches of KEGG orthologues (Kanehisa *et al.,* 2016a) (as opposed to the use of BLAST utilised in Section 3.3.1).

3.3.2.1 – Thiamine B_1

The results of the thiamine analysis at the large-scale largely corroborates that of the initial small-scale screen with a few exceptions and are presented in Figure 3.6. There are still very few hits to the pathway enzymes in the Oomycota, Labyrinthulea, Bikosia, Nanomonadea (MAST3) or the Platysulcea. Collectively these groups will be referred to as the early diverging lineages (EDLs) as they diverged prior to the plastid acquisition event that gave rise to the Ochrophyta. The only enzymes consistently found across the EDL classes are iscS and TPK. This is unsurprising given their essential roles in metabolism; iscS is also involved Fe-S cluster formation (Schwartz *et al.,* 2000) and TPK is the enzyme that converts thiamine to the biologically active form thiamine pyrophosphate (TPP). The lack of homologues to thiamine biosynthesis enzymes in the EDLs suggests that the earliest ancestor of the stramenopiles lacked the ability to perform *de novo* biosynthesis of B_1, and that the ability of some species in the Ochrophyta, in particular within the Bacillariophyceae, Eustigmatophyceae, Raphidophyceae and Xanthophyceae has been an acquired trait.

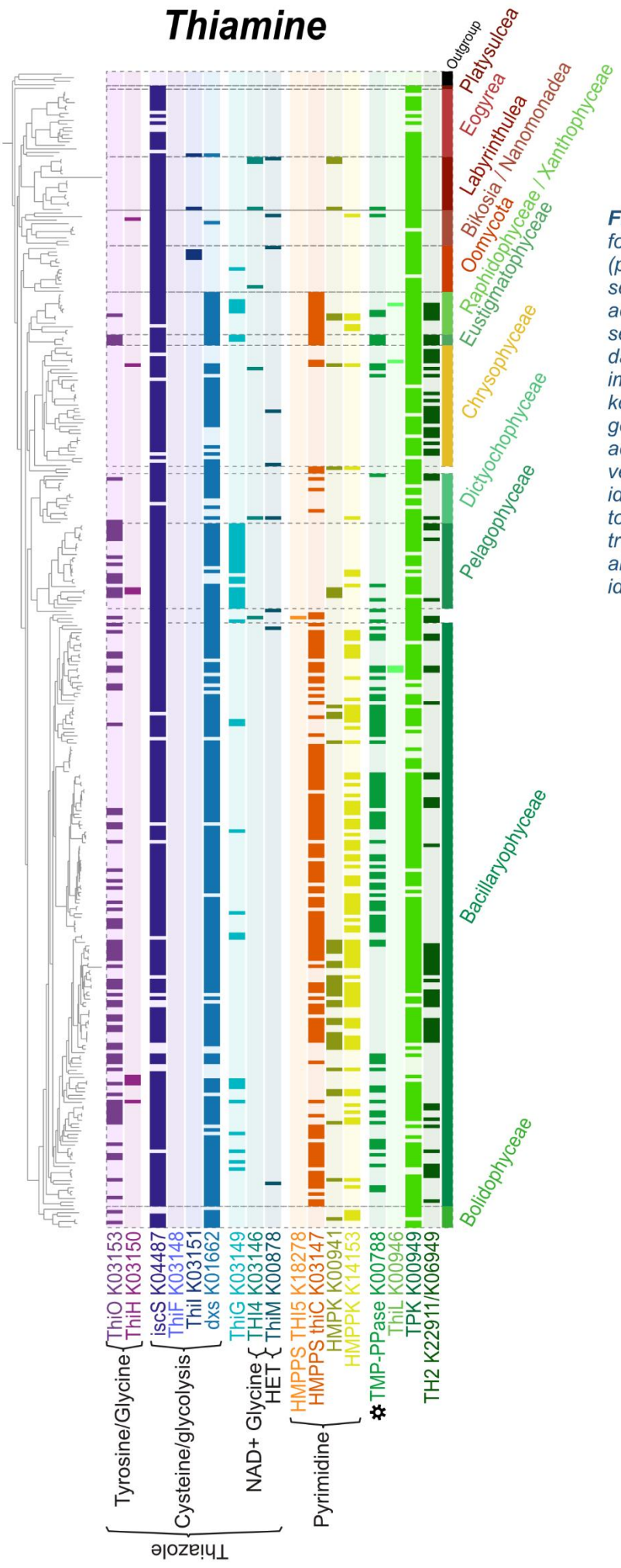


Figure 3.6. *Presence and absence results for the enzymes of the thiamine pathway (pathway presented in Figure 2). Large scale analysis of 321 stramenopile accessions where homologues were searched for using the KEGG orthologue database and HMMER search, implemented by Anvi'o (anvi-run-kegg-kofams). Where a KO is present in the genomic/transcriptomic data for an accession, it is represented by a solid vertical bar. When no homologue was identified the bar is absent. The tree at the top of the figure is a maximum-liklihood tree built with IQTree using a concatenated alignment of single-copy orthologues identified using Anvi'o.*

The form of HMPPS encoded by the NMT1/THI5 gene, typical of fungi (Wightman and Meacock, 2003) was elusive in the large-scale analysis despite being identified in each of the Pelagophyceae and Bacillariophyceae species in the initial screen. This is likely due to an overly stringent threshold for this HMM within the KO fam database. This is a prime example of the limitations of this approach and the difficulty to curate false negatives in an effort to minimise false positives. The same is true for ThiF which was found almost ubiquitously in the preliminary analysis but scarcely identified in this analysis.

3.3.2.2 – Riboflavin B2

Riboflavin synthase (RS) was found in the majority of species in each of the classes analysed (Figure 3.7a). The presence of this key enzyme would suggest that the ability to perform biosynthesis is also widespread in the stramenopiles despite, in some cases, not identifying other enzymes in the pathway. DMRLS and GTPCH2 were also widely identified. pyrP and RIBD1 and 2 were not identified in many species, however these enzymes also proved difficult to identify with confidence in the initial screening, even with manual curation of weak hits. The most obvious difference between the results of the initial screen and the large-scale analysis is the presence of DHBPS which only had a positive hit in two species in the large-scale analysis despite being identified in 17 of the 24 species in the initial screen. This again could be due to an overly stringent threshold or a poorly representative HMM for the sequence of this enzyme.

3.3.2.3 – Niacin B_3

The pattern identified in the initial screen that suggested a dichotomy across the group with regard to initial niacin biosynthesis pathway (i.e., aspartate versus kynurenine pathways) was upheld and reinforced by the large-scale analysis as displayed in Figure 3.7b, illustrated by the red and green tracks, which represent the aspartate and kynurenine pathways respectively. One enzyme was an exception to the trend of a distinctive split between pathway alternatives. Aspartate dehydrogenase (AD) was identified in some members of the EDLs who otherwise displayed evidence of the kynurenine pathway. Generally speaking, the classes Platysulcea, Eogyra, Labyrinthulea, Bikosia, Nanomonadea, Chrysophycea and Pelagophyceae and half of the Oomycota displayed homologues to the enzymes in the kynurenine-initiated branch. The classes Bacillariophyceae, Bolidophyceae, Raphidophyceae, Xanthophyceae and the other half of the Oomycota displayed homology to the enzymes in the L-aspartate initiated pathway. Two groups,

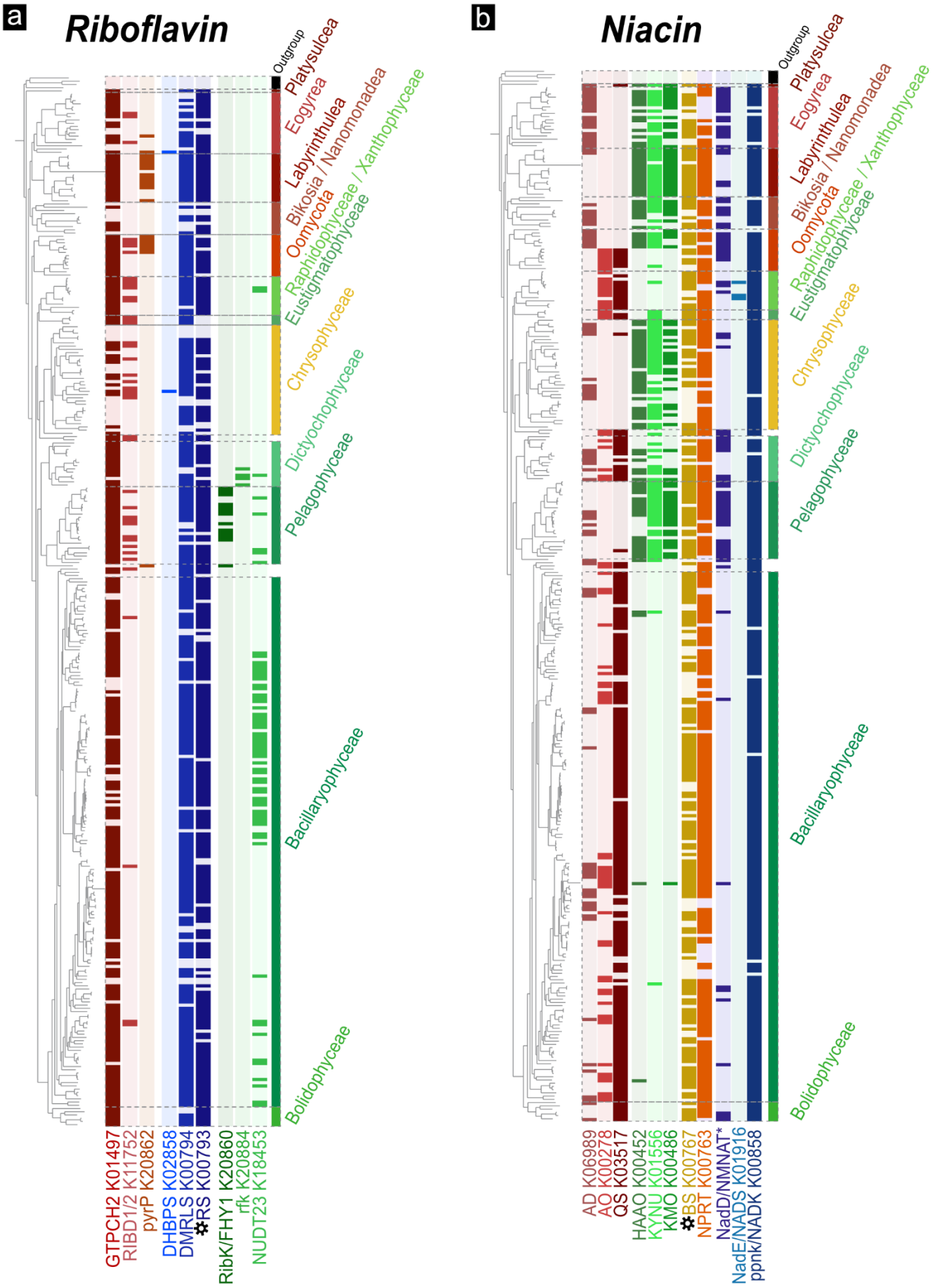


Figure 3.7 (a-b). *Large scale presence and absence results for (a) riboflavin - B2 and (b) niacin - B3. See Figure 6 for full description.*

the Dictyochophyceae and the Platysulcea hinted at members of the group having both pathways present, something not previously reported (Ternes and Schönknecht, 2014). The advantage of having both pathways is not clear and may be evidence of transition from one to the other in these lineages.

3.3.2.4 – Pantothenate B_5

The enzymes required for the production of the intermediate metabolite pantoate are found almost ubiquitously in every group analysed (BCAT, KPHMT, KPR) (Figure 3.8a). The same is true for the critical final enzyme pantothenate synthase (PS) indicating that in general all stramenopiles have the capability of *de novo* B_5 biosynthesis. The enzymes required for β-alanine production, however, were elusive in this analysis despite homologues being identified readily in the species in the preliminary screen (a homologue for POA was identified in 19/24 species and 23/24 for ALDH, Figure 3.4a). As previously discussed, this is likely a systematic error in thresholding or HMM modelling. Despite this it is likely that all stramenopile groups biosynthesise B_5.

3.3.2.5 – Pyridoxal B_6

Although it may not be true for a given individual, the majority of species from across the group had homologues to the enzymes required for pyridoxal biosynthesis as well as those required for cycling of isoforms of the vitamin (Figure 3.8b). There was no clear indication at this macro level of classification, that pyridoxal biosynthesis had been lost or was not present, suggesting that the last common ancestor to the stramenopiles also had the ability to perform B_6 biosynthesis *de novo.*

3.3.2.6 – Biotin B_7

The critical final enzyme in the biosynthesis pathway for B_7 is biotin synthase. Homologues to this diagnostic enzyme were identified readily in the plastid containing groups as well as the Labyrinthulea (Figure 3.9a). In contrast, it was identified in only one accession in the Bikosia, a few in the Chrysophyceae and in only half of the Oomycota group. DTBS was not regularly found in either sets of analysis although, given that it was identified in the Bacillariophyceae in the initial screen and many of these organisms are known not to require exogenous biotin for growth in axenic culture, it is surprising how few hits were found. This and the low level of identification for KAPAS (readily identified in the initial screen) are attributed to the limitations outlined in previous sections.

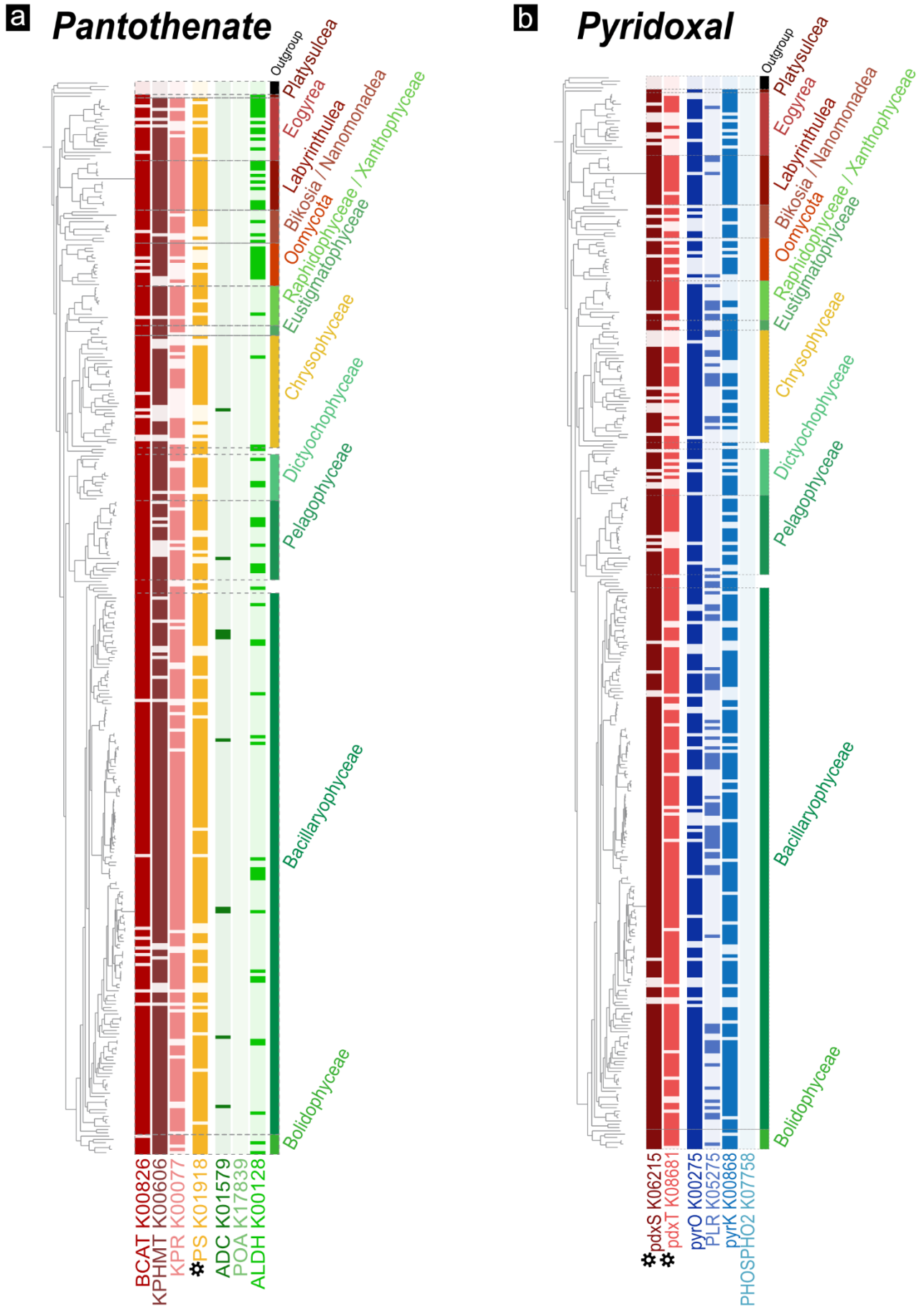


Figure 3.8 (a-b). *Large scale presence and absence results for (a) pantothenate - B5 and (b) pyridoxal - B6. See Figure 6 for full description.*

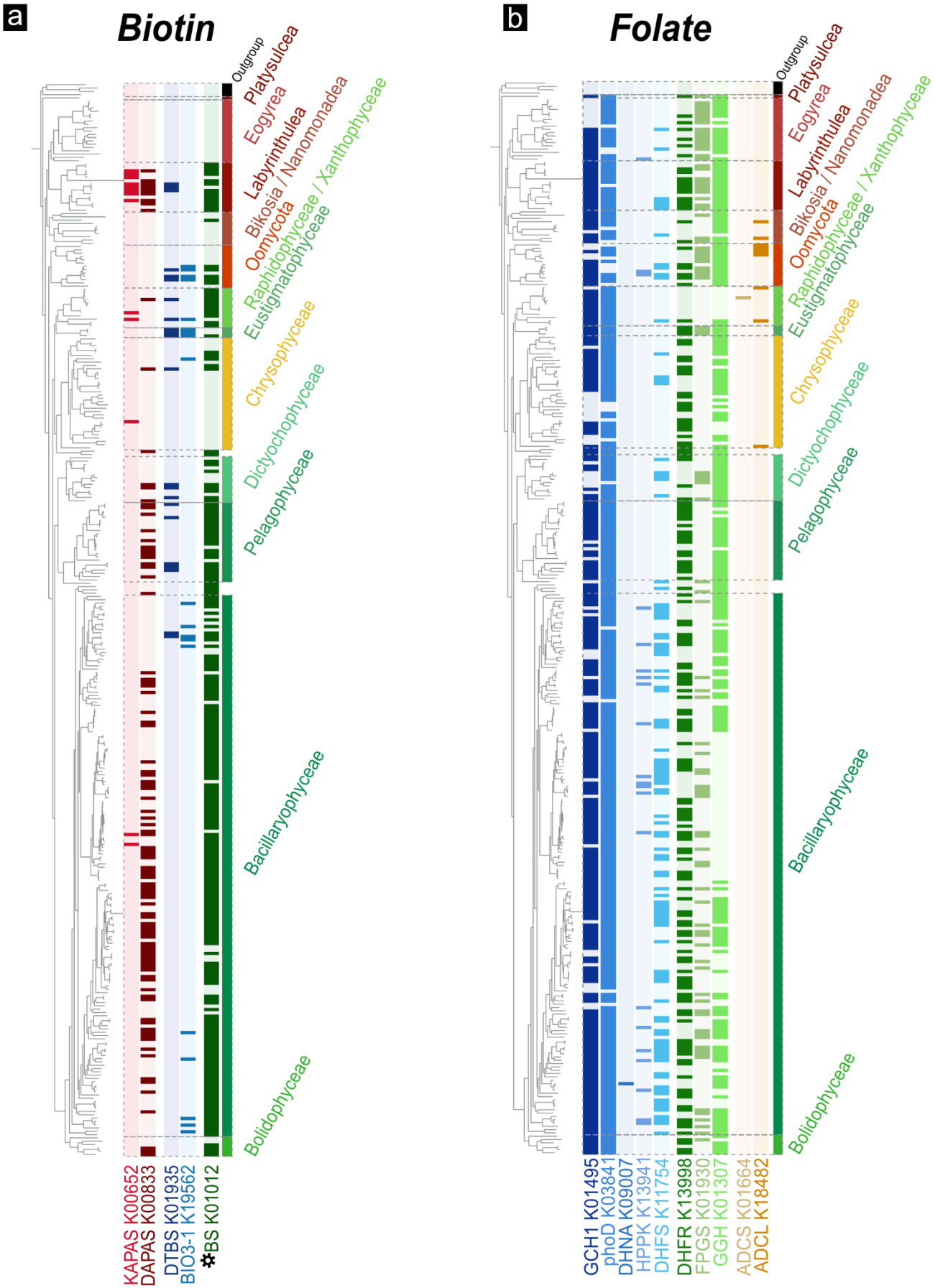


Figure 3.9 (a-b). *Large scale presence and absence results for (a) biotin - B7 and (b) folate - B9. See Figure 6 for full description.*

3.3.2.7 – Folate B_9

With the exception of the enzymes required in the pABA branch of the pathway, DHNA and HPPK homologues to the enzymes of folate biosynthesis are found in all groups analysed, as displayed in Figure 3.9b. No discernible patterns were identified when comparing one group to another suggesting that there have been no historic loss events of the full pathway in the evolution of the extant stramenopiles. Species from each class are likely to biosynthesise B_9 and any non-producers are likely to have evolved independently. The frequency with which DHFS was identified demonstrates that the majority of the species analysed have the ability to form the biologically active tetrahydrofolate.

3.3.2.8 – Cobalamin B_{12}

Homologues to METH are found in almost every accession in the large-scale analysis (91.6%). However, instances of METE identification were less common, present in only 93 of 321 accessions. The instances of METE were spread across the supergroup with species in every class displaying B_{12}-independence with the exception of the Eogyra where none of 19 species encoded METE. The percentage of species that only had a METH present (making them obligate B_{12} auxotrophs) was 64%. MCM was also found in all classes across the group and did not correlate with the presence or absence of the two isoforms of methionine synthase. Almost half of accessions (45.8%) had a homologue to MCM. These results are displayed in Figure 3.10.

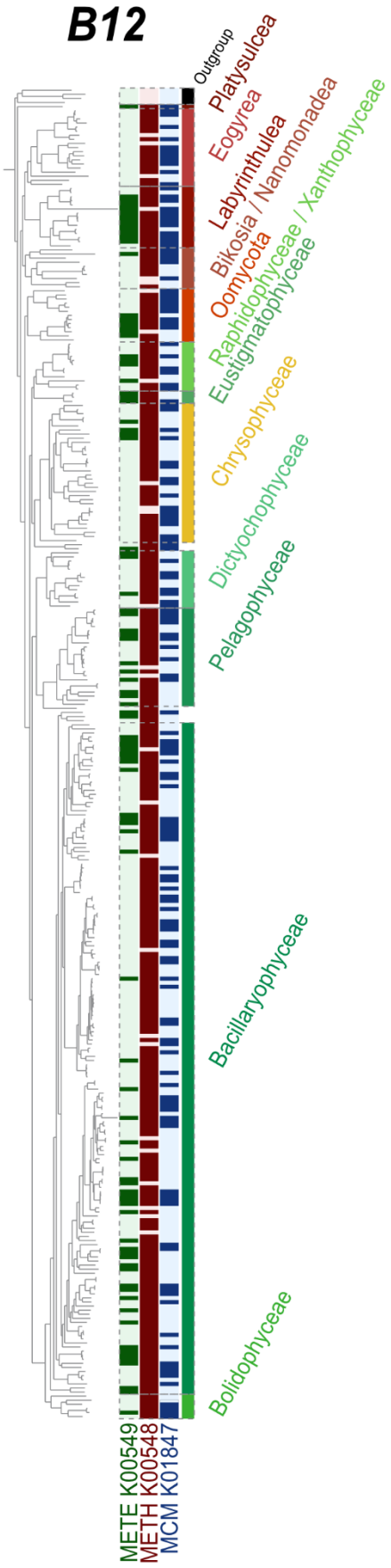


Figure 3.10. *Large scale presence and absence results for cobalamin - B12. See Figure 2 for full description.*

3.4 – Discussion

Across the stramenopiles, many traits show high levels of variation. This includes morphological traits such as the number of flagella possessed by a cell as well as lifestyle traits such as heterotrophy versus phototrophy. Here I have demonstrated that the traits of B vitamin biosynthesis for a number of the B vitamins also display high multiplicity that reflects the group diversity.

According to the large-scale analysis the ability to perform *de novo* biosynthesis of B_2, B_3, B_5, B_6 and B_9 is generally ubiquitous. Although for a given individual accession there may not have been a homologue found for every enzyme in the pathway of each of these vitamins, the patterns observed suggest that for these vitamins the ability to synthesise is a common trait and has been for a long period of evolutionary time.

Thiamine – The results from the large-scale analysis for B_1 biosynthesis are difficult to interpret. Many diatoms, from culturing, are known to be prototrophic for thiamine. However, the identification of homologues for many of the thiamine biosynthesis enzymes is inconsistent across the accessions analysed in this study. This may be due to various systematic reasons (see below). In the EDL groups however, almost none of the enzymes were identified in any of the accessions. This again demonstrates that the rate of auxotrophy in heterotrophic species is higher than that of phototrophs. This is perhaps not surprising given the relative complexity of the molecule in enzymatic steps if not in structure and the likely availability of the vitamin in the diet. The ability to biosynthesise B_1 appears to be unique, within the stramenopiles, to the classes Bacillariophyceae, Eustigmatophyceae, Raphidophyceae and Xanthophyceae based on both the large scale and initial screening analysis. The fact that these lineages all occur after the purported plastid acquisition event suggests this is a trait that might have been acquired in the same event. However, there is a low number of accessions involved in the large-scale analysis for the latter three of these groups and if this was to be a trait acquired with the plastid acquisition, subsequent gene loss events must have occurred to explain the apparent lack of B_1 biosynthesis in a the Pelagophyceae and the Chrysophyceae.

Niacin – The B_3 pathway in particular displays a fascinating dichotomy in the utilisation of alternative early pathway routes, the aspartate pathway or the kynurenine pathway. Here I demonstrate that in almost every class, the enzymes of one of these routes is found and none of the other. The exception to this is the Dictyochophyceae and the Platyculsea who have members with

homologues to the enzymes of both pathway routes. This may represent a transitional state, where recent acquisition of the aspartate pathway enabled selective loss of genes from the kynurenine pathway. This warrants further investigation, including analysis of the homologous sequences to identify any evidence that one or other of the pathway routes is being lost (pseudonisation of genes).

That groups have one or the other pathway corroborates previous analysis made by Ternes and Schönknecht (2014). However, this analysis complicates the story further. Where they suggest the oomycete group utilise the aspartate pathway, in this analysis I show that although this is true for the well-studied soil pathogen oomycetes the opposite is true for the oceanic species assembled as MAGs from TARA. Based on analysis of the sequences of individual enzyme homologues Ternes and Schönknecht (2014) propose a series of putative horizontal gene transfer (HGT) events to explain the dichotomic nature of pathway utilisation across eukaryotes. However, they suggest that a HGT occurred within the stramenopiles at the branching point between the Gyrista and the Bigyra. In Figure 3.7b, it is apparent that half the oomycetes analysed have the aspartate pathway and the other half the kynurenine one. This split corresponds to the terrestrial and the oceanic oomycetes respectively suggesting and indicating that this HGT occurred later in time or multiple HGT events occurred independently within this class and that this transition from one pathway to the other may have coincided with the lifestyle change.

Biotin – When assessed by Croft et al. (2006) only around 5% of algal species required exogenous B_7 for growth. The results here suggest that the proportion of heterotrophic species that are auxotrophic for biotin is much higher. Of the 29 accessions of Eogyra, Bikosia and Nanomonadea, only one had a homologous sequence to biotin synthase. *P. tardus* also had no homologues to any of the B_7 pathway enzymes. This would suggest that the common ancestor of the stramenopiles also lacked biotin synthesis capabilities and that the genes required for the *de novo* biosynthesis pathway have been subsequently acquired by later evolving lineages. The fact that the Labyrinthulea appear to have biotin synthase, but the other basal lineages do not, suggests this might be an independent HGT event to the one that might have given rise to B_7 biosynthesis in the Gyrista. Equally, the lack of biotin synthase identification in the majority of the chrysophytes might suggest more independent gene acquisition events or subsequent loss events, or both. Indeed it has been suggested that this is the case in strains of yeast that display different auxotrophy states for biotin (Hall and Dietrich, 2007).

Cobalamin – Previous analysis has shown that instances of B_{12} auxotrophy (sole presence of METH isoform of methionine synthase) is spread across the EToL in a polyphyletic manner (Croft *et al.,* 2005; Helliwell *et al.,* 2011). Here I demonstrate, on a previously unparalleled scale, that this is true within the stramenopiles too. The entirety of the Eogyra group appears to be auxotrophic for B_{12} and, in the Oomycota class, as with the dichotomy in niacin pathways, half the group appears to be B_{12} dependent and the other independent. As mentioned previously this dichotomy matches the lifestyle split in this group with the oceanic species being auxotrophic for cobalamin and the soil pathogen species encoding METE as well as METH. This renders them B_{12} independent, mirroring their plant hosts and may be indicative of the availability of B_{12} in these environments (E. M. Bertrand *et al.,* 2012). One would speculate that METE must have been gained or lost from one of these groups. Gain of METE might have occurred by HGT in the rhizosphere, allowing these organisms to survive in a B_{12} deplete environment. Alternatively, METE may have been lost from the oceanic species when B_{12} was abundant.

3.4.1 – Limitations and mitigations

Any analysis based purely on bioinformatic analysis, as with the work detailed in this chapter, has limitations. Here I will discuss these limitations and their implications for the interpretation of these results as well as possible mitigation steps.

The two approaches used here, the initial small-scale screen and the large-scale analysis, are limited in different ways. However, the use of both of these approaches in parallel does mitigate some of the limitations of each individual approach. The initial screen was limited for a number of reasons. Firstly, the number of accessions that were used in the analysis does not allow for robust patterns to be identified across the groups. This was addressed by conducting the large-scale analysis which increases the number of accessions by over 12 times. This did to an extent allow for the identification of more robust patterns in presence and absence of enzymes at the class level. However, the data set is still biased towards the Bacillariophyceae. As discussed in Chapter 1, this is a well-established research bias and the sooner this is addressed the sooner more robust analysis can be carried out.

A limitation of both studies is the quality of the data to start with. Although many of the accessions used in the initial screen have genome completeness levels approaching 100% (by BUSCO analysis), others, namely those of the less well studied groups within the Bigyra, are closer to 40% or worse. This leaves a level of ambiguity in the analysis of an individual as the inability to identify a

gene encoding a particular enzyme may be due to missing data rather than true absence. This is mitigated to an extent by increasing the number of accessions analysed in a class. Assuming the missing data in each genome is random when analysing, six very closely related species for example, and an enzyme is purportedly missing in all of them, the likelihood of this being a genuine absence is higher than when this absence is observed in just one of the six. This assists in identifying patterns across classes but does not improve the analysis at the individual level. To have greater trust of the results at an individual level there must be an improvement in the genomic data available (closer to 100% completeness). That said, BUSCO analysis for completeness is not without its own limitations, especially for lesser-studied organisms.

The large-scale analysis is limited by the nature of the homologue search. In the initial in-depth study, with fewer accessions to analyse, it was possible to manually verify every hit for each enzyme from each accession. This was achieved by analysing bit scores and e-values as well as subsequent supplementary analysis such as protein domain prediction and cross referencing as well as phylogenetic tree building and multiple-sequence alignment where necessary. Given the number of accessions and queries used in the large-scale analysis, this level of scrutiny was not feasible (subsequent detailed analysis could be undertaken in cases of interest such as the results of the B_3 pathway). The analysis pipeline used for the large-scale analysis utilised hmmer (Eddy, 2011) search with the KEGG orthologue database as queries. This itself has limitations. Firstly, the KO database is a collection of HMMs for each protein sequence of interest. These HMMs are built with available confirmed sequencing data for a given protein, given enough similarity between sequences. The implication for this in cases where the given protein has not been readily identified in a variety of species, or where the protein is highly variable in sequence, is a poorly representative HMM that may not allow for the identification in an organism that is distantly related to those from which the sequences to make the HMM were taken. This is constantly updated and improved with experimental validation of sequences when they become available, however, it will still present an issue when attempting to identify sequences in a novel genome. There is also a predetermined bit-score threshold level for each HMM which may be too stringent for distantly related sequences, resulting in false negatives. This is particularly apparent when comparing the results of the two approaches. For example, in the in-depth study I was able to identify homologues to THI5 from the thiamine biosynthesis pathway in each of the Bacillariophyceae. However, none of these were identified in the large-scale analysis. Another example is the KAPAS enzyme of the B_7 pathway. This was identified in every accession in the initial study but scarcely at all in the large-scale analysis.

These issues will only be improved with a concerted effort to characterise the protein complement of a greater variety of species and rectify the research bias that is currently reflected in these types of databases.

The limitation outlined above are some of the reasons for which the threshold of 40% was chosen when selecting the data for the large-scale analysis. If the threshold had been at 70% the data set would have been reduced by ~20% and the reduction would have had a disproportionate effect on the earlier diverging lineages, for which we were most interested in for this study (for example, ~35% of MAST4 species would be excluded compared to only ~15% of Bacillariophyta). Presence and absence was assessed in the same way at this higher threshold and the results for Biotin and Niacin can be seen in Appendix Section 2.1. These figures demonstrate that the patterns of presence and absence are already evident at this higher completeness threshold and are not introduced by the lower completeness genomes.

Some possible further explorations of these results would be to identify the likelihood of the HGT events discussed above. This might be done with a phylogenetic approach and a multiple sequence alignment approach. This would require a large number of sequences from a range of species both eukaryotic and prokaryotic to identify the likely source of the transferred genes. To do this for all of these pathways would be a large undertaking due to the numerous putative HGT events. Further to this a deeper exploration of the evolution of these traits, both at the pathway scale and at the scale of presence or absence of each individual protein, could be carried out with an ancestral state reconstruction. This could provide insight into the potential trophic modes of ancestors or even their likely habitats. For example, a hypothesis might be that the ancestor of the oomycetes was marine dwelling, an ancestral state of B_{12} dependence might indicate this as the extant species that live as terrestrial plant pathogens are B_{12} independent presumably as a result of limited B12 sources on land.

As with any results obtained by bioinformatic analysis, the predicted auxotrophy status of the species should be experimentally verified. This may be possible for a number of the species involved in the initial in-depth study. However, many of the accessions analysed in the large-scale analysis are uncultured. This reiterates the need to improve our understanding of culturing the currently unculturable.

Despite these limitations, the work here represents the potential for utilising careful manual analyses with large-scale analysis platforms, such as Anvi'o, to gain meaningful insight into the

genomics and underlying biology of these elusive organisms. It also shows how iterative approaches can yield results as well as highlight the limitations as well as the potential of using this type of analysis.

4 - Physiology, Imaging and Metagenomics of the MAST3 species *Incisomonas marina* and associated bacterial consortia

4.1 Introduction

4.1.1 MArine STramenopiles

As previously discussed, the stramenopile group of the EToL harbours a large number of taxa who demonstrate a wide range of lifestyles and inhabit marine, freshwater and terrestrial environments. A large proportion, if not the majority, of this diversity comes not from the traditionally well-studied photosynthetic lineages within the group but from the heterotrophic lineages. This is demonstrated primarily with environmental sequencing efforts. Amplicon sequencing of the 18S rRNA gene from oceanic sampling has identified previously unidentified species. These were phylogenetically consolidated into 18 distinct monophyletic "ribogroups", collectively named MASTs (MArine STramenopiles) (Massana *et al.,* 2004; 2014). These groups have routinely been detected in environmental sequencing efforts, including the global TARA oceans expedition (Bork *et al.,* 2015; Sunagawa *et al.,* 2015; de Vargas *et al.,* 2015; Seeleuthner *et al.,* 2018) and in localised experiments (e.g. Taylor and Cunliffe 2014). Similarly, the mesocosm experiments described in Chapter 2 found stramenopiles to be one of the top 10 families identified in this environmental sequencing (Figure 2.10) and MAST genera to be among the top 70 abundant in the experiment (Chapter 2 Appendix Section 1.7). These MAST groups are predominantly thought to be heterotrophic and reside in the basal lineages of the stramenopiles, although Massana et al. (2014) also identified five novel marine ochrophyte groups, one of which was also determined to be heterotrophic despite phylogenetically grouping in the lineages whose evolution arose after the plastid acquisition event in the stramenopile history.

Only a handful of species from these MAST groups have been isolated and successfully cultured. *P. vesiculosus* (MAST6) a phagotrophic bi-flagellate (Shiratori *et al.,* 2017) and *P. tardus,* a gliding flagellate with typical stramenopile flagella (two of unequal length), which in phylogenomic studies appears sister to all stramenopiles (Shiratori *et al.,* 2015; Thakur *et al.,* 2019). Kolodziej and Stoeck (2017) also described a MAST12 species from an environmental sample and performed SEM imaging on the species but used culture-independent techniques to do so. There is only one publicly available MAST in culture. This is *I. marina* (MAST3) (CCAP 977/1) isolated from an estuarine environment in Nova Scotia, Canada (Cavalier-Smith and Scoble, 2013).

As a consequence, little genomic, physiological or metabolic information is available for these organisms. Thakur et al. (2019) produced transcriptomic data sets for *I. marina*, *P. vesiculosus* and *P. tardus.* Derelle et al. (2016) carried out a genomic sequencing effort on *I. marina* however despite publishing the raw read data, no assembly has been published. Instead, the majority of sequence information for these groups comes from MAGs and SAGs recovered from TARA oceans samples (Delmont *et al.,* 2021; Seeleuthner *et al.,* 2018). Indeed, these data have been analysed and displayed in Chapter 3 Section 3.3.

4.1.2 – Nutrient exchange, interaction and signalling across kingdoms and within communities

In the environment, stramenopiles, just as all microbes, live not in isolation but as members of communities, which include both prokaryotic and eukaryotic members. Microbial interaction is a vital component to these communities being able to function at full capacity. Interactions can be categorised into predatory, mutualistic or parasitic but more broadly they can be described as active or passive (Kazamia *et al.,* 2016). A passive interaction would constitute the uptake of nutrients from the dissolved pool of nutrients that reached this pool by "leaking" from living cells or as a result of cell lysis. Active interactions would include trophic interactions as well as symbiotic interactions where both parties actively engage in the interaction. Where these are positive interactions, they are defined as mutualism, for which many examples have been identified. Often the interaction is based on the provision of fixed carbon by an alga to a bacterium and a different nutrient in the reverse direction. An example of this is the association of some cyanobacteria who provide fixed nitrogen to their diatom associates (Amin, Parker, and Armbrust 2012). Iron is a key component of metabolism for almost all cells. However, its relative scarcity in the environment means various strategies to acquiring this nutrient have evolved. One suggested strategy might be associating with vibrioferrin producing bacteria such as *Marinobacter* spp. Vibrioferrin is a stable siderophore that can supply algal cells with soluble Fe(III), likely in return for organic carbon (Amin, Parker, and Armbrust 2012; Amin *et al.,* 2009).

Mutualisms of this nature also occur for vitamin exchange. Examples of which are discussed in Chapter 1 Section 1.1.3 but include for example the exchange of vitamin B_{12} for organic carbon which has been documented for the red alga *P. purpureum* and the bacterial species *Halomonas sp.* (Croft *et al.,* 2005). Direct interactions like this may be essential in the environment as many algal species have been determined to be auxotrophic for a variety of vitamins, which are found in

the environment at concentration below the required level for growth (see Chapter 2, (Croft *et al.,* 2006)). Correlation and interaction analysis also demonstrate the close coupling of eukaryotic and prokaryotic microbes (Lima-Mendez *et al.,* 2015).

As well as these B vitamin-based exchanges (introduced fully in Chapter 1 Section 1.3.3), amino acid metabolism is likely to be another source of nutrient exchange. Many organisms do not produce the full suite of amino acids, especially microbes living in diverse communities (e.g., Strauss 1979; Hubalek *et al.,* 2017; Seif *et al.,* 2020; Ferrario *et al.,* 2015). Signalling is another essential interaction within microbial communities, which allows members of the community to efficiently share resources or recruit beneficial partners within the community. Signalling is introduced in full in Chapter 1 Section 1.1.4.2.

4.1.3 Sequencing technologies

Advancements in sequencing technologies has allowed greater volumes of higher quality data to be produced at a reduced cost, leading to a boom in the amount of genomic data generated. These technologies have allowed projects such as the TARA ocean expedition (Bork *et al.,* 2015) to generate data and release it to the scientific community for analysis. A key example is Delmont et al. (2021) who used the TARA Oceans data to identify and characterise SMAGs (used in Chapter 3 Section 3.3.2). The TARA metagenomes were all produced via next-generation (short-read) sequencing (NGS) (Sunagawa *et al.,* 2015) via Illumina sequencing. However, since this time, long-read sequencing has been developed, which provides a complementary system to short-read sequencing.

Long-read sequencing, alternatively known as third-generation sequencing, is primarily achieved by two sequencing technologies: Oxford Nanopore Technologies (ONT) or Pacific Biosciences sequencing (PacBio) (Wang *et al.,* 2021; Rhoads and Au 2015). These technologies have advantages over NGS because they can produce reads that are many thousands of bases long. This characteristic is desirable for genome sequencing for a number of reasons. Firstly, it makes the assembly process easier due to an increase in the overlap span of the reads. As a result, the assemblies are much more contiguous than assemblies produced using short-read data. Secondly, long-read sequencing has the ability to span highly repetitive regions of a genome. This allows more accurate assembly of regions that have traditionally been troublesome for assembly algorithms, which find it difficult to identify the correct length of highly repetitive stretches of DNA (van Dijk *et al.,* 2018). Of note is the recent reporting of telomere to telomere resequencing of the

Arabidopsis chromosomes, revealing for the first time the complete structure of eukaryotic centromeres (Naish *et al.,* 2021) and the following year those in humans (Altemose *et al.,* 2022).

On the other hand, third generation sequencing, especially for ONT, has a high level of error in base calling. In some cases, up to 1 in 10 bases can be called incorrectly. For this reason, in some instances it may be beneficial to use assembly methods that utilise both short and long read sequencing techniques, integrating the benefits from both. For example, when the long-read sequencing quality is low (i.e., high error rates) or when attempting to identify intraspecies genomic variation (Brown *et al.,* 2021; Miller *et al.,* 2017; Chen, Erickson, and Meng 2020).

4.1.4 Aims of this chapter

In this chapter, I will outline some physiological and imaging work carried out to further our understanding of the MAST3 species *I. marina*. As described above this is one of only a handful of culturable species belonging to the basal lineages of the stramenopiles and could provide valuable insight into the evolution of the diversity of this group, including their vitamin requirements. *I. marina* was isolated and has subsequently been maintained with an assemblage of bacteria (Cavalier-Smith and Scoble, 2013). The interactions, signalling and nutrient transfer between the eukaryote and the bacteria in this system, and indeed between the bacteria themselves, are currently unknown and have not been previously investigated. To investigate this further a sequencing effort was performed to improve the genomic information available for *I. marina* in conjunction with the bacterial species it was co-isolated and subsequently cultured with. Long-read sequencing was employed along with short-read sequencing. The genome was subsequently mined for indications of various traits such as biosynthesis pathways for B vitamins, amino acids and signalling molecules, in an effort to shed light on the metabolism of *I. marina* and the possible interactions that it may have with the bacterial consortium.

4.2 Materials and Methods

4.2.1 – Culturing of *Incisomonas marina*

I. marina was acquired from the Culture Collection of Algae and Protozoa (culture number CCAP 997/1). The culture was maintained in artificial seawater for protazoa (ASWP), the recipe can be found here https://www.ccap.ac.uk/index.php/media-recipes/. A barley grain was included in the culture which was sterilised by boiling (10 minutes) before being added. Sub-culturing was performed at a frequency of anywhere from 2 weeks to 16 weeks in an inoculation ratio of 1 ml of culture to 24 ml of fresh media. Cultures were grown at 15 °C in a light dark cycle of 12:12 light to dark.

4.2.1.1 – Antibiotic treatment

To assess the dependence of *I. marina* on the bacteria in the culture community, a set of antibiotic treated cultures were set up to compare to untreated cultures. Culture conditions as above. Two different antibiotic cocktails were used:

1. Rifampicin (10 µg/ml), streptomycin (100 µg/ml), gentamycin (100 µg/ml), ampicillin (100 µg/ml)
2. Streptomycin (30 µg/ml), neomycin (60 µg/ml), kanamycin (50 µg/ml), ampicillin (50 µg/ml), chloramphenicol (25 µg/ml)

Light microscopy images were taken at 20x and 40x magnification on the Rebel microscope in the inverted configuration (see below) directly of the cultures every other day over 7 days.

4.2.2 – Imaging of *Incisomonas marina* cultures

4.2.2.1 – Light microscopy

To check on the health of the cultures, regular imaging was carried out. Since the community form a biofilm-like growth complex on the bottom surface of the growth vessel, inverted microscopy was used.

4.2.2.2 – Scanning Electron Microscopy

Cells were grown on Melinex plastic coverslips (Agar Scientific). Then, samples were very briefly dipped twice in cold, de-ionised water to remove any buffer salts and quickly plunge-frozen by dipping into liquid nitrogen-cooled ethane. Then, samples were transferred to liquid nitrogen-cooled brass inserts and freeze-dried overnight in a liquid nitrogen-cooled turbo freeze-drier (Quorum K775X). Samples were mounted on aluminium SEM stubs using conductive silver pain (Agar Scientific) and coated with 15 nm iridium using a Quorum K575X sputter coater. Samples were viewed using a FEI Verios 460 scanning electron microscope run at 2.00 keV and 50 pA probe current. Secondary electron images were acquired using either an Everhard- Thornley detector in field-free mode (low resolution) or a Through-Lens detector in full immersion mode (high resolution).

4.2.3 – Bacterial isolation and colony PCR

In an initial attempt to identify the bacteria present in the *I. marina* cultures, samples were spread on Marine broth agar plates and incubated in the same conditions as the cultures (Section 4.2.1). Subsequently colony PCR was performed on colonies of bacteria with distinct morphological

appearances to amplify 16S rRNA sequences. Primers used as in Section 2.2.4. PCR was carried out using Red Taq kit (Sigma-Aldrich – R2523) and the cycling conditions were:

1. 95°C – 1 minute
2. Cycle x30 of:
    a. 95°C – 0.5 minutes
    b. 50°C – 0.5 minutes
    c. 72°C – 1.5 minutes
3. 72 °C – 3 minutes

PCR products were analysed using a 2% agarose gel. Illustra GFX Gel Band Purification kit was used to purify PRC products and subsequently send for Sanger sequencing. Resulting sequences were analysed in Geneious Prime and SINA taxonomy identification tool (Pruesse *et al.,* 2012).

4.2.4 – Metagenomics of *Incisomonas marina* culture

4.2.4.1 – DNA extraction

Prior to DNA extraction cultures were grown at a bulk volume (6 x 200 ml cultures) before then concentrating the cells by a series of centrifugation steps. Cells were scraped off the bottom of the growth vessels to suspend them before transferring to 50 ml falcon tubes (four per culture). These were centrifuged at 2000 xg for 20 minutes and repeated an additional two times. The supernatant was discarded with the exception of some to resuspend the pellet and consolidate into single 15 ml falcon tubes (1 per culture). These were then centrifuged in the same conditions. Finally, cells were again resuspended in 1 ml of supernatant and transferred to a 1.5 ml Eppendorf tube where they were centrifuged at 8,000 xg for 10 minutes. After centrifugation the cells separate into two fractions, the top layer was brown in colour and the bottom white in colour. Microscopy of these two fractions confirmed the majority of *I. marina* cells to be in the brown fraction. To reduce the bacterial load so as to not flood the sequencing library with bacterial sequence, these two fractions were manually separated. DNA was extracted using a Phenom-chloroform mini-prep method optimised in house by Pawel Mordaka and Andre Holzer and adapted from Zhang et al. (2014).

4.2.4.2 – Long-read sequencing

Library prep and sequencing was carried out by collaborators Adam Monier and Victoria Jackson at the University of Exeter. Sequencing was performed on a minION flongle and base calling was performed with guppy in high-accuracy mode.

4.2.4.3 – Sequencing data quality control and processing

The quality of the Long-read data was assessed with LongQC (Fukasawa *et al.,* 2020) and Nanoplot2 (De Coster *et al.,* 2018), both of these were run with default parameters. Trimmomatic (Bolger *et al.,* 2014) was used for read trimming and porechop (https://github.com/rrwick/Porechop) was used to remove adapter sequences.

4.2.4.4 – Assembly

Four different assemblers were tested initially with draft assemblies being compared to assess suitability of assembler. The assemblers tested were metaFlye (Kolmogorov *et al.,* 2020), Shasta (https://github.com/chanzuckerberg/shasta), Canu (Koren *et al.,* 2017) and MaSuRCA (Zimin *et al.,* 2017). The quality of the draft assemblies produced were assessed and compared to each other using Quast (Gurevich *et al.,* 2013). The assemblers were run with the following parameters: Flye was run with the read input flag "--nano-raw" and the flag "--meta", Shasta was run in default parameters, Canu was run with the options genomeSize=167000000, purgeOverlaps=aggressive. The full MaSuRCA parameter file can be found in Appendix Section 2.1. Quast was run with default parameters.

4.2.4.5 – Anvi'o metagenomics workflow

The general workflow followed can be found here: https://merenlab.org/2016/06/22/anvio-tutorial-v2/

Briefly, the Long-reads used for the initial assembly were aligned back to the final assembly using LongReadAligner (lra) (Ren and Chaisson, 2021). This allows the calculation of read recruitment for future bins and for Anvi'o to perform hierarchical clustering. The resulting SAM file was then converted to a BAM file and a BAM index file using samtools (Li *et al.,* 2009). From then Anvi'o (Eren *et al.,* 2021) was used to create a contigs database (anvi-gen-contigs-database), followed by searching for HMMs of rRNA genes and single-copy genes for protists and bacterial lineages (anvi-run-hmms). An Anvi'o profile (anvi-profile) was then created for the contig database to allow the use of the Anvi'o interactive interface for supervised binning (anvi-interactive (Eren *et al.,* 2015)). Binning was performed manually based on read clustering, presence of rRNA genes, differences in GC content and read coverage. MAGs were extracted by summarising the binning effort (anvi-summarize).

4.2.4.6 – Anvi'o MAG analysis workflow for metabolism analysis

Further analysis of the MAGs was also performed with the Anvi'o suite of analysis programs. A contig database was created for each MAG (anvi-gen-contigs-database) before repeating the HMM search for rRNA genes and single-copy genes (anvi-run-hmms). Next, the level of completeness was assessed for each bin by running anvi-estimate-genome-completeness which measures the number of single-copy orthologues identified compared to the full set.

Next, each bin was annotated with KEGG orthologues achieved by running anvi-run-kegg-kofams. This uses hmmsearch to annotate contig databases with KOfam HMM hits. This ensures the necessary information is present in the contig database for each MAG to allow metabolism estimation (anvi-estimate-metabolism). This takes pathways defined in the KEGG pathways database and assesses the completeness of these pathways by the presence or absence of the KOfam HMM hits provided in the previous step of the process.

4.2.4.7 – Decontamination, Repeat analysis and eukaryotic gene prediction

To remove potential co-assembled bacterial content from the *I. marina* bin, the NCBI decontamination tool, FCS, was used to remove suspected contamination. (https://github.com/ncbi/fcs). The repeat content of the *I. marina* bin was assessed using RepeatModeler and RepeatMasker. (https://www.repeatmasker.org). Although Anvi'o performs ab initio gene calling on binned contigs, the program employed is prodigal (Hyatt *et al.,* 2010) which is optimised for prokaryotic gene calling. To achieve an appropriate set of gene calls for the eukaryotic MAG, contigs were submitted to the AUGUSTUS web-server (https://bioinf.uni-greifswald.de/webaugustus/).

4.3 Results

4.3.1 Physiology and imaging of *Incisomonas marina* and associated bacteria

4.3.1.1 Culturing and nutrient limitation of I. marina cultures

I. marina was isolated from an estuarine environment with a consortium of bacteria (Cavalier-Smith and Scoble, 2013). It is maintained in a medium that resembles the oceanic environment – ASWP. This is a complex medium that includes a soil extract (boiled and filtered sample of soil). The culture is also provided with a boiled barley grain, which is thought to provide a carbon source for the bacterial species in the culture. Figure 4.1a shows a wide field image of a culture, individual circular cells can be seen covering the field of view. The community primarily grows in a benthic-like state on the bottom surface of the growth vessel, with the cells regularly forming "clumps"

(Figure 4.1a-b) as well as living individually both on the surface and swimming in the water column (Figure 4.1c-d). The cells of *I. marina* are circular, around 2 µm in diameter and they possess a single flagellum, the length of which can be many times the cell diameter (Cavalier-Smith and Scoble, 2013). This can be observed in Figure 4.1d and is indicated with arrows.

I. marina is presumed to be bacterivorous as the only source of food in the original growth medium was the bacterial co-isolates (although subsequently the community is grown with the addition of a boiled cereal grain, which may provide carbon to *I. marina* as well as the bacteria). To investigate the dependency of *I. marina* on the bacterial species present in the culture a series of cultures were set up containing antibiotics, (two alternative cocktails contained (1) rifampicin, streptomycin, gentamycin and ampicillin, (2) streptomycin, neomycin, kanamycin, ampicillin and chloramphenicol) (Section 4.2.1.1). The cultures were then imaged and compared to a control with

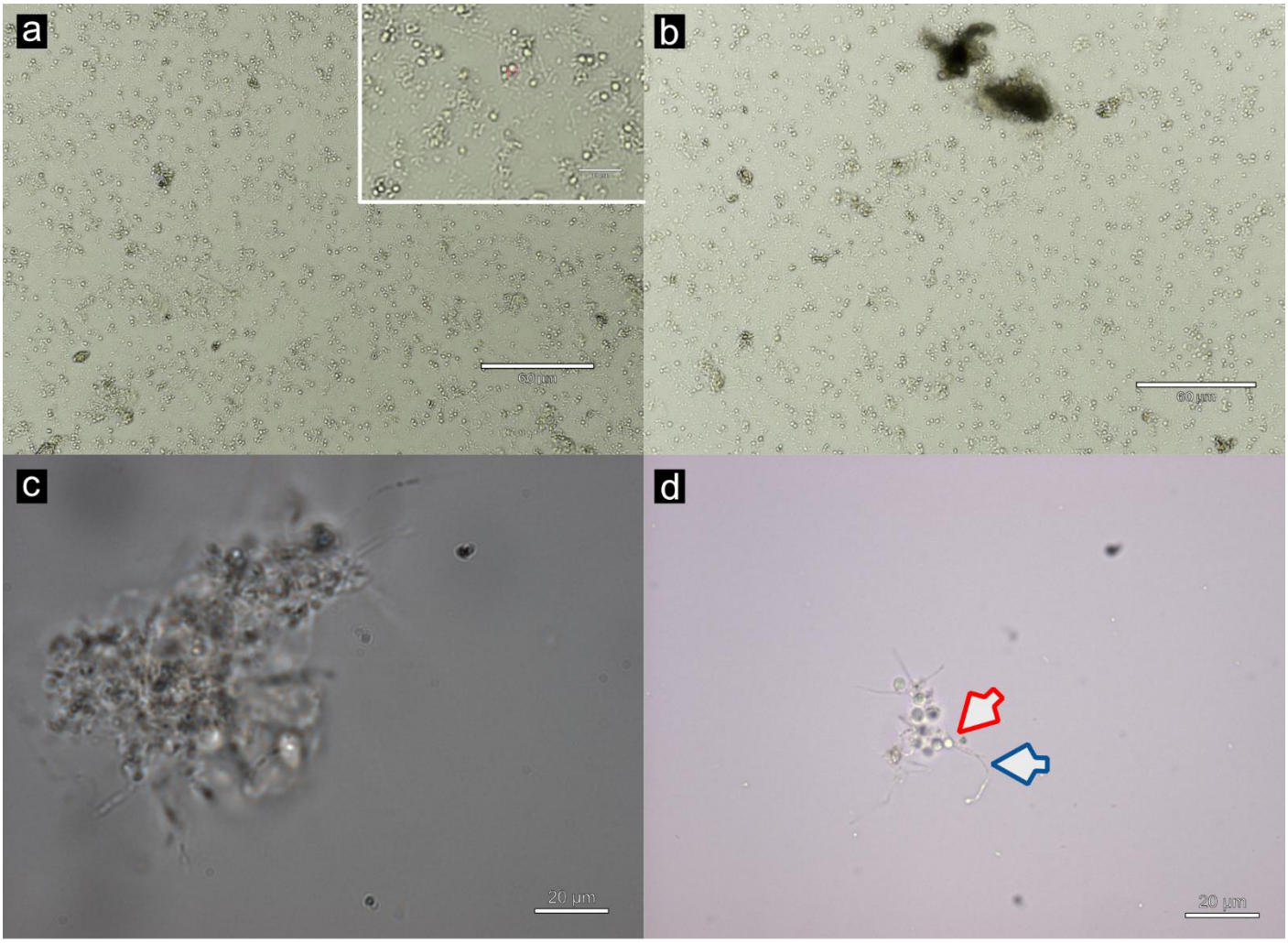


Figure 4.1 - Light microscopy images of I. marina cultures. *(a-b) Inverted microscope images directly of I. marina cultures at x40 magnification. These images show a healthy culture with cells forming a layer on the bottom surface of the growth vessel. The insert in (a) is a zoomed in section with an annotated cell demonstrating the cell size. (c-d) Oil immersion x100 magnification of culture samples. These images show typical clumps that form in the medium column. Red arrow indicates a cell body of I. marina and the blue arrow indicates the flagellum.*

no antibiotics, as well as a control of media only (no cells), over a time course of 7 days (Figure 4.2). It has not been possible to develop a technique to monitor the growth of the *I. marina* culture quantitatively thus far because of the nature of the cells and how they grow as a mixture of a single benthic layer and as clumps, but it is clear that the appearance of the culture over these days is very different in the treated cultures versus the untreated culture. After 5 days of antibiotic treatment there are no cells that could be identified as *I. marina* in the former. Indeed, there appear to be no cells at all. The lack of bacteria in the treated cultures was demonstrated by plating out on marine broth agar plates, which showed no growth, in comparison bacterial colonies were abundant on the plates spread with samples from the untreated control culture (Figure 4.3). This suggests that when there are no bacterial cells there are also no *I. marina* cells. However, this does not preclude the possibility that the antibiotics killed the *I. marina* cells directly as well as the bacteria. However, two cocktails of antibiotics were used, one of which is routinely used for treatment of bacterial contamination in cultures of *P. tricornutum* and the other used in reducing bacterial load in cultures of *C. roenbergensis.* Neither of these relatives of *I. marina* are reported to be affected by these treatments (Hackl *et al.,* 2020) which would suggest that removing the bacterial community from the culture by antibiotic treatment has the knock-on effect of killing *I. marina* by removal of a nutritional source.

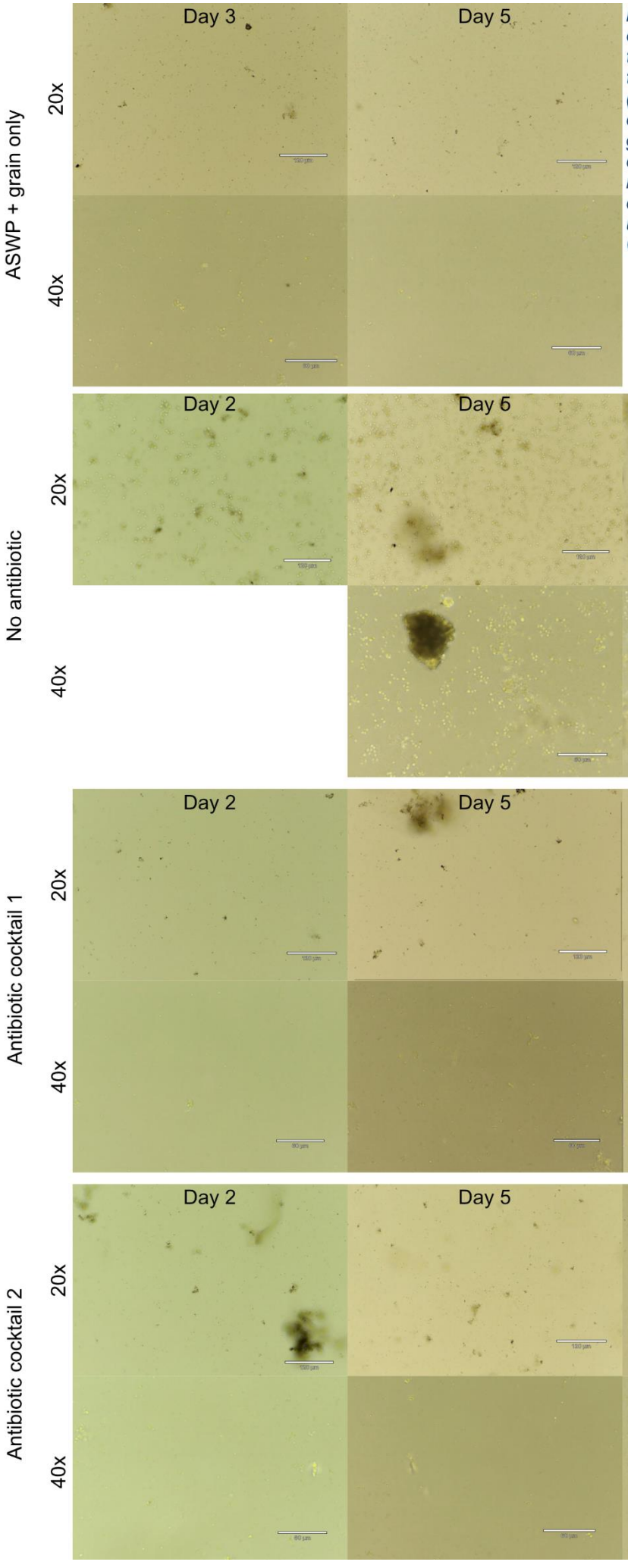


Figure 4.2 - Light microscopy example images from antibiotic treated I. marina cultures versus no treatment versus media only control (no cells). *Antibiotic cocktail 1 consisted of rifampicin, streptomycin, gentamycin and ampicillin. Antibiotic cocktail 2 consisted of streptomycin, neomycin, kanamycin, ampicillin, and chloramphenicol. Cultures were grown under normal conditions (15 degrees Celsius, 12:12 Dark:Light cycle).*

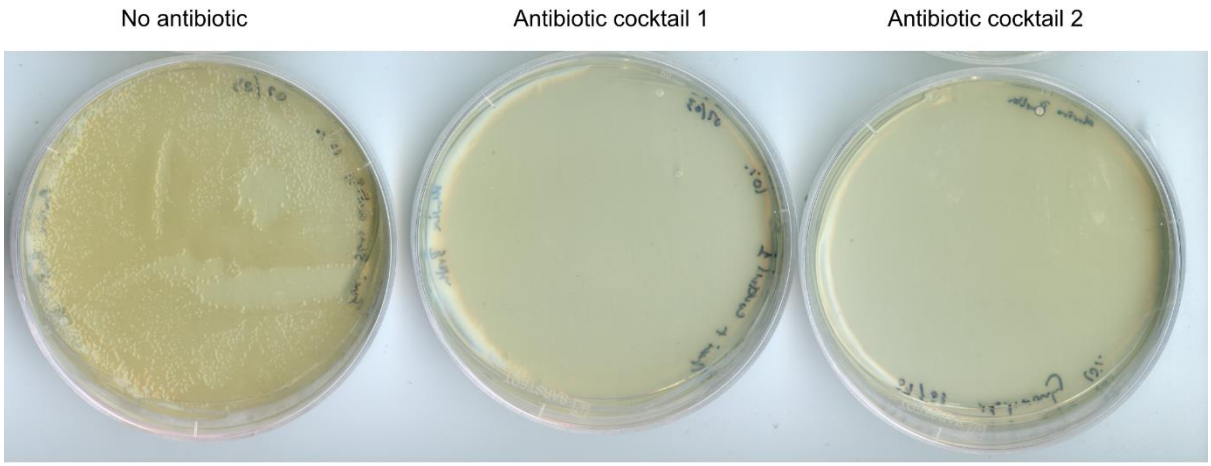


Figure 4.3 - Samples taken from I. marina cultures treated with antibiotic cocktails versus no treatment (see Figure 2) plated on marine broth agar plates. *No growth was observed on plates spread with samples from antibiotic treated cultures after 7 days. In comparison bacterial growth was observed on plates spread with samples from untreated cultures over the same time period.*

4.3.1.2 Scanning electron microscopy

Samples of *I. marina* and associated bacterial isolates from untreated cultures were imaged using SEM. Various views of individual *I. marina* cells with associated bacteria can be seen in Figure 4.4 a-d. This imaging effort substantiates the transmission electron microscopy performed in the original strain identification paper by Scoble and Cavalier-Smith (2013). It confirms the presence of a single smooth flagellum. This is in direct contradiction to the etymology of either name of this group. Stramenopile derives from the Latin *stramen* = straw and *pilus* = hair referring to the characteristic mastigonemes (lateral hairs protruding from the flagella) of many stramenopiles such as *C. roenbergensis* (Webster and Weber, 2007). The name Heterokonta, the original name for the stramenopiles (Green , Leadbeater and Diver, 1989), means possessing two flagella of unequal length (Webster and Weber, 2007). Other key observations made from the SEM imaging include the recurrent presence of an opening at the base of the flagellum where it meets the cell body, indicated in Figure 4.4 a-c by a white and red arrow, and the alignment of bacteria along the length of the flagellum of *I. marina* cells. Images of "clumping" behaviour exhibited by the culture can be seen in Figures 4.4 e-h. This behaviour is empirically associated with culture stress as was generally observed after long periods without sub-culture. The appearance of the cells in this benthic state is indicative of growth in culture where the majority of cells are adhered to the surface of the growth vessel, with a shift towards independent cells and mobile cells with a healthier culture condition. This corroborates the report by Cavalier-Smith and Scoble (2013) who observed individual

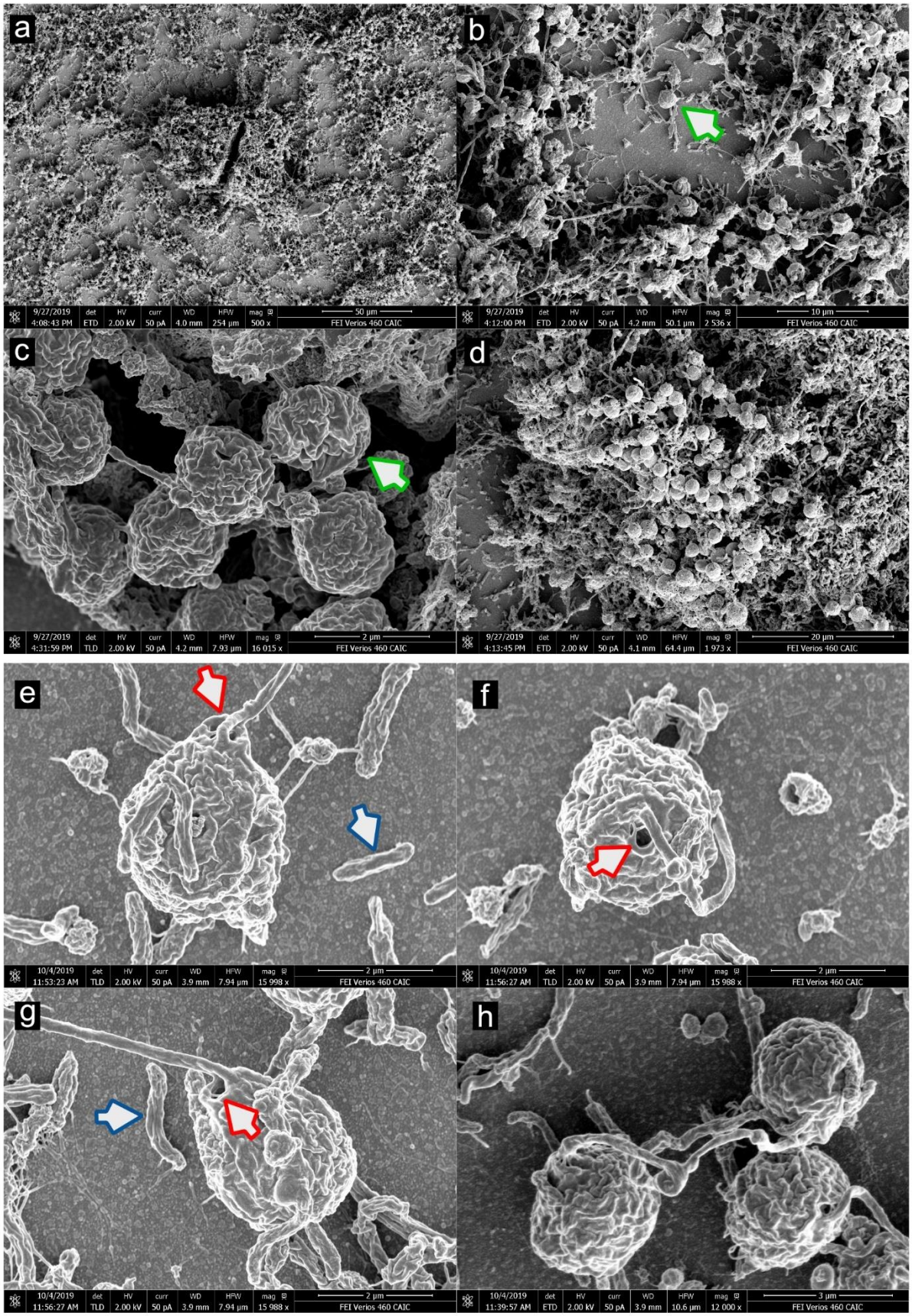


Figure 4.4 - Scanning Electron Microscopy of I. marina culture. *(a - d) Wide field view of I. marina colonies and associated bacteria displaying benthic growth and clumping behaviour. Green arrows indicate individual I. marina cells. (e-h) Individual I. marina cells displaying single flagellum as well as physically close associations with bacteria. Red and white arrows indicate a consistently observed opening at the base of the flagella and a protrusion at the base of the cell body as observed by Cavalier-Smith and Scoble (2013). Blue arrows indicate bacterial cells.*

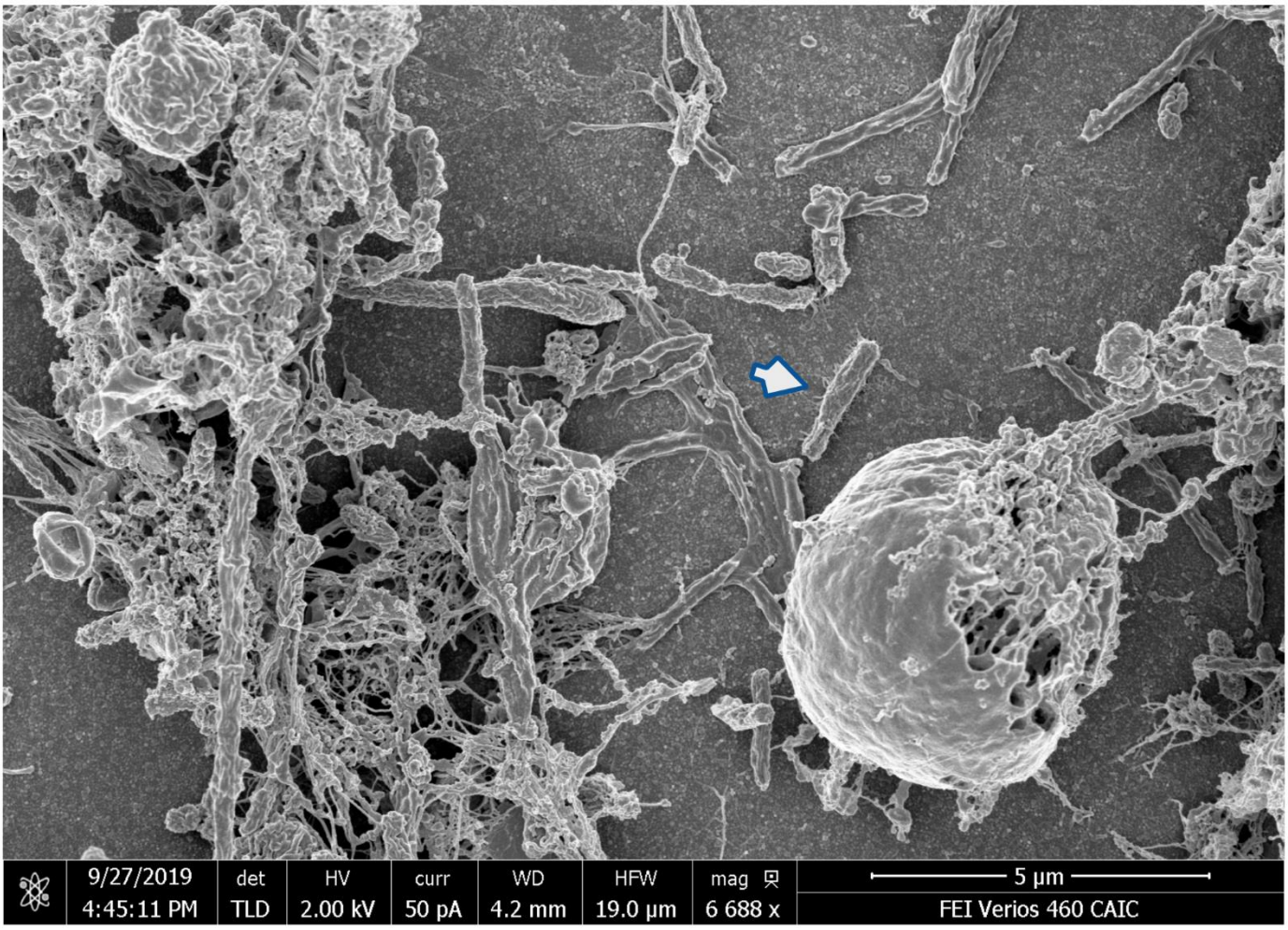


Figure 4.5 - SEM image of an I. marina culture which appears to display a dead and decaying I. marina cell (bottom right quadrant) with associated bacteria in close proximity potentially feeding on this nutrient pool.

swimming cells more commonly in well-fed cultures. Figure 4.5 depicts what appears to be the degradation of an *I. marina* cell by bacteria suggesting that not only does the eukaryote predate the bacteria in the community but also that the bacteria may gain a nutritional advantage from the death of the eukaryotic cells.

4.3.2 Identification of bacterial members of the consortia

To identify what species of bacteria are living in the culture community, two initial approaches were taken: (1) physical isolation of bacterial colonies from marine broth plates followed by colony PCR and Sanger sequencing of the 16S rRNA gene; and (2) amplicon sequencing of the 16S rRNA gene directly from DNA extractions performed on the liquid culture.

Samples of the culture plated on marine broth showed several morphologically-different bacterial species (images of the plates and colonies picked are in Figure 4.6). The first approach performed

colony PCR on four unique (by appearance) colonies in duplicate. The four distinct colony types were pink, white, yellow and clear. Two of each colony types were used for colony PCR. PCR products of 16S rRNA gene were sent for Sanger sequencing, this did not include a sequence for the pink colony as this failed to amplify during colony PCR. This process yielded four unique sequences (Appendix Section 2.2). The SINA taxonomy assignment tool (Pruesse *et al.,* 2012) uses a last common ancestor approach to assign to the lowest taxonomic rank available to the query sequence. The genera assignments for these four sequences were: *Marinobacter* (clear), *Rhodobacteraceae* (white), *Alteromonas* (white) and *Winogradskyella* (yellow). Full taxonomic predictions can be found in Table 4.1.

The amplicon sequencing sample was processed using the pipeline utilised for the analysis performed in Chapter 2 (described in section 2.2.5) with taxonomic assignment being performed using SILVA v132 (Yilmaz *et al.,* 2014; Quast *et al.,* 2013). Thirty-five sequences were classified, with 26 unique classifications (Table 4.2).

All four of the taxonomic predictions for the physically isolated species were also predicted in the taxonomic assignment of the amplicon sequencing effort performed directly on the culture. This indicates that the assignments of these species are likely to be accurate, at least to the genus level.

Table 4.1 – SINA Taxonomic predictions from colony PCR (16S rDNA gene) of isolate bacteria from plating of I. marina culture

Seq	Phylum	Class	Order	Family	Genus
1	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Winogradskyella*
2	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
3	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Alteromonadaceae*	*Alteromonas*
4	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Marinobacteraceae*	*Marinobacter*

Table 4.2 – SILVA v132 Taxomnomic predictions of 16S amplicon sequences from I. marina cultures

Seq	Domain	Class	Order	Family	Genus	Species
1	*Bacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Alcanivoracaceae*	*Alcanivorax*	*NA*
2	*Bacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Alcanivoracaceae*	*Alcanivorax*	*venustensis*
3	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas*	*NA*
4	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas*	*stellipolaris*
5	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas*	*NA*
6	*Bacteria*	*Bacteroidia*	*Flavobacteriales*	*Crocinitomicaceae*	*NA*	*NA*
7	*Bacteria*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Arenibacter*	*NA*
8	*Bacteria*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Muricauda*	*aquimarina*
9	*Bacteria*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Polaribacter*	*NA*
10	*Bacteria*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Winogradskyella*	*NA*
11	*Bacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Hyphomonadaceae*	*Hyphomonas*	*NA*
12	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	*Idiomarina*	*NA*
13	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	*Idiomarina*	*NA*
14	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	*Idiomarina*	*NA*
15	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Idiomarinaceae*	*Idiomarina*	*NA*
16	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Marinobacteraceae*	*Marinobacter*	*NA*
17	*Bacteria*	*Gammaproteobacteria*	*Oceanospirillales*	*Marinomonadaceae*	*Marinomonas*	*pontica*
18	*Bacteria*	*Gammaproteobacteria*	*Nitrosococcales*	*Methylophagaceae*	*Methylophaga*	*NA*
19	*Bacteria*	*Subgroup_6*	*NA*	*NA*	*NA*	*NA*
20	*Bacteria*	*Gammaproteobacteria*	*NA*	*NA*	*NA*	*NA*
21	*Bacteria*	*Gammaproteobacteria*	*Cellvibrionales*	*NA*	*NA*	*NA*
22	*Bacteria*	*Alphaproteobacteria*	*Parvibaculales*	*Parvibaculaceae*	*Parvibaculum*	*NA*
23	*Bacteria*	*Phycisphaerae*	*Phycisphaerales*	*Phycisphaeraceae*	*SM1A02*	*NA*
24	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas*	*NA*
25	*Bacteria*	*Gammaproteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas*	*NA*
26	*Bacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*	*Hoeflea*	*NA*
27	*Bacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*	*Hoeflea*	*NA*
28	*Bacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Antarctobacter*	*heliothermus*
29	*Bacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*NA*	*NA*
30	*Bacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Phaeobacter*	*NA*
31	*Bacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Roseovarius*	*halotolerans*
32	*Bacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Sulfitobacter*	*NA*
33	*Bacteria*	*Gammaproteobacteria*	*Cellvibrionales*	*Spongiibacteraceae*	*Zhongshania*	*NA*
34	*Bacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Thalassospiraceae*	*Thalassospira*	*NA*
35	*Bacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Thalassospiraceae*	*Thalassospira*	*NA*

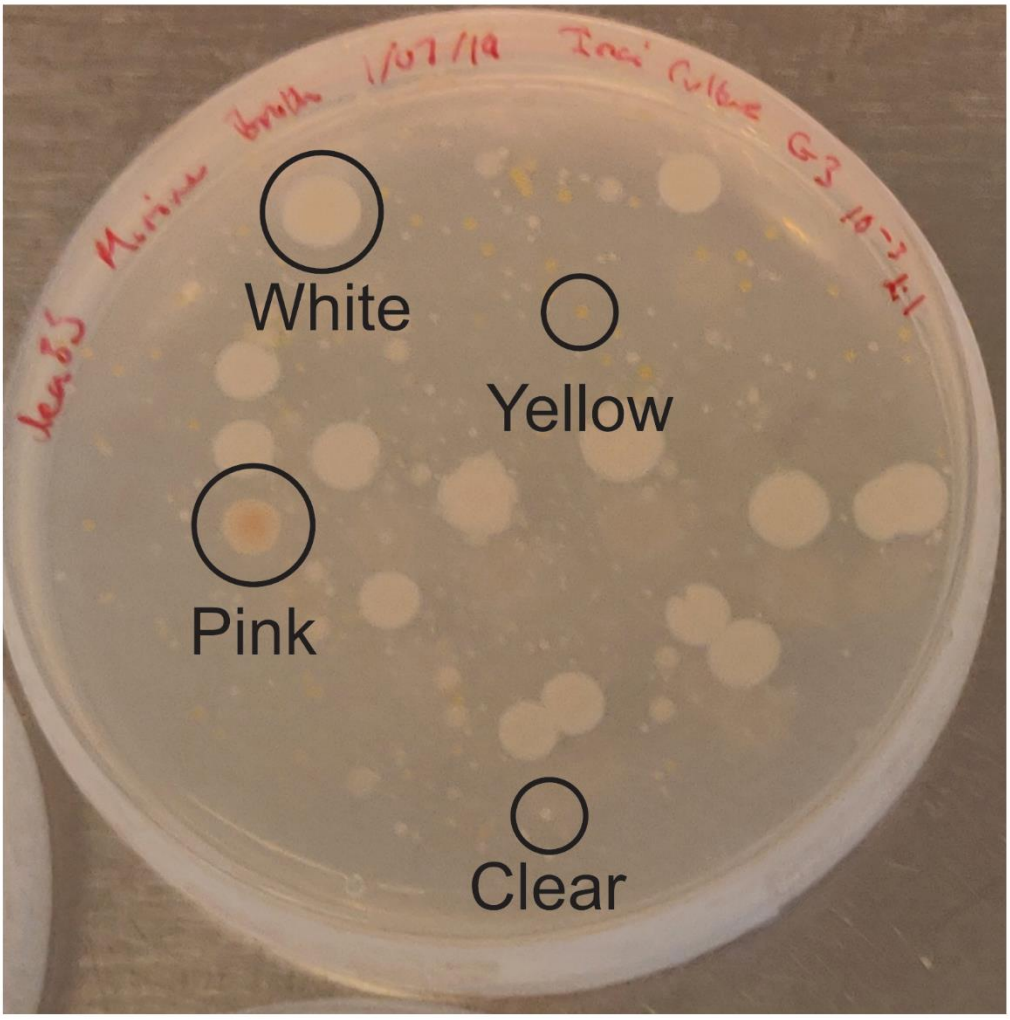


Figure 4.6 - Images of marine broth agar plates spread with I. marina culture to isolate bacterial species from the culture. *Circled are four unique by appearance colonies which were used for colony PCR and 16S Sanger sequencing for taxonomic identification. No PCR product was achieved for the pink colony. The others were identified as: Marinobacter (clear), Rhodobacteraceae (white), Altermonas (white) and Winogradskyella (yellow).*

4.3.3 Long-read sequencing

As discussed in the Introduction, *I. marina* is the only MAST3 species publicly available in culture. It has previously had a transcriptome published (Thakur *et al.,* 2019) and genomic sequencing effort (Derelle *et al.,* 2016) but no published assembly. Therefore, in an effort to improve the quality of the genomic data available for *I. marina*, as well as gain insight into the metabolic potential of the community within which this species operates in a cultured environment, a long-read sequencing effort was undertaken. DNA was extracted from cultures of *I. marina* (including bacterial co-inhabitants) and sent to the University of Exeter where ONT sequencing was performed by collaborators Victoria Jackson and Adam Monier.

Four sequencing runs resulted in 16.3 Gb of data. Base-calling was conducted with guppy in high-accuracy mode. The sequencing effort yielded 1,415,560 reads spanning a total of 16,635,245,180

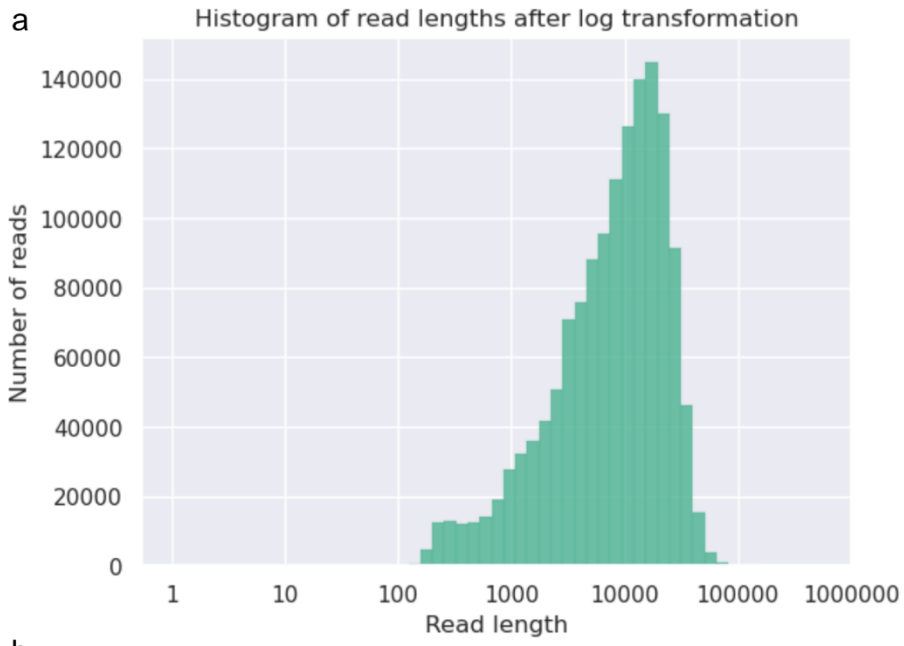
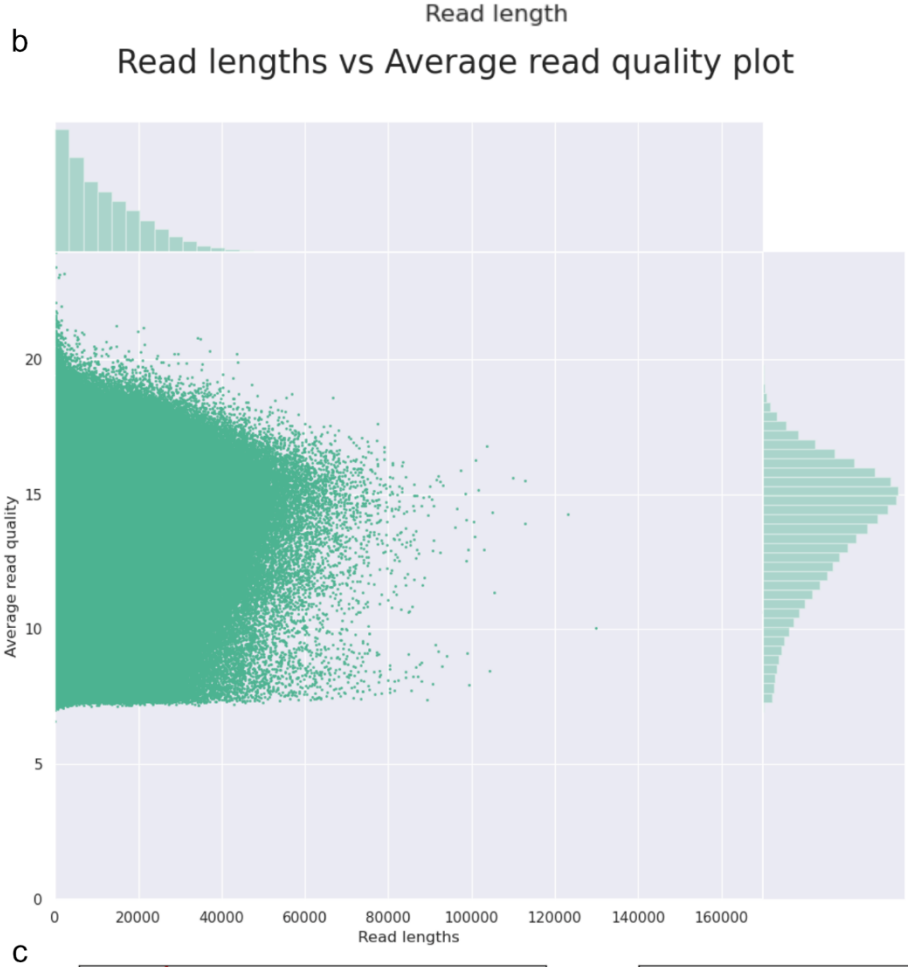
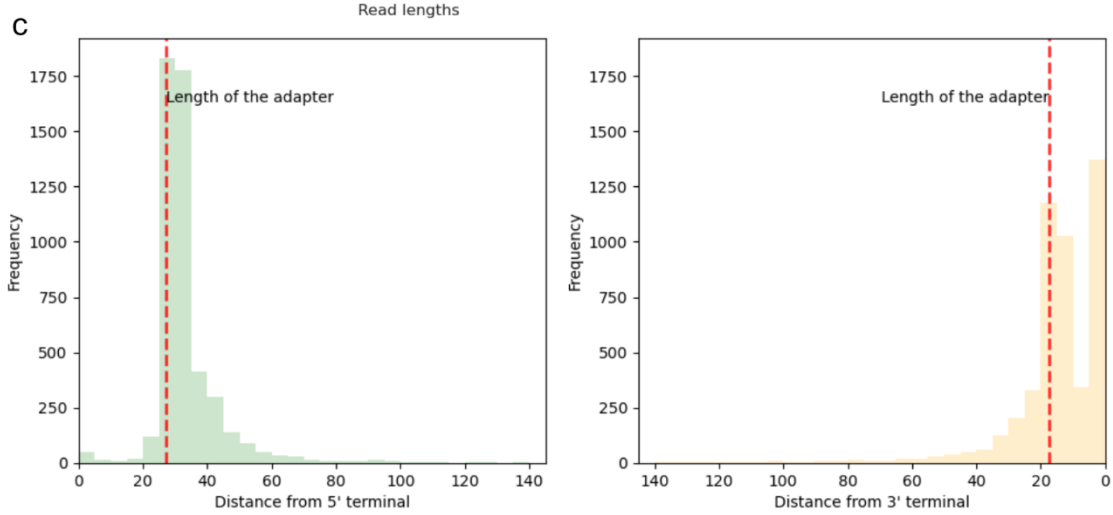


Figure 4.7 - Quality Control plots of long-read data. *(a) Histogram of read length (bp) after log transformation. (b) Scatter plot of read length (bp) against phred quality (Q) score for each read. (c) Frequency of adapter sequences identified and proximity to terminal of read (bp).*





bases. The mean read length was 11,752 bases and the N50 was 18,816 bases. The distribution of read lengths can be seen in Figure 4.7a. All the reads were above phred quality value Q7 and 93.5% were above Q10 (i.e., over 90% of reads had a base call accuracy of >90%). Read length versus read quality is shown in Figure 4.7b.

Figure 4.7c displays the frequency of sites at which adapter sequences were identified. This information informed the subsequent trimming operations. Two different trimming approaches were taken, one simply removed the first and last 50 bases from every read, the other used a specific adapter removal tool. The two resulting read files were exactly the same size (in terms of Gb) indicating that the different trimming approaches had little difference in the resulting reads. An alternative approach of filtering for read quality was also taken, where reads with an assigned quality of <Q10 were removed. Two sets of these quality-controlled reads were used for *de novo* metagenomic assembly: the trimmed reads (at 50 bases) and the quality filtered reads.

4.3.4 Metagenome Assembly

After taking relevant quality control steps with the long-read ONT data, both the raw reads and the two QC set of reads were used to create draft metagenome assemblies. The general workflow of the assembly process is outlined in Figure 4.8. As is good practice (Latorre-Pérez *et al.,* 2020), a number of different assemblers were used to create *de novo* assemblies from the processed reads. A comparison of the initial results of three assemblers was created using Quast (Gurevich *et al.,* 2013) (Table 4.3). The different variations of trimming and quality filtering parameters were also tested with the different assemblers. Table 4.3 shows various measures of assembly quality. Each column represents an assembly run including the assembler used and the reads used as input, either all reads (raw, unfiltered), trimmed (50 bases of each end of each read) or quality filtered (>=Q10). Quality parameters that were of particular note were number of contigs, N50 (length of contig at which 50% of bp are assembled) and largest contig (bp). The Flye assembly with the trimmed input data was selected to be carried forward for its superior N50 value (1,725,137 bp) and because it had the largest contig of all the assemblies (5,850,649 bp), it also had a lower rate of unassigned bases in the assembly (#N's Table 4.3), in comparison to the assemblies with the other read sets.

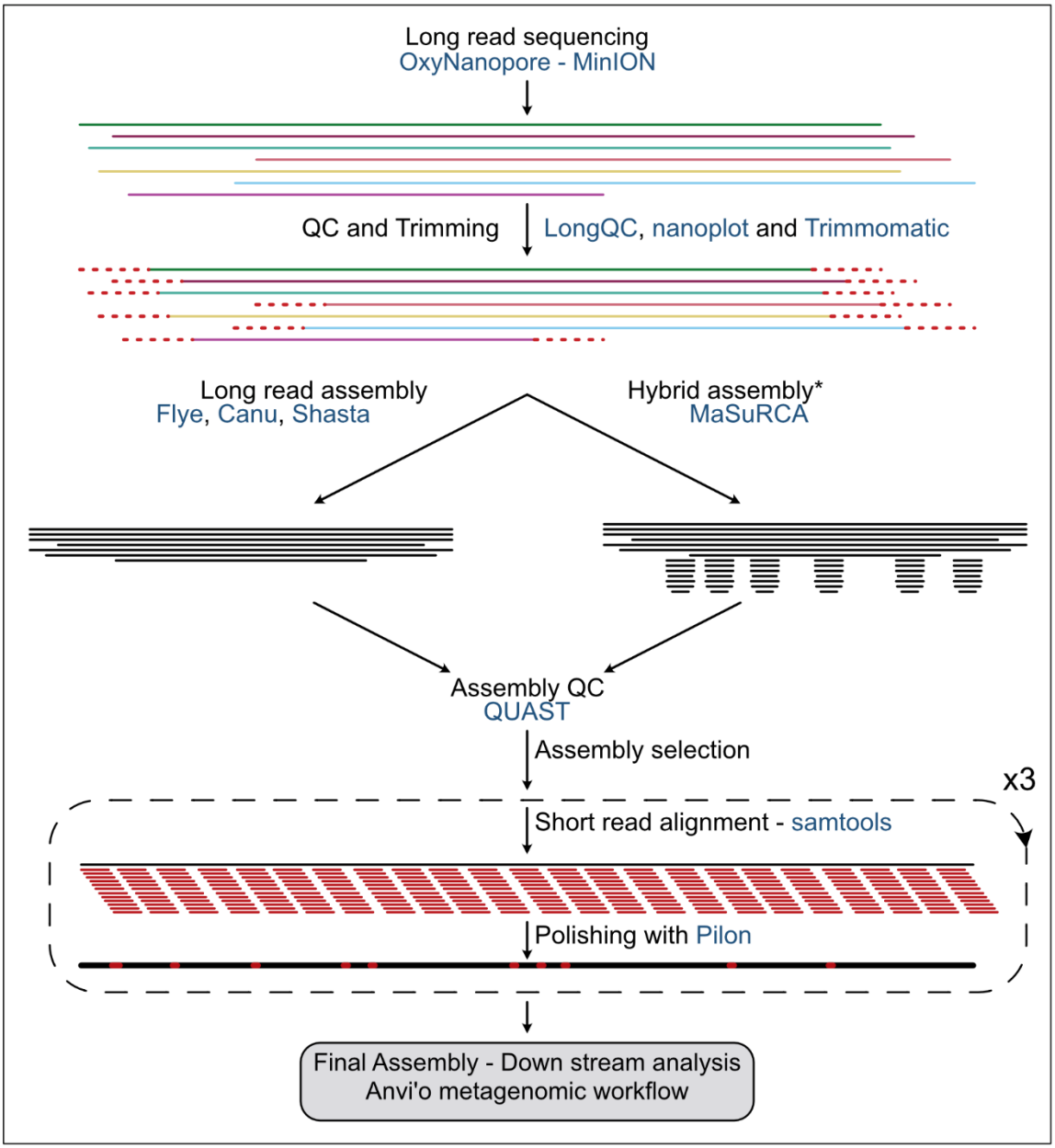


Figure 4.8 - Schematic of the metagenome assembly workflow followed. *Where appropriate the tool used for each step in blue.*

Initially the MaSuRCA assembly was also carried forward because it allows a hybrid method of assembly, so in this case it was possible to combine both the raw long-read ONT data generated in this project and the short-read data generated by Derelle et al. (2016) simultaneously. However, the MaSuRCA assembly was significantly shorter in total length than the Flye assembly: 102,483,909 bp

Table 4.3. Comparison of assembly statistics from different assemblers with different read treatments. Generated using QUAST.

Assembler	Flye	Flye	Flye	Flye	Shasta	Shasta	Canu	Canu	MaSuRCA
Reads	All	Trimmed	Pore-chop	Quality filt.	All	Trimmed	All	Trimmed	All
# contigs (>= 0 bp)	424	474	474	450	649	761	383	383	297
# contigs (>= 1000 bp)	400	431	437	411	560	608	383	383	297
# contigs (>= 5000 bp)	295	317	308	303	443	503	383	383	295
# contigs (>= 10000 bp)	250	264	256	254	379	436	379	379	293
# contigs (>= 25000 bp)	208	212	213	217	303	348	364	364	280
# contigs (>= 50000 bp)	172	178	172	184	254	280	270	268	239
Total length (>= 0 bp)	166,489,361	166,470,680	166,481,360	165,754,362	136,402,352	142,046,290	154,509,381	155,231,615	102,483,909
Total length (>= 1000 bp)	166,474,012	166,441,316	166,456,026	165,727,290	136,373,347	141,994,472	154,509,381	155,231,615	102,483,909
Total length (>= 5000 bp)	166,177,124	166,118,840	166,097,958	165,420,102	136,036,892	141,703,730	154,509,381	155,231,615	102,477,312
Total length (>= 10000 bp)	165,859,370	165,740,664	165,733,640	165,058,791	135,567,404	141,216,209	154,476,584	155,196,071	102,464,958
Total length (>= 25000 bp)	165,158,370	164,863,205	165,008,841	164,442,524	134,351,684	139,796,340	154,238,579	154,956,243	102,194,984
Total length (>= 50000 bp)	163,858,054	163,619,514	163,472,994	163,192,425	132,620,798	137,344,728	150,534,449	151,088,580	100,500,369
# contigs	422	472	473	446	584	648	383	383	297
Largest contig	5,420,675	5,850,649	5,420,681	5,469,241	4,793,997	4,793,851	4,819,579	5,557,920	4,671,608
Total length	166,488,735	166,469,932	166,480,931	165,752,586	136,391,666	142,023,952	154,509,381	155,231,615	102,483,909
GC (%)	52.4	52.41	52.41	52.39	52.92	52.39	52.51	52.5	52.7
N50	1,659,535	1,725,137	1,715,585	1,634,624	852,164	845,682	1,347,720	1,397,238	853,059
N90	515,414	481,297	507,755	402,658	185,867	167,950	186,283	176,800	131,360
L50	23	23	23	24	45	48	30	27	31
L90	91	92	90	97	162	182	131	127	137
# N's per 100 kbp	0.18	0.12	0.36	0.24	0	0	0	0	1.66

compared to 166,469,932 bp respectively. When a draft binning effort was performed on the MaSuRCA assembly, it was clear that the quality of the assembly was inferior to that of the Flye assembly, most obviously due to the low number of recovered bacterial bins. For this reason, the rest of the analysis was performed on the Flye assembly only.

To improve the quality of the *de novo* long-read assembly, short-reads were used to "polish" errors (as mentioned above, potential error rate of base calls for ONT was as high as 10%). This was achieved using Pilon (Walker *et al.,* 2014), which allowed the short-read data from Derelle et al. (2016) to be aligned with the initial Flye assembly, thereby enabling the identification and correction of errors. This process was repeated three times, each time using the latest most corrected assembly. After three rounds of polishing the workflow moved on to further downstream analysis.

4.3.5 Metagenomic analysis and metagenome assembled genome recovery

After final assembly the raw long-read data was mapped back to the final assembly to allow calculation of read assignment and clustering of sections of the metagenome, which in turn informed the binning of the metagenome into MAGs. The full workflow of this process, including tools and Anvi'o (Eren *et al.,* 2021) commands required is shown in Figure 4.9. It follows a guide to metagenomics by the developers of Anvi'o (https://merenlab.org/2016/06/22/anvio-tutorial-v2/). After aligning the reads back to the assembly Anvi'o is used to analyse the metagenome for single-copy orthologues and ribosomal rRNA genes of both bacteria and eukaryotes. This information also informs the binning and subsequent statistical calculations of each bin.

An Anvi'o profile database was generated with the minimum contig length of 1,000 bp. In total it contained 431 contigs, which corresponded to 90% of all contigs of the assembly and 99% of all nucleotides. A supervised binning effort utilising the Anvi'o interactive interface (Eren *et al.,* 2015) generated 24 bins, accounting for 161,776,004 nucleotides, which represented 97.16% of all nucleotides stored in the contigs database and 97.18% of nucleotides stored in the profile database. The binning process is based on a number of data types: hierarchical clustering, read coverage, GC content as well as the presence and location of domain specific rRNA genes. Figure 4.10 displays the binning of the *I. marina* metagenome with the data types visualised as tracks around a central tree (hierarchical clustering tree). All but one of the bins has 16S/23S rRNA sequences and so are likely to be bacterial. Summary statistics (Tables 4.4 and 4.5) for each bin show that three of these bins consist of one contig and are predicted to be 100% complete

Table 4.4. Summary statistics (1/2) of metagenome bins of Incisomonas marina and isolates. Metagenome assembled using OxyNanopore long-read data and polished using Derrelle et al., (2016) Illumina short-read data with Flye and Pilon respectively. Supervised binning performed using the Anvi'o interactive interface.

Bin No.	Variability	Relative abundance	Std coverage	Mean coverage Q2Q3	Mean coverage	Max normalised ratio	Abundance	Bins % recruitment	Detection %
1	51.26	1	8.88	85.66	85.48	1	0.90	36.83	99.72
2	50.00	1	30.61	426.92	424.67	1	4.49	11.18	99.99
3	32.73	1	5.27	57.19	57.14	1	0.60	1.65	100.00
5	2.05	1	2.83	19.06	19.07	1	0.20	0.35	100.00
6	27.53	1	18.68	193.69	192.72	1	2.04	3.49	99.61
7	26.08	1	9.94	153.42	152.93	1	1.62	3.79	100.00
8	0.14	1	2.48	4.80	5.15	1	0.05	0.10	99.85
9	35.68	1	5.52	61.74	61.68	1	0.65	2.16	100.00
10	28.52	1	24.68	554.37	551.61	1	5.83	13.94	100.00
11	0.60	1	2.02	9.32	9.36	1	0.10	0.25	100.00
14	18.33	1	3.93	35.26	35.25	1	0.37	0.80	100.00
15	44.31	1	9.59	150.10	149.88	1	1.58	3.55	99.82
16	45.63	1	6.62	87.08	86.96	1	0.92	2.28	100.00
17	7.19	1	3.33	24.37	24.33	1	0.26	0.72	100.00
18	48.14	1	13.63	234.79	234.19	1	2.47	6.64	100.00
19	51.77	1	7.06	86.61	86.54	1	0.91	2.78	100.00
20	0.55	1	30.05	6.88	16.83	1	0.18	0.35	99.92
21	131.61	1	7.72	74.34	73.84	1	0.78	2.39	100.00
23	6.92	1	2.64	18.50	18.53	1	0.20	0.42	100.00
25	45.23	1	8.49	130.26	129.93	1	1.37	3.97	100.00
26	2.37	1	2.09	9.48	9.53	1	0.10	0.18	99.91
27	0.64	1	1.75	6.58	6.60	1	0.07	0.18	99.91
28	3.61	1	2.38	13.98	14.02	1	0.15	0.39	99.99
29	1.79	1	2.52	14.57	14.59	1	0.15	0.35	100.00
Un-binned	-	-	-	-	-	-	-	1.27	-

Table 4.5. Summary statistics (2/2) of metagenome bins of Incisomonas marina and isolates. Metagenome assembled using OxyNanopore long-read data and polished using Derrelle et al., (2016) Illumina short-read data with Flye and Pilon respectively. Supervised binning performed using the Anvi'o interactive interface.

Bin No.	Total length (Bp)	Num. contigs	N50	GC content	Percent completion	Percent redundancy
1	68,357,064	213	1215950	52.5	68.7	16.9
2	4,108,619	5	4006367	40.9	100	5.6
3	4,510,811	3	4487859	43.3	100	2.8
5	2,913,340	1	2913340	44.8	98.6	4.2
6	2,900,267	63	94070	39.4	0	0
7	3,899,295	2	3863077	33.4	97.2	16.9
8	3,085,312	21	266584	39.0	78.9	15.5
9	5,505,433	2	5469367	39.7	98.6	2.8
10	3,945,902	1	3945902	41.6	98.6	4.2
11	4,219,735	4	2482457	49.6	95.8	7.0
14	3,566,829	1	3566829	63.6	97.2	2.8
15	3,706,920	3	3586063	57.0	100	2.8
16	4,096,611	1	4096611	57.4	100	0
17	4,658,551	1	4658551	61.5	100	0
18	4,425,738	1	4425738	61.0	97.2	8.5
19	5,035,835	4	4700859	61.7	97.2	1.4
20	4,175,398	17	605885	60.5	90.1	9.9
21	5,058,777	2	4789643	59.1	98.6	0
23	3,516,122	1	3516122	58.0	97.2	7.0
25	4,785,063	2	4571768	53.5	97.2	0
26	2,943,871	8	842461	46.9	93.0	5.6
27	4,331,361	10	543885	44.9	98.6	9.9
28	4,324,906	2	4265747	66.8	97.2	7.0
29	3,704,244	1	3704244	62.7	100	0

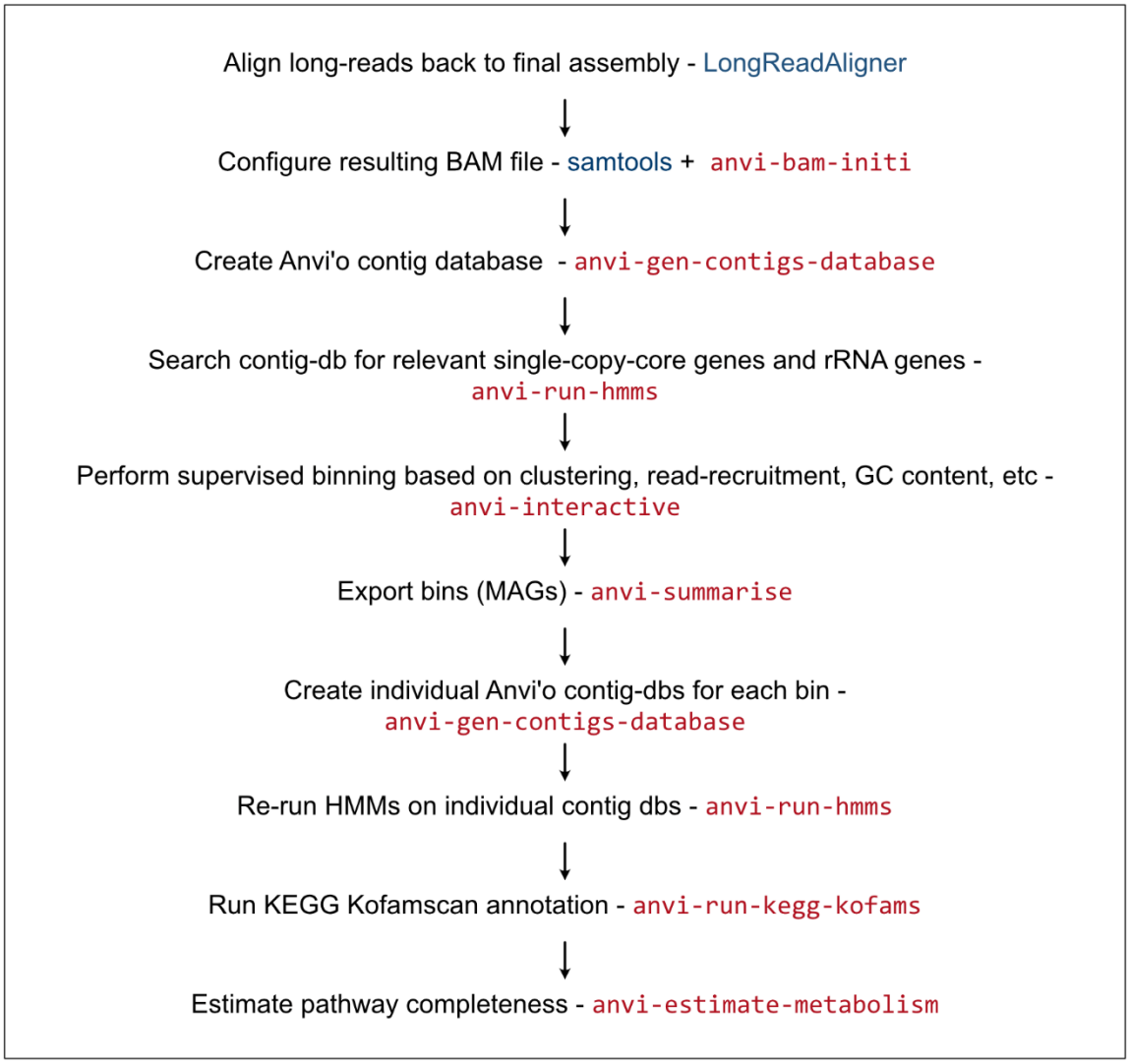


Figure 4.9 - Schematic of the downstream metagenome processing workflow. *This process took the draft assembly and allowed the recovery of metagenome assembled genomes (MAGs), providing individual genomic data for each species in the community. Bioinformatic tools used are in blue and Anvi'o commands used are in red.*

suggesting their genome was sequenced in full. An additional five bins have 1 contig and are over 97% complete demonstrating the quality of the metagenomic assembly. By utilising the taxonomy prediction function in Anvi'o the closest taxonomic rank was assigned to each bin (Table 4.6). Three of the predicted species from these MAGs match those bacteria that were previously isolated and identified from the culture using 16S sequencing: *Winogradskyella sediminis*, *Marinobacter hydrocarbonoclasticus* and *Pseudooceanicola marinus*. Conversely, three of the MAGs have no taxonomic prediction at the rank of species suggesting potentially novel species: Bin 5 – genus = *Methylophaga*, Bin 10 – genus = *Muricauda* and Bin 17 – genus = *Alcanivorax*. Bin 29 represents a possible novel genus, with taxonomic prediction only reaching the family level - *Parvibaculaceae*.

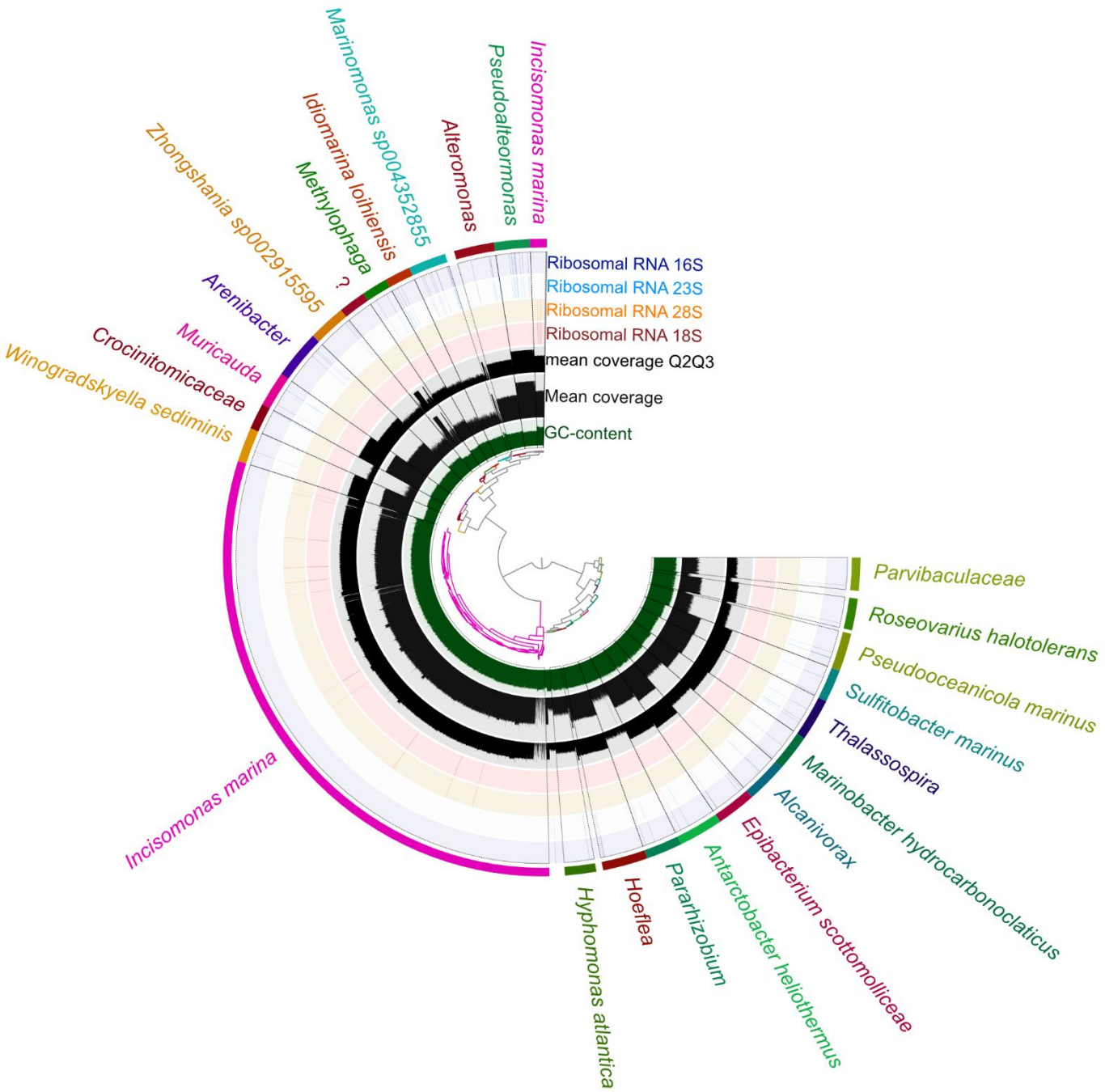


Figure 4.10 - A view of the Anvi'o interactive interface used to perform supervised binning for MAG recovery. *The central tree is a hierarchical clustering of re-aligned long reads to the final metagenome assembly. Each track surrounding this tree represents the following data used to inform binning. In order from inside out: GC-content, mean coverage (read recruitment), mean coverage for Q2 and Q3, locations of rRNA genes for 18S, 28S, 16S and 23S. Each bin is coloured individually by the final outermost track with an associated taxonomic prediction where possible.*

Table 4.6 – Taxonomic predictions of each bin as predicted by Anvi'o taxonomic prediction

Bin No.	Domain	Phylum	Class	Order	Family	Genus	Species
1	-	-	-	-	-	-	-
2	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Alteromonadaceae*	*Pseudoalteromonas*	-
3	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Alteromonadaceae*	*Alteromonas*	-
5	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Nitrosococcales*	*Methylophagaceae*	*Methylophaga*	-
6	-	-	-	-	-	-	-
7	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Winogradskyella*	*Marinobacter hydrocarbonoclasticus **
8	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Crocinitomicaceae*	-	-
9	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Arenibacter*	-
10	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Muricauda*	-
11	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Spongiibacteraceae*	*Zhongshania*	*Zhongshania sp002915595*
14	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Roseovarius*	*Roseovarius halotolerans*
15	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Sulfitobacter*	*Sulfitobacter marinus*
16	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Oleiphilaceae*	*Marinobacter*	*Marinobacter hydrocarbonoclasticus**
17	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Alcanivoracaceae*	*Alcanivorax*	-
18	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Epibacterium*	*Epibacterium scottomollicae*
19	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Antarctobacter*	*Antarctobacter heliothermus*
20	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*	*Pararhizobium*	-
21	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*	*Hoeflea*	-
23	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Hyphomonadaceae*	*Hyphomonas*	*Hyphomonas atlantica*
25	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Thalassospiraceae*	*Thalassospira*	-
26	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Alteromonadaceae*	*Idiomarina*	*Idiomarina loihiensis*
27	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Marinomonadaceae*	*Marinomonas*	*Marinomonas sp004352855*
28	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Pseudooceanicola*	*Pseudooceanicola marinus**
29	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Parvibaculales*	*Parvibaculaceae*	-	-

There is no taxonomic prediction for bin 1 as this is the eukaryotic bin belonging to *I. marina*, indicated by the identified locations of 18S/28S rRNA genes. This bin is >10x larger than the next largest bin. It also accounts for more than a third of the genetic information in the metagenome (36.83% read recruitment). Note in Figure 4.10 there are two sections of the tree that are labelled as *I. marina.* This is likely due to a misalignment of the smaller section in the metagenome. However, as 18S rRNA sequences were identified in this smaller section and all 18S sequences recovered had a closest species match with *I. marina*, this section was included in with the rest of the *I. marina* bin.

4.3.6 Downstream analyses of *Incisomonas marina* and culture community

4.3.6.1 Further characterisation of bacterial species

To consolidate the bacterial species prediction in the light of the recovered MAGs a reanalysis of the data was performed to identify overlap from the three approaches taken: (1) colony PCR of 16S from physical isolates, (2) amplicon sequencing from direct DNA extraction of liquid culture and (3) binning of the metagenome. For each approach a different tool for taxonomic prediction was used on the resulting data: SINA prediction tool (Pruesse *et al.,* 2012) was used for the first approach, SILVA (Quast *et al.,* 2013; Yilmaz *et al.,* 2014) was used for amplicon analysis and the built-in taxonomic prediction from Anvi'o was used for metagenomic bins.

To make a more valid comparison between the data types, taxonomic prediction was repeated for each approach using the SINA tool. This required the extraction of 16S sequences from each bacterial MAG which resulted in a different taxonomic prediction to that of Anvi'o for the MAG as a whole, listed in Table 4.7. A summary of the predictions and overlaps in predictions from each approach is made in Table 4.8. The presence of 28 different taxa were predicted by at least one of the approaches taken. Of these, three of the taxa were predicted in all of the data sources (*Rhodobacteraceae*, *Alteromonas* and *Marinobacter*). Eleven were predicted in two of the three data sources and fourteen were only predicted from one of the data sources. That half of the species were identified in at least two of the three approaches gives confidence in the validity of their identity. The general nature of the species identified (i.e., marine species) can also be considered a good corroboration for the analysis. The amplicon sequencing approach returned the highest number of individual predictions whilst the plating out method only yielded four. This suggests that some of the bacterial species have identical appearance when grown on solid media and that only picking two colonies resulted in a subset of the species being identified by sequencing, or that some of the species cannot grow on the media used (marine broth).

Table 4.7 – SINA Taxonomic predictions from 16S sequences extracted from bins

Seq	Domain	Phylum	Class	Order	Family	Genus
1	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Arenibacter*
2	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Joostella*
3	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Muricauda*
4	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Parvibaculales*	*Parvibaculaceae*	*Parvibaculum*
5	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Devosiaceae*	*Devosia*
6	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*	
7	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Antarctobacter*
8	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Epibacterium*
9	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Sulfitobacter*
10	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	
11	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Thalassospiraceae*	*Thalassospira*
12	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Alteromonadaceae*	*Alteromonas*
13	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Idiomarinaceae*	*Idiomarina*
14	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas*
15	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Nitrosococcales*	*Methylophagaceae*	*Methylophaga*
16	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Alcanivoracaceae1*	*Alcanivorax*
17	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Marinobacteraceae*	*Marinobacter*
18	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Marinomonadaceae*	*Marinomonas*
19	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Spongiibacteraceae*	*Zhongshania*

Table 4.8 – Taxonomic predictions from each bacterial classification approach – each using the SINA taxonomic prediction tool

Seq	*Kingdom*	*Phylum*	*Class*	*Order*	*Family*	*Genus*	Plating out	MAG 16S seqs	Amplicon
1	*Bacteria*	*Acidobacteriota*	*Vicinamibacteria*	*Vicinamibacterales*	*uncultured*				✓
2	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Crocinitomicaceae*				✓
3	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Arenibacter*		✓	✓
4	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Joostella*		✓	
5	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Muricauda*		✓	✓
6	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Polaribacter*			✓
7	*Bacteria*	*Bacteroidota*	*Bacteroidia*	*Flavobacteriales*	*Flavobacteriaceae*	*Winogradskyella*	✓		✓
8	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Hyphomonadaceae*	*Hyphomonas*			✓
9	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Parvibaculales*	*Parvibaculaceae*	*Parvibaculum*		✓	✓
10	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Devosiaceae*	*Devosia*		✓	
11	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*			✓	✓
12	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Antarctobacter*		✓	
13	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Epibacterium*		✓	
14	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Roseovarius*			✓
15	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Sulfitobacter*		✓	
16	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*		✓	✓	✓
17	*Bacteria*	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillales*	*Thalassospiraceae*	*Thalassospira*		✓	✓
18	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Alteromonadaceae*	*Alteromonas*	✓	✓	✓
19	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Alteromonadaceae*				✓
20	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Idiomarinaceae*	*Idiomarina*		✓	✓
21	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas*		✓	✓
22	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Nitrosococcales*	*Methylophagaceae*	*Methylophaga*		✓	✓
23	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Alcanivoracaceae1*	*Alcanivorax*		✓	✓
24	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Marinobacteraceae*	*Marinobacter*	✓	✓	✓
25	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Marinomonadaceae*	*Marinomonas*		✓	✓
26	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Spongiibacteraceae*	*Zhongshania*		✓	
27	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*	*Spongiibacteraceae*				✓
28	*Bacteria*	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonadales*					✓

The difference in the number of predicted species made via the amplicon sequencing approach and the metagenome binning approach either suggests that the whole community is not represented by the metagenome or that the number of species predicted by the amplicon sequencing is inflated from the true number. This could be caused by the characterisation of ASVs in the analysis pipeline being too strict (e.g., characterising sub-species as separate ASVs).

4.3.6.2 General statistics and annotation of I. marina

The *I. marina* MAG (herein referred to as genome) consisted of 213 contigs with a total length of 68,357,064 bp. The average contig length was 320,925 bp with an N50 of 1,215,950 bp and a GC content of 52.5%. To account for any co-assembly with bacterial reads, the NCBI decontamination tool, FCS, was used to remove any potential contaminating contigs. The report recommended the removal of one contig giving a total remaining 212 contigs with a total length of 68,108,360bp. After decontamination, RepeatMasker (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at http://repeatmasker.org) was used to assess the repeat content of the genome. It was determined that 20.5% of the genome was made up of repeats. This consisted predominantly of simple repeat (5,834,774 bp – 8.57%) and retroelements (2,698,197bp - 3.96%, including 1.27% large tandem repeat elements). DNA transposons made up 1.23% of the genome. The full table of output can be

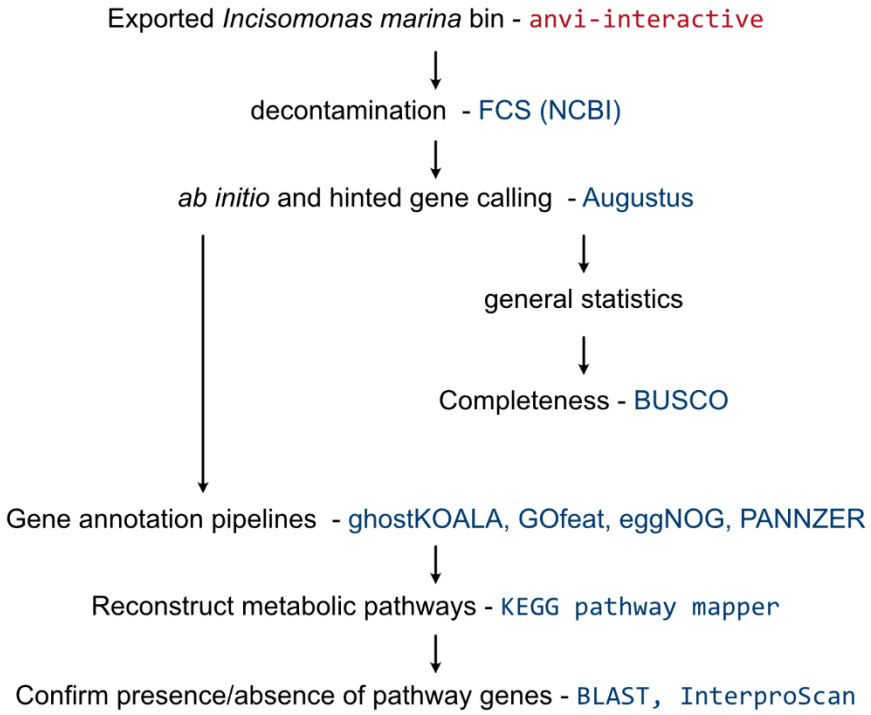


Figure 4.11 - Schematic of the down stream analysis of the I. marina MAG including gene calling, completeness estimation, annotation and pathway analysis.

found in Appendix Section 3.3. Anvi'o predicts the original bin to be 68.7% complete based on the in-built eukaryotic single-copy orthologue assessment. However, the Anvi'o functionality makes gene calls that are optimised for prokaryotic genes (Hyatt *et al.,* 2010). To achieve a more accurate assessment of the completeness of this genome the contigs were processed into gene-calls using the eukaryotic optimised Augustus gene caller (Stanke and Morgenstern, 2005). Both the genome sequence generated here, and the transcriptome of *I. marina* produced by Thakur et al. (2019) were used as input for gene prediction. Figure 4.11 shows a schematic of this workflow. The output from this was a eukaryotic-optimised set of gene calls totalling 20,238 predicted genes. When BUSCO was run on this with stramenopile_odb10 the completion was calculated at 93%. A meta comparison of the completeness of a large number of stramenopiles (498 accessions from TARA, MMETSP and JGI) can be seen in Figure 4.12.

The completeness of 93% puts the *I. marina* genome in the top 10% of these stramenopile genomes and allows us to analyse the data with a relatively high degree of confidence that missing data will not have a great influence on results.

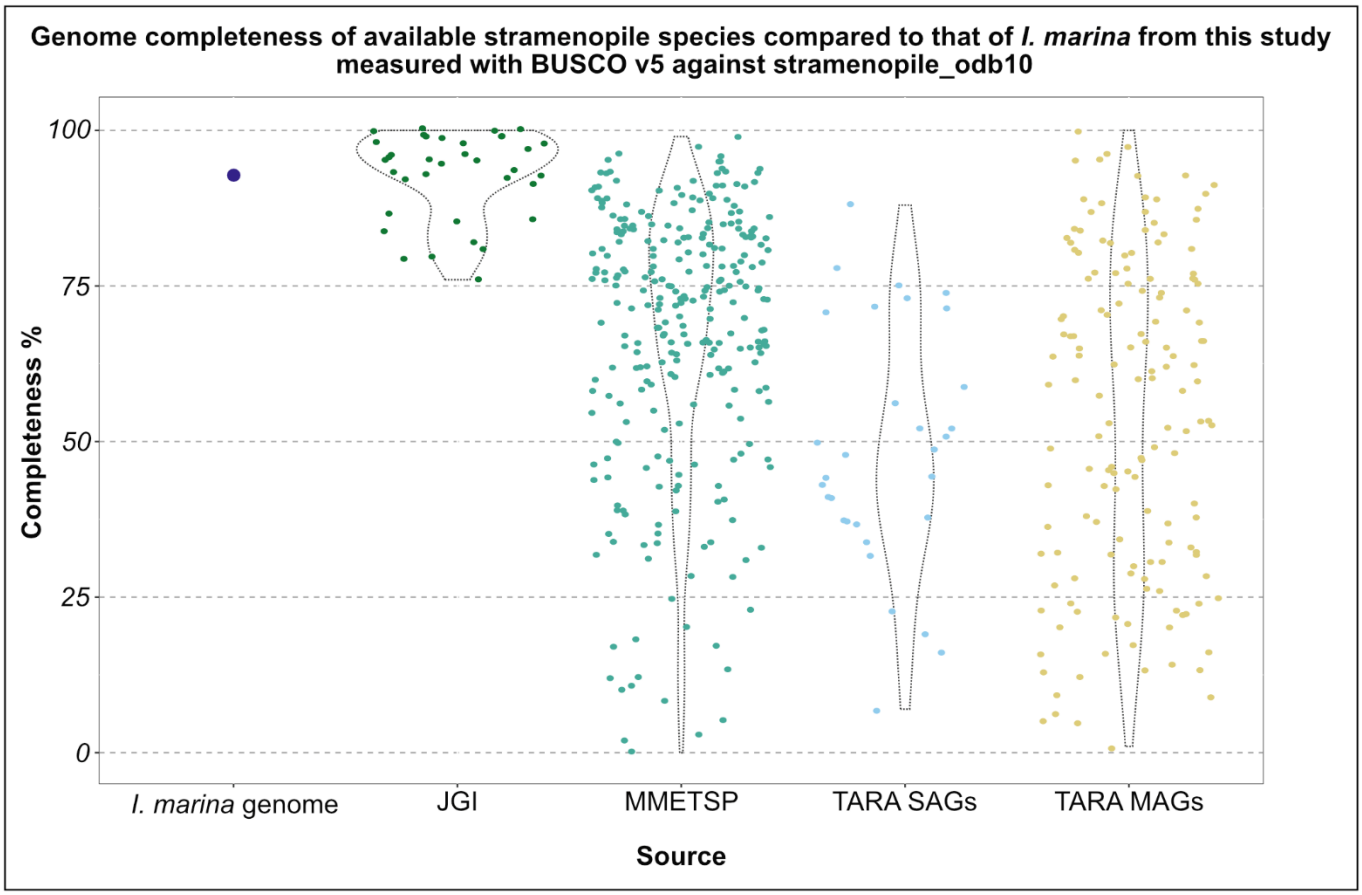
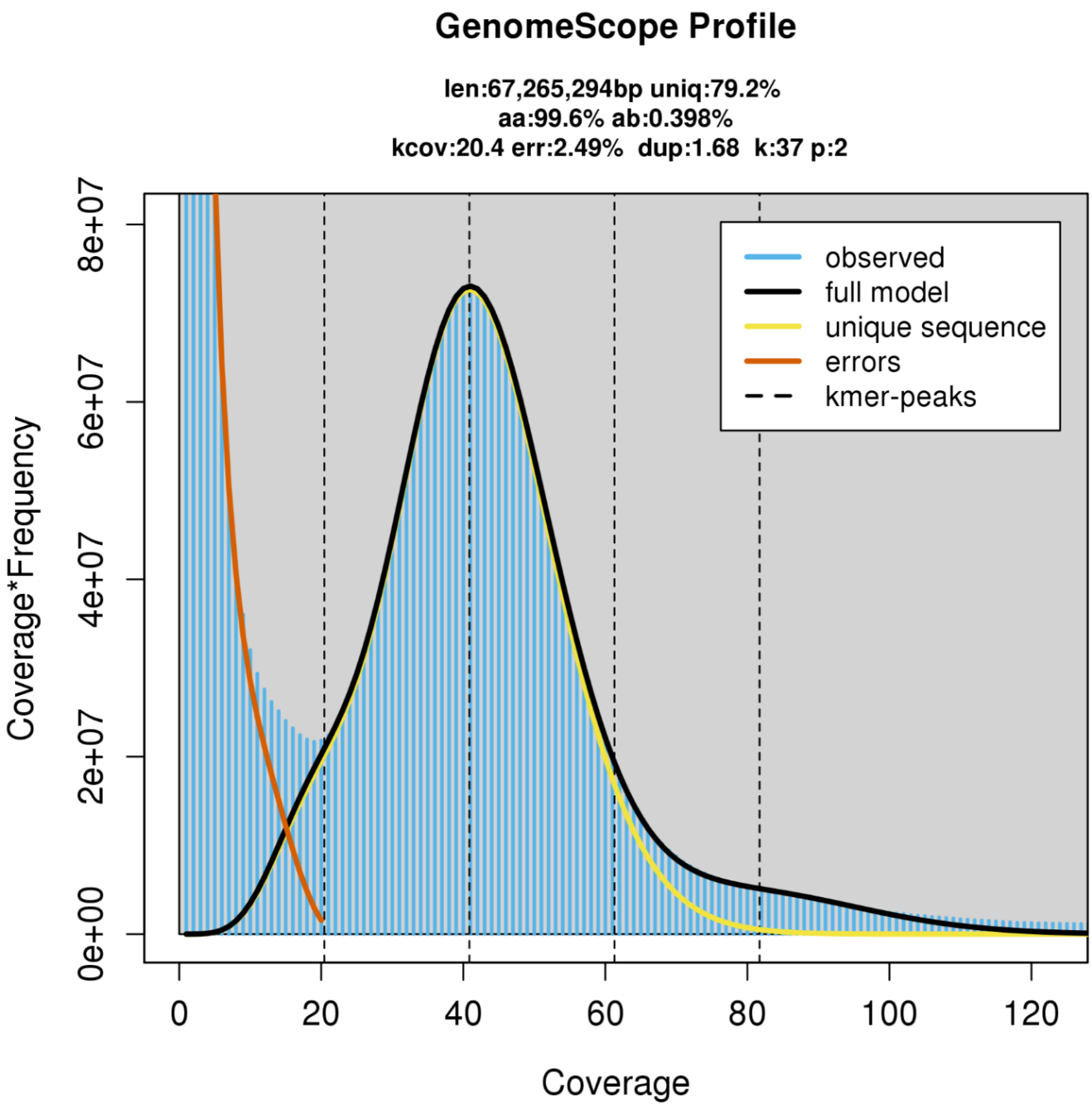


Figure 4.12 - Comparison of the genome completeness estimations for available stramenopile genomes and transcriptomes versus the estimated completeness I. marina MAG generated in this study (93%). *Measured by BUSCO (v5) against stramenopile_odb10.*




```
GenomeScope version 2.0

p = 2
k = 37

property                  min                 max
Homozygous (aa)           99.5901%            99.6141%
Heterozygous (ab)         0.3859%             0.40985%
Genome Haploid Length     67,125,052 bp       67,265,294 bp
Genome Repeat Length      13,953,246 bp       13,982,398 bp
Genome Unique Length      53,171,806 bp       53,282,896 bp
Model Fit                 85.3674%            99.0997%
Read Error Rate           2.48741%            2.48741%
```

***Figure 4.13 - Plot of k-mer frequencies in the reads used for the I. marina MAG**. k-mer size of 38. Single peak demonstrates the haploid nature of the genome. Plot generated by Genomescope2.*

diatoms, oomycetes and the closely related *C. roenbergensis* (Albertin and Marullo, 2012). To check the ploidy level of the genome reads mapping to the *I. marina* bin of the metagenome were recovered and k-mer frequencies were calculated from these *I. marina* specific reads. Figure 4.13 shows a plot of k-mer frequencies generated using genomescope2 (Ranallo-Benavidez, Jaron, and Schatz 2020; Vurture *et al.,* 2017). The single peak indicates a ploidy level of one indicating a haploid genome. Genomescope2 also estimates genome size as well as amount of unique sequence and repeat sequence. The results of this agree with the assessment of a haploid genome (predicted haploid genome size equal to that of the assembled genome). Note a relatively large k-mer size was used to account for the relatively high rate of errors in the long-read data. Two annotation pipelines were used to provide function predictions for the predicted genes. The first tool used was GhostKOALA which annotates the predicted protein sequences with KEGG orthologues (Kanehisa *et al.,* 2016b). Only 30.8% of predicted protein sequences were functionally annotated, this is comparable to annotating the *P. tricornutum* genome by the same method which yields 32.7% of proteins annotated. The context and importance of some of the annotated functions are in the results sections that follow.

GhostKOALA also annotates each predicted protein sequence with a taxonomic assignment based on the best hit in the KEGG GENES database for each query. A summary of the annotations can be found in Figure 4.14. The taxonomy of the majority of the closest hits to the *I. marina* genome are from other protist species (67.6%). The next most frequent group to be annotated as closest hit come from the animalia (28.4%). Around 20% of sequences had a closest hit from a bacterial sequence in the database, although this only represents 6.3% of all genes. This could either indicate HGT events or contamination in the assembly. Interestingly, when analysed the same way, *P. tricornutum* has a closest hit to stramenopile sequences for nearly all of its proteins, indicating a bias in the KEGG GENES database towards the Gyrista (no representatives from the Bigyra).

The second annotation pipeline employed was EggNOG mapper (Cantalapiedra *et al.,* 2021; Huerta-Cepas *et al.,* 2019). EggNOG provides annotations from a variety of functional databases. One of these is the NCBI COG database (Galperin *et al.,* 2021). A summary of COG annotations of the *I. marina* genome can be seen in Figure 4.15 where the COG annotations of three other stramenopiles, of varying evolutionary distance from *I. marina*, are provided as a comparison of the overall functional landscape of each genome. The functional repertoire of *I. marina* when analysed at this macro level appears to be similar to these other stramenopiles. Notably the most frequent

annotation in each of the genomes is “Function Unknown”. This percentage is on top of the genes that have no annotation at all, which in the *I. marina* genome makes up ~70% of all sequences (i.e., only 20% of *I. marina* genes are annotated with a known function).

4.3.6.3 B vitamin biosynthesis

To assess the interdependencies of the community with regards to B vitamin metabolism an assessment of each member of the community’s ability to biosynthesise the B vitamins was made. The results of this analysis can be seen in Figure 4.16. Biosynthesis pathway completeness was estimated using the Anvi’o metabolism estimation tool after annotation with KEGG orthologues. In the 21 bacterial bins with over 90% completion 6 were predicted to have a full thiamine biosynthesis pathway (3 others possessed the thiamine salvage pathway - *Muricauda* (bin 10), *Hoeflea* (bin 21) and *Thalassospira* (bin 25)). Nineteen of the 21 were predicted to have the ability to biosynthesise B_2. Thirteen were predicted to biosynthesise niacin with 4 of these predicted to have both aspartate and tryptophan (kynurenine) pathway routes. 13 were predicted to biosynthesise vitamin B_5. Almost every species was predicted to convert B_5 into CoA. Only 5 of the bacterial MAGs were predicted to have complete biosynthesis pathways for pyridoxal. 9 were predicted to produce biotin with a further 2 converting the intermediate pimeloyl-ACP/CoA to biotin. 13 were predicted

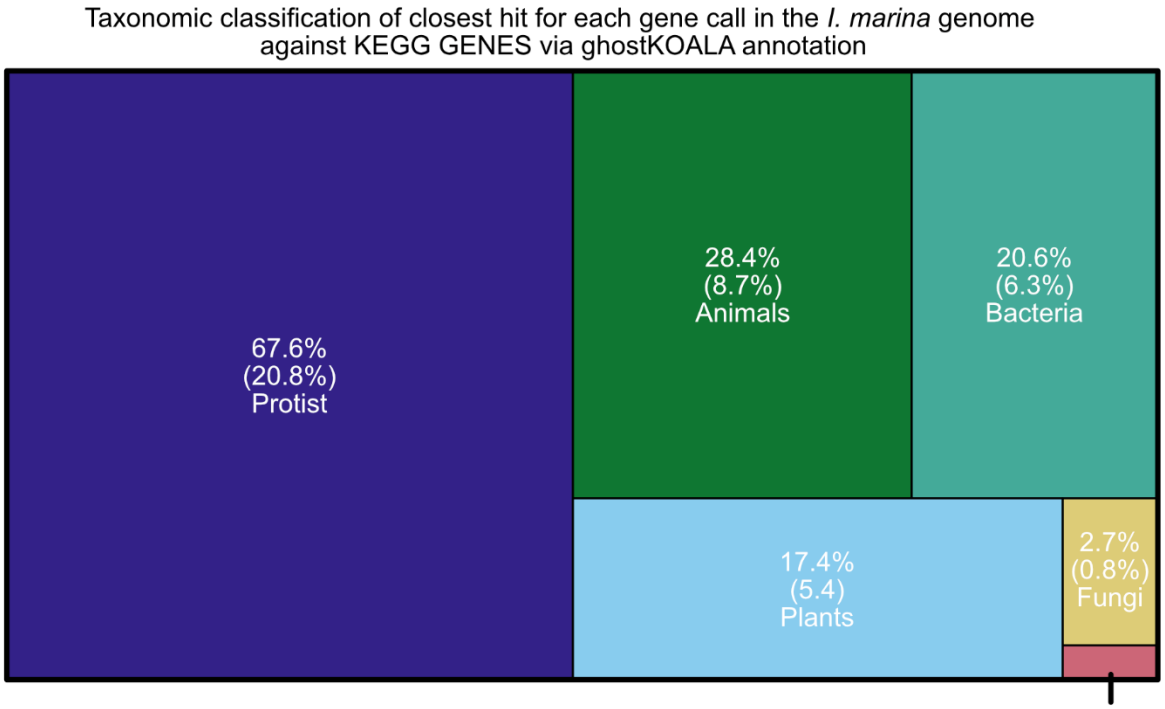


Figure 4.14 - Graphical representation of the kingdom-level taxonomic rank of the closest hit to each annotated gene in the I. marina genome against the KEGG GENES database. Percentages reported are the percentage of that classification for genes that had a functional classification (6225 genes). Percentage in brackets is the percentage of that classification in the whole data set (20238 genes)

to produce folate. 6 of the bacteria had full pathways for B_{12} production. A further 5 had partial pathways, forming cobalamin from the intermediate metabolite cobyrinate.

No single species was predicted to produce the full suite of B vitamins. The most a single species had was 7 plus a partial pathway. This occurred in four of the species, one of which was *Pseudoaltermonas sp,* predicted to have full pathways for B_1, B_2, B_3, B_6, B_7, B_9 and to produce B_{12} from the metabolite cobyrinate, but did not encode either route to pantothenate (B_5) production (aspartate or spermine), although genes for all the enzymes to CoA were present. *Aclanivorax sp.* and *M. hydrocarbonoclasticus* were similar with the same predicted pathways with the exception that they encoded a full pathway for pantothenate biosynthesis but not thiamine. Finally, *Alteromonas sp.* was predicted to have the pathways B_1, B_2, B_3, B_5, CoA, B_6, B_7 from pimeloyl ACP/CoA and B_9. The species with the fewest predicted biosynthesis pathways (of species with >90% completion) was *Pararhizobium* sp, which was only predicted to produce vitamins B_2 and B_7.

The *I. marina* genome was also analysed for B vitamin biosynthesis enzyme genes. This was carried out for two main reasons, firstly to corroborate the results seen in the transcriptomic data analysed in Chapter 3 and secondly to match the methodology used in analysing the pathways in the bacterial MAGs. *I. marina* was predicted to have complete pathways for the vitamins B_3, B_5 (as well as conversion to CoA) and B_9. Those pathways that were predicted to be complete by Anvi'o were also manually verified using BLAST and InterPro domains (matching the methodology of Chapter 3). These manual checks were performed on the Augustus gene calls rather than the Anvi'o gene calls (as used for Anvi'o metabolism estimation of the bacterial bins). The biosynthesis capabilities assessed from this genome assembly corroborate the predictions made using the *I. marina* transcriptome in Chapter 3. From these two lines of evidence, it is likely that *I. marina* is auxotrophic for the vitamins B_1, B_2, B_6 and B_7. On top of these the presence of METH but not METE in the *I. marina* genome means that it will also require an external supply of B_{12}.

4.3.6.4 Nitrogen assimilation, urea cycle and amino acid metabolism

Nitrogen is an essential macronutrient for all organisms due to its use in a plethora of biomolecules. A significant proportion of the source of N in the aquatic environment is made up of the molecules nitrate and ammonia and the processing of these by marine microbes contributes significantly to the cycling of nitrogen in the atmosphere (Flynn and Butler, 1986; Pajares and Ramos, 2019). Processing of ammonia into organic nitrogen is performed by glutamine synthase and glutamine oxoglutarate aminotransferase (GOGAT). These two reactions subsequently feed

into the biosynthesis of amino acids and thereby into other N-containing compounds (Miflin and Lea, 1980). To gain an appreciation of the nitrogen metabolism capabilities in *I. marina*, homologues to enzyme required by the urea cycle and nitrogen/ammonia assimilation were searched for by sequence similarity. The required enzymes for the conversion of ammonia to glutamine and from glutamine to glutamate were identified in *I. marina.* This demonstrates the ability to assimilate ammonia into amino acid metabolism. The relevant KO, EC numbers and *I. marina* protein sequence IDs can be found in Table 4.9.

-

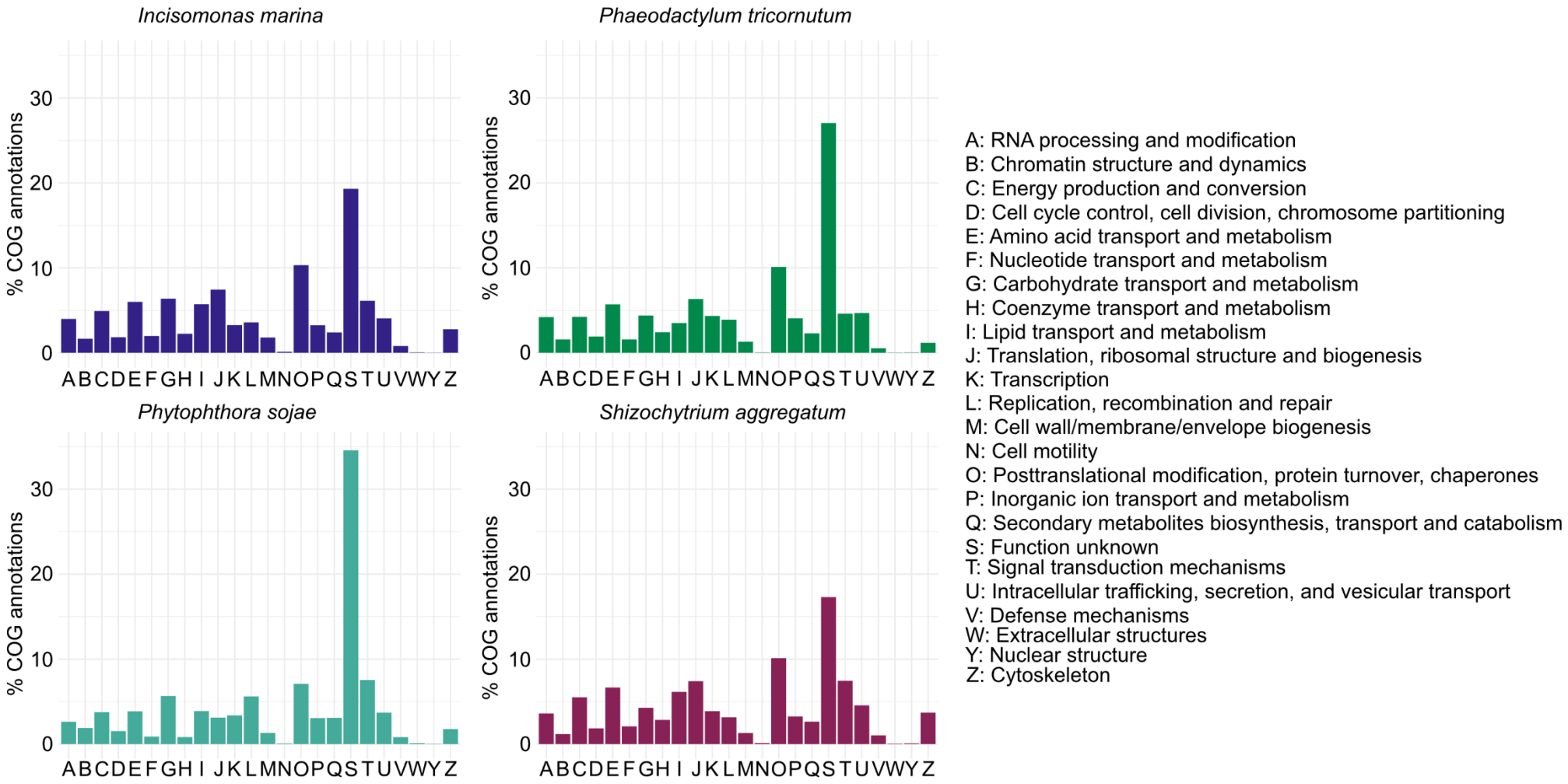


***Figure 4.15 - Comparison of the percentage of COG category assignments across the annotated genes of four stramenopiles**: I. marina from this study, P. tricornutum, P. sojae and S. aggregatum. These represent groups spread across the stramenopiles: nanamonadea, bacillariophyceae, oomycota and labyrinthulae respectively. Each organism displays a relatively similar functional landscape at this macro-scale of annotation.*

Figure 4.16 - Predictions of full B vitamin biosynthesis pathways *in each bin of the metagenome where a circle represents instances where a pathway is predicted to be complete. Predictions performed using the Anvi'o metabolism estimation workflow. Completness data is as determined using Anvi'o completness function. The I. marina completness also displays BUSCO completness (green).*

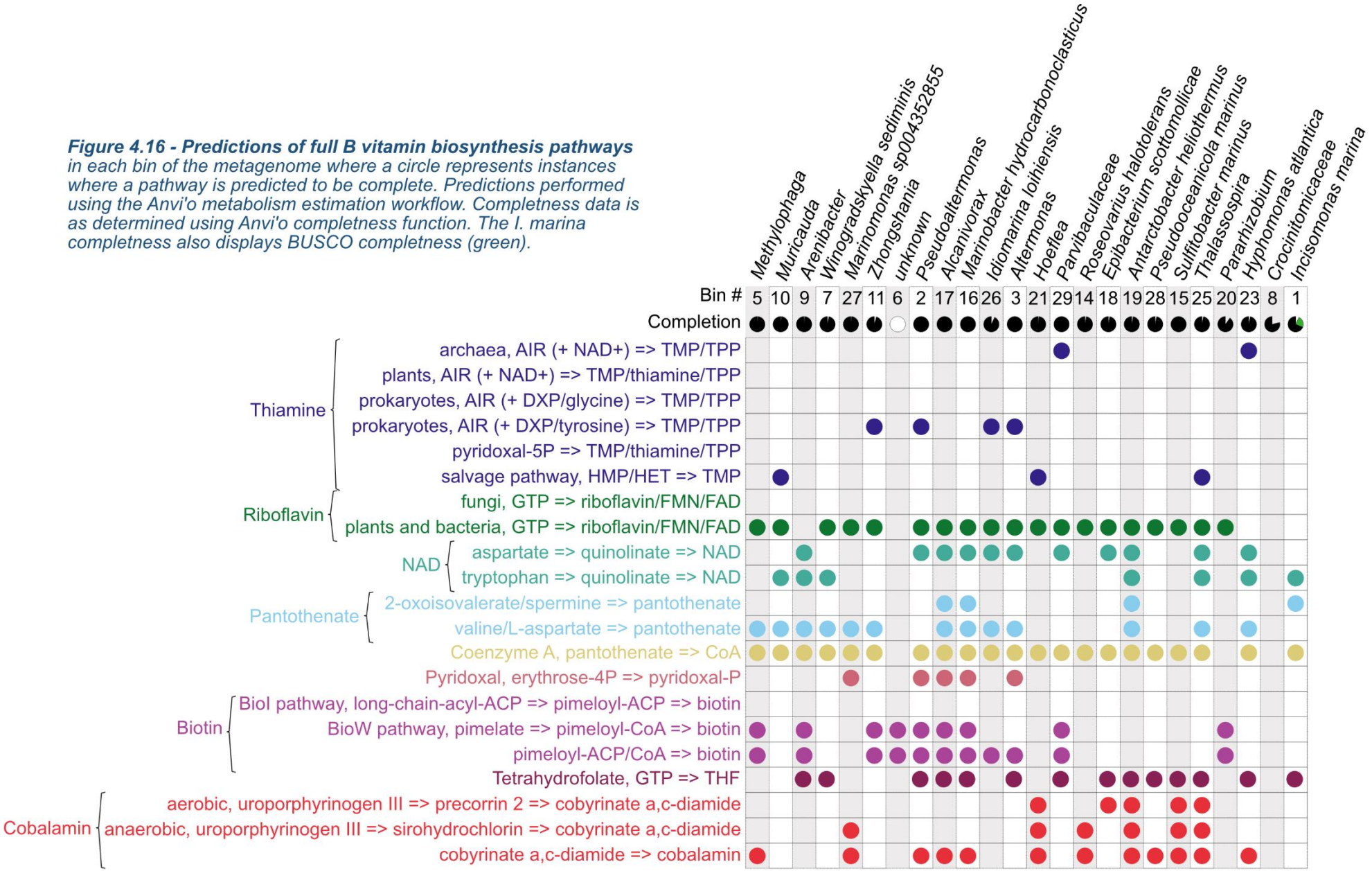


Table 4.9. Nitrogen assimilation enzymes identified in the genome of Incisomonas marina

Enzyme	E.C No.	KEGG orthologue No.	Reaction	*Incisomonas marina* protein ID
Glutamine synthetase	6.3.1.2	K01915	Ammonia -> Glutamine	g18223.t1, g7363.t1, g14366.t1
GOGAT	1.4.1.13	K00265	Glutamine -> Glutamate	g9516.t1

In metazoans ammonia is fed into the urea cycle as a means to prevent the build-up of ammonia to toxic levels. It was thought that this metabolic cycle evolved within the metazoans. However, the urea cycle was recently identified in diatoms (Allen *et al.,* 2011), suggesting this pathway evolved long before the metazoan split in the EToL. Although the metabolic steps appear the same as in the metazoan group the function of the pathway seems to be different, performing a nitrogen packaging and distribution role in diatoms as opposed to the excretion role in metazoans. A complete urea cycle was also identified here demonstrating the ability of *I. marina* to feed ammonia into the citrate cycle as well as synthesise urea.

Despite their importance in metabolism, in a similar fashion to the B vitamins, many organisms are auxotrophic for various amino acids (AAs). For this reason, and to develop a fuller picture of which species in this system are providing which nutrients, predictions of complete AA metabolism pathways for each MAG were made. The results of these predictions can be found in Figure 4.17 a-b. In similar fashion to the results of the B vitamin biosynthesis pathways for this community no single organism was predicted to biosynthesise every AA. Not least because two of the AAs, the aromatic AAs tyrosine and phenylalanine, were predicted to be biosynthesised only by the organism represented by bin 27 (predicted to be a *Marinomonas* species).

Similarly, *I. marina* appears to lack the ability to biosynthesise most of the AAs *de novo* suggesting it satisfies its requirement for these nutrients via its diet. Those AAs that *I. marina* was predicted to biosynthesise were arginine, tryptophan, leucine, methionine, histidine, glutathione and serine. In contrast the bacterial species in the community cover the biosynthesis of all of the AAs with a high level of redundancy in the system with the exception of tyrosine and phenylalanine, as previously stated. For those species that are predicted to biosynthesise very few AAs they must source these from elsewhere, potentially via exchange with the other bacteria in the community or from the complex growth medium.

Figure 4.17 a-b - Predictions of full amino acid metabolic pathways *in each bin of the metagenome where a circle represents instances where a pathway is predicted to be complete. Predictions performed using the Anvi'o metabolism estimation workflow.*predicted to be complete if threshold is lowered to 70% for completness assignment.*

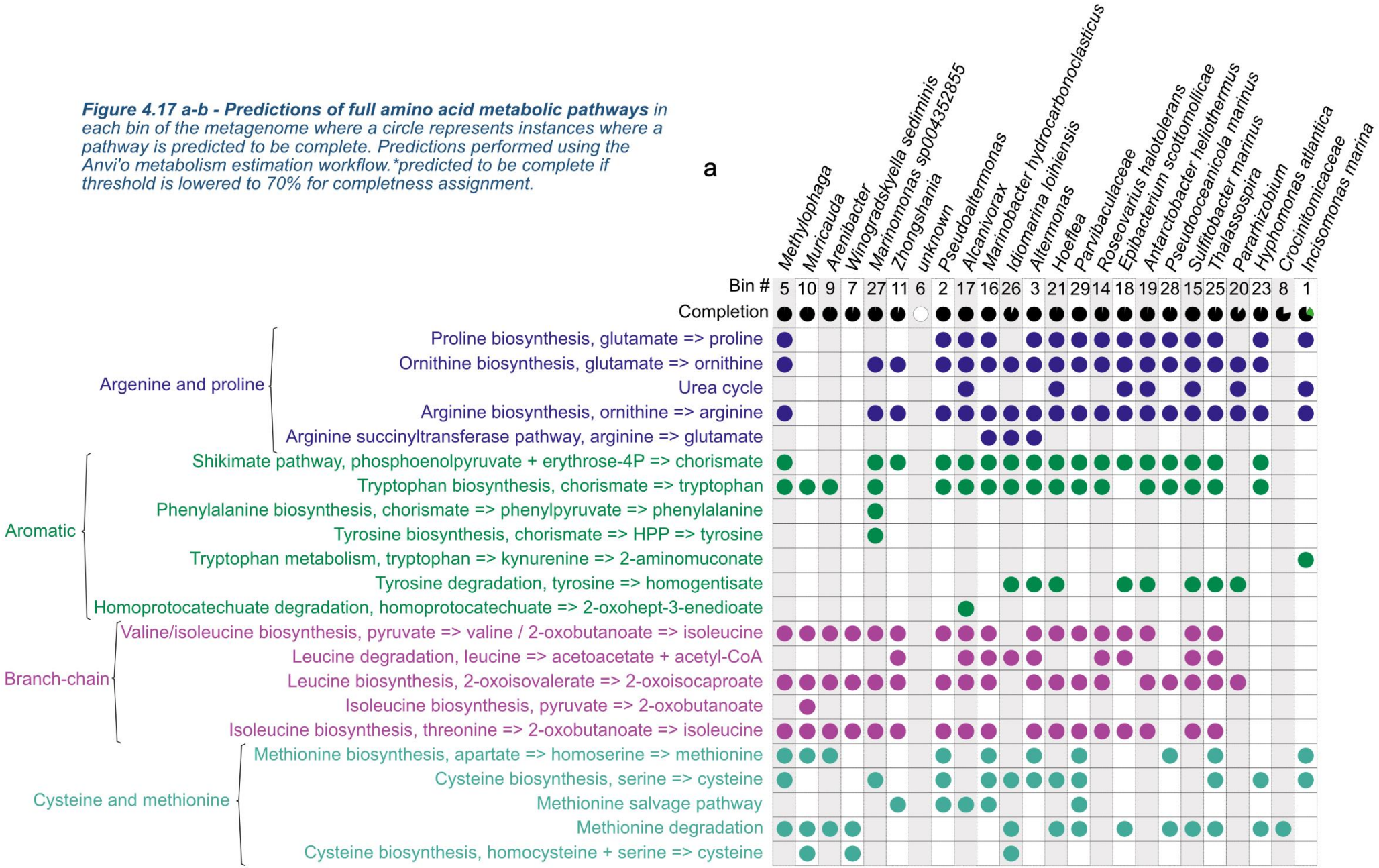


Figure 4.17 b *- see previous figure.*

An example of such an organism is from bin 7, taxonomically predicted to be *W. sediminis*, which is predicted to have full biosynthesis pathways for only five AAs (valine, leucine, isoleucine, cysteine and serine).

4.3.6.5 Eukaryotic-Bacterial Signalling

Cross-kingdom interactions are thought to be commonplace in the marine microbial environment. These types of interactions have been characterised to some degree in the more readily studied, photosynthetic members of the stramenopiles, the diatoms (Bacillariophyceae). Rosmarinic acid has been suggested as a signalling molecule in the diatom *A. glacialis* (Shibl *et al.,* 2020). It was demonstrated that this metabolite reduced motility of associated bacteria enhancing favourable interactions for the diatom. Given the close proximity with which *I. marina* and its bacteria reside a similar mechanism would be conceivable. However, as with the results from Shibl et al. (2020) for the diatom the biosynthesis pathway for this secondary metabolite was not found to be present in *I. marina*. In both cases this genome mining was performed by looking for homologues of known plant biosynthesis enzymes. The biosynthesis pathway in *A. glacialis* is therefore different to the plant pathway and yet to be identified. If *I. marina* does use this molecule in the same way, it too would likely use a different pathway from the plant pathway. Tryptophan has also been reported to be an important signalling molecule in the marine environment and suggested as a potential bacteria-diatom messenger (Amin *et al.,* 2015). As discussed above the biosynthesis pathway for this AA was also not found to be present in *I. marina* although it was present in the majority of the bacterial MAGs.

DMSP has also been suggested as in important signalling molecule for algae to bacteria. DMSP is broken down by bacteria to form dimethyl sulfide (DMS)(Zhang *et al.,* 2019). DMSP from the eukaryotic cell can act as a chemoattractant and stimulates the production of quorum sensing molecules (Cirri and Pohnert 2019; Seymour *et al.,* 2010; Johnson, Kido Soule, and Kujawinski 2016). DMSP is synthesised via three alternative pathways, however the one found in algae is the transamination pathway (Shaw *et al.,* 2022). This pathway can be seen in Figure 4.18. The only characterised enzyme in this pathway is the methylation enzyme. Two alternative forms of this enzyme have been identified. DSYB is found in many DMSP producing algal species (Curson *et al.,* 2018). A putative DSYB protein sequence was identified in the genome of *I. marina*. Figure 4.19 shows this putative DSYB protein in a phylogenetic tree with other putative DSYB sequences

identified by Curson et al. (2018). The tree shows the *I. marina* sequence grouping with other stramenopile species, including that of *F. cylindrus,* which has been functionally validated

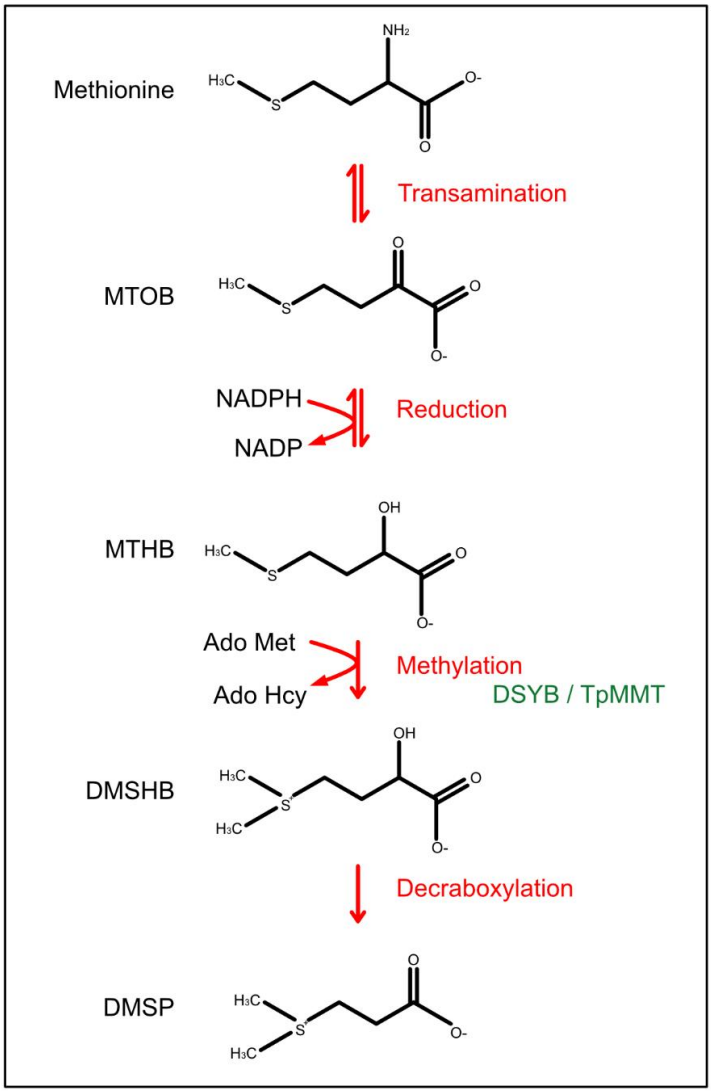


Figure 4.18 - Proposed transamination biosynthesis pathway for DMSP from methionine in algae. *The proposed methylation enzyme DSYB, indicated in green, has been identified in a number of species (see Figure 4.19). The MMT enzyme identified by Kagyama et al. (2018) in Thalassiosira pseudonana is an alternative proposed methylation enzyme.MTOB - 4-methylthio-2-oxobutyrate; MTHB - 4-methylthio-2-hydroxybutyrate; DMSHB - 4-dimethylsulphonio-2-hydroxybutyrate. Pathway constructed from Stefels (2000), Curson et al. (2017) and Shaw et al. (2022).*

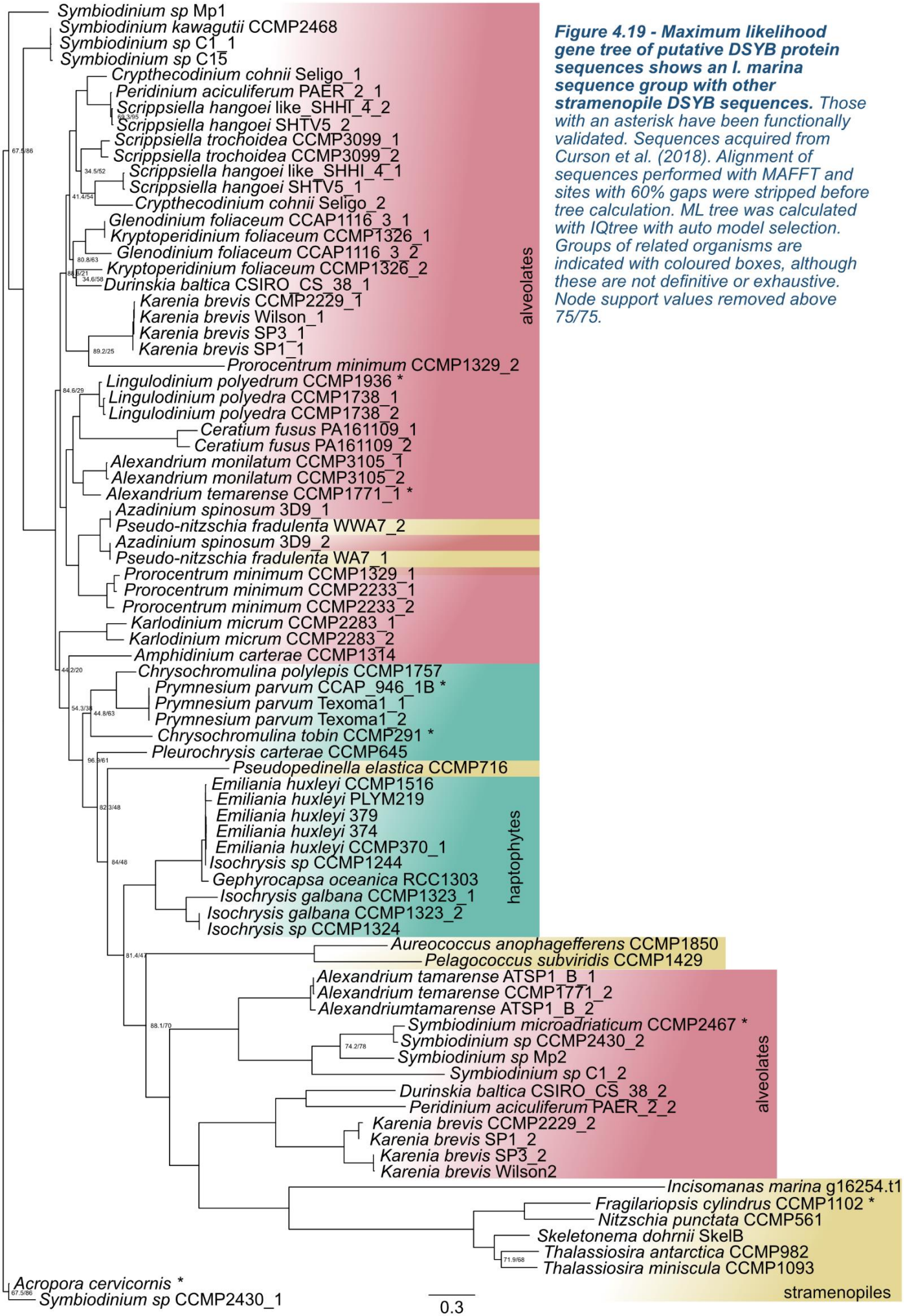


Figure 4.19 - Maximum likelihood gene tree of putative DSYB protein sequences shows an I. marina sequence group with other stramenopile DSYB sequences. *Those with an asterisk have been functionally validated. Sequences acquired from Curson et al. (2018). Alignment of sequences performed with MAFFT and sites with 60% gaps were stripped before tree calculation. ML tree was calculated with IQtree with auto model selection. Groups of related organisms are indicated with coloured boxes, although these are not definitive or exhaustive. Node support values removed above 75/75.*

This suggests that *I. marina* has the potential to biosynthesise DMSP and this might function as a molecule for signalling in the community. The other enzyme is referred to as MTHB-methyltransferase (MMT) and has been identified in *T. pseudonana* by Kagyama et al. (2018). A putative homologue of this enzyme was identified in *I. marina* (g8551.t1) when the TpMMT sequence was used as a query in a BLAST search. However, the hit was relatively low in similarity (e val = $3.25e^{-07}$), especially compared to the putative DSYB sequence which had strong similarity to the query DSYB sequence from *F. cylindrus* (e val = $1.27e^{-55}$).

DMS is a volatile form of sulphur, highly abundant in the oceans , and along with DMSP contribute significantly to the global cycling of sulphur in the atmosphere (Zhang *et al.,* 2019). DMS is produced by the cleavage of DMSP. Only one of the bacterial MAGs contained a homologue to the DMSP cleavage enzyme dddL (EC:4.4.1.3) – bin 14, *Roseovarius halotolerans*. The eukaryotic DMSP cleavage enzyme, dimethylsulfonioproprionate lyase, was identified in the haptophyte *Emiliania huxleyi* (Alcolombri *et al.,* 2015). No homologue to this enzyme was identified in the eukaryotic bin belonging to *I. marina*. This suggests that either *I. marina* does not perform the conversion of DMSP to DMS or it uses an alternative lysase enzyme yet to be identified.

4.3.6.6 Red algal linked genes and the chromalveolate hypothesis

The ability to photosynthesise is a trait that is distributed across the EToL in a polyphyletic manner. The trait first developed in eukaryotes by a process of endosymbiosis of a then free-living photosynthetic cyanobacterium by a heterotrophic eukaryotic cell. This is known as the primary endosymbiosis event which gave rise to the green and red algae (see Chapter 1 Section 1.1) (Gould *et al.,* 2008). For other lineages, however, the acquisition of this trait is not so clear. It is likely due to a secondary endosymbiosis event and in some cases tertiary or serial endosymbiosis (e.g., members of the dinoflagellates). Groups which have this red-algal derived plastid include the photosynthetic stramenopiles, haptophytes and dinoflagellates, each of which fall within different parts of the EToL (Burki *et al.,* 2020).

There are two conflicting proposals for the evolution of plastids within these groups. One is the independent acquisition hypothesis which reasons multiple individual endosymbiosis events in the early history of each of the phylogenomic groups mentioned above (Archibald, 2009). The other hypothesis, the chromalveolate hypothesis, was first proposed by Cavalier-Smith (1999). Because of the commonality of a chloroplast with red algal characteristics he proposed a common ancestor for all of these groups with a single red-algal endosymbiosis event from which these groups

subsequently evolved. The rationale was that this type of event is too complex and unlikely to occur multiple times independently (Green 2011). The problem with this hypothesis is the nature of the other members of the groups in which these photosynthetic organisms are found. Many heterotrophic species are closely related to these phototrophs: ciliates are heterotrophic members of the Alveolata group to which the photosynthetic dinoflagellates also belong, oomycetes are heterotrophic members of the stramenopile group related to the photosynthetic diatoms. However, neither ciliates nor oomycetes have any evidence of ever possessing a plastid (Stiller *et al.,* 2009; Wang, Sun, and Huang 2017), bringing this hypothesis in to question. In light of this, evidence for a red-algal like plastid was searched for in the genome of *I. marina*, an organism in the basal lineages of the stramenopiles. If the chromalveolate hypothesis were correct one would expect to find some relics of the plastid in these basal heterotrophic stramenopile groups.

Of the 4208 "Protist" annotated genes from the ghostKOALA annotation results, 113 genes had a red alga as the highest hit against the KEGG GENES database. A full list of red-algal annotated genes can be found in Table 4.10. Of these 117 genes, 78 had functional annotations with 50 having unique KO numbers assigned. Of the 50 unique annotations with red-algal taxonomy only one was identified as having a predicted protein product with localisation to the chloroplast of photosynthetic species: *I. marina* gene ID - g17092.t1, KO number - K15918, GLYK; D-glycerate 3-kinase [EC:2.7.1.31]. Tyler et al. (2006) identified 12 gene products in the non-photosynthetic stramenopile *P. ramorum* that were predicted to have mitochondrial localisation in that species but plastid localisation in algae/plants. They suggest that genes with this type of history are most likely to provide evidence of a common photosynthetic ancestor to the stramenopiles. None of these twelve genes identified match the red-algal annotated genes in the *I. marina* genome. It seems unlikely therefore that the MAST3 lineage had chloroplasts at any point during their evolutionary history.

Table 4.10 – Red algal annotated genes in I. marina. Annotated by ghostKOALA. Evidence for the chloroplast associated gene https://www.arabidopsis.org/servlets/TairObject?type=locus&name=AT1G80380

Gene ID	KO number	Description	Plastid linked?
g44.t1	K08288	PRKCSH; protein kinase C substrate 80K-H	n
g504.t1	K19612	PDE12; 2',5'-phosphodiesterase [EC:3.1.13.4 3.1.4.-]	Mitochondrial
g569.t1	K18156	ATP23, XRCC6BP1; mitochondrial inner membrane protease ATP23 [EC:3.4.24.-]	Mitochondrial
g1394.t1	K01875	SARS, serS; seryl-tRNA synthetase [EC:6.1.1.11]	Mitochondrial
g1556.t1	K00942	gmk, GUK1; guanylate kinase [EC:2.7.4.8]	n
g1679.t1	K05917	CYP51; sterol 14alpha-demethylase [EC:1.14.14.154 1.14.15.36]	n
g1705.t1	K07305	msrB; peptide-methionine (R)-S-oxide reductase [EC:1.8.4.12]	n
g1872.t1	K18156	ATP23, XRCC6BP1; mitochondrial inner membrane protease ATP23 [EC:3.4.24.-]	Mitochondrial
g1888.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g1969.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g2756.t1	K20347	TMED2, EMP24; p24 family protein beta-1	n
g3076.t1	K12821	PRPF40, PRP40; pre-mRNA-processing factor 40	n
g3633.t1	K08288	PRKCSH; protein kinase C substrate 80K-H	n
g3942.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g4827.t1	K11498	CENPE; centromeric protein E	n
g4368.t1	K08796	BRSK; BR serine/threonine kinase [EC:2.7.11.1]	n
g4619.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g4872.t1	K11498	CENPE; centromeric protein E	n
g4885.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g5477.t1	K11498	CENPE; centromeric protein E	n
g5867.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g6309.t1	K08288	PRKCSH; protein kinase C substrate 80K-H	n
g7000.t1	K20347	TMED2, EMP24; p24 family protein beta-1	n
g7257.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g7265.t1	K13703	ABHD11; abhydrolase domain-containing protein 11	n
g8047.t1	K12821	PRPF40, PRP40; pre-mRNA-processing factor 40	n
g8708.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g9448.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g9925.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g10230.t1	K17777	TIM9; mitochondrial import inner membrane translocase subunit TIM9	Mitochondrial
g10403.t1	K12821	PRPF40, PRP40; pre-mRNA-processing factor 40	n
g10658.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g11052.t1	K03469	rnhA, RNASEH1; ribonuclease HI [EC:3.1.26.4]	n
g11660.t1	K12821	PRPF40, PRP40; pre-mRNA-processing factor 40	n
g11580.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g11705.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g12317.t1	K18703	SUGCT; succinate---hydroxymethylglutarate CoA-transferase [EC:2.8.3.13]	Mitochondrial

g13694.t1	K11498	CENPE; centromeric protein E	n
g16639.t1	K01875	SARS, serS; seryl-tRNA synthetase [EC:6.1.1.11]	Mitochondrial
g12792.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g12837.t1	K12821	PRPF40, PRP40; pre-mRNA-processing factor 40	n
g13091.t1	K24887	GTPBP1; GTP-binding protein 1	n
g13095.t1	K00939	adk, AK; adenylate kinase [EC:2.7.4.3]	n
g13290.t1	K06675	SMC4; structural maintenance of chromosome 4	n
g13486.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g13520.t1	K12821	PRPF40, PRP40; pre-mRNA-processing factor 40	n
g13581.t1	K15111	SLC25A26; solute carrier family 25 (mitochondrial S-adenosylmethionine transporter), member 26	Mitochondrial
g13694.t1	K11498	CENPE; centromeric protein E	n
g13783.t1	K12812	DDX39B, UAP56, SUB2; ATP-dependent RNA helicase UAP56/SUB2 [EC:3.6.4.13]	n
g13784.t1	K12812	DDX39B, UAP56, SUB2; ATP-dependent RNA helicase UAP56/SUB2 [EC:3.6.4.13]	n
g13943.t1	K01726	GAMMACA; gamma-carbonic anhydrase [EC:4.2.1.-]	Mitochondrial
g13948.t1	K08288	PRKCSH; protein kinase C substrate 80K-H	n
g14169.t1	K11090	LA, SSB; lupus La protein	n
g14773.t1	K11498	CENPE; centromeric protein E	n
g15451.t1	K00942	gmk, GUK1; guanylate kinase [EC:2.7.4.8]	n
g15645.t1	K06002	PGA; pepsin A [EC:3.4.23.1]	n
g16100.t1	K08288	PRKCSH; protein kinase C substrate 80K-H	n
g16227.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g16320.t1	K18932	ZDHHC; palmitoyltransferase [EC:2.3.1.225]	n
g16412.t1	K03469	rnhA, RNASEH1; ribonuclease HI [EC:3.1.26.4]	n
g16639.t1	K01875	SARS, serS; seryl-tRNA synthetase [EC:6.1.1.11]	n
g17092.t1	K15918	GLYK; D-glycerate 3-kinase [EC:2.7.1.31]	Chloroplast, mitochondria, nucleus
g17171.t1	K15111	SLC25A26; solute carrier family 25 (mitochondrial S-adenosylmethionine transporter), member 26	n
g17341.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g17981.t1	K05906	PCYOX1, FCLY; prenylcysteine oxidase / farnesylcysteine lyase [EC:1.8.3.5 1.8.3.6]	n
g18240.t1	K11498	CENPE; centromeric protein E	n
g18296.t1	K18932	ZDHHC; palmitoyltransferase [EC:2.3.1.225]	n
g18501.t1	K18932	ZDHHC; palmitoyltransferase [EC:2.3.1.225]	n
g17341.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g18514.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g18609.t1	K08592	SENP1; sentrin-specific protease 1 [EC:3.4.22.68]	n
g19413.t1	K07305	msrB; peptide-methionine (R)-S-oxide reductase [EC:1.8.4.12]	n
g19479.t1	K01726	GAMMACA; gamma-carbonic anhydrase [EC:4.2.1.-]	n
g19493.t1	K08288	PRKCSH; protein kinase C substrate 80K-H	n

4.3.6.7 Evidence of viral DNA in the community

To achieve a well-rounded assessment of the biotic interactions of the community analysis was performed to identify evidence for viral DNA. Virus nucleotide sequences from the NCBI RefSeq database (579,256 sequences) were used as query sequences in BLAST searches against the whole metagenomic assembly as well as against the *I. marina* genome separately. No hits were detected in the search against *I. marina*. However, within the metagenome as a whole there was significant evidence of viral sequences. The full list of hits can be seen in Appendix Section 3.4. The most frequently occurring accession was on contig 453 where there were 20 hits from the virus *Pseudoalteromonas* phage C5a (accession no. NC_047790). There was also a combined total of 33 hits on contig 596 for enterobacteria-type phages (accession numbers NC_019723.1, NC_001416.1, NC_019711.1, NC_049951.1).

The second largest virus ever detected (by genome size) was identified from the organism *C. roenbergensis* (Fischer *et al.,* 2010), a heterotrophic flagellate , and one of the closest relatives of *I. marina* that is kept in culture and has a sequenced genome (Hackl *et al.,* 2020). To look for any evidence of this giant-virus in the *I. marina* genome the predicted AA sequences from the genome of *C. roenbergensis* virus (CroV) were used as queries in a BLAST search against the metagenome. Of the 544 viral predicted AA sequences 86 had hits (e-value threshold of <10-10). The full list of BLAST hits for these top hits as well as protein domain predictions from InterPro can be found in Appendix Section 3.5. That strong hits were not identified for all of the CroV proteins suggests that this virus is not infecting *I. marina* as half of the CroV proteins are known to hit eukaryotic sequences (Fischer *et al.,* 2010).

A different class of virus, the mavirus, has been found to infect CroV itself and has been proposed as an anti-viral system for the protist that it infects. It purportedly integrates into the genome of the host (Fischer and Hackl, 2016) and is transcriptionally activated during CroV infection. Fischer and Hackl (2016) found a 19,055bp section of 100% identity to the reference genome of the mavirus in question (GenBank accession number HQ712116) in the genome of *C. roenbergensis.* To identify any mavirus-like sequences in the *I. marina* genome the reference genome mentioned above was used as a query to search against the *I. marina* genome (nucleotide sequences using blastn). No hits were found. On top of this the predicted protein sequences from this mavirus were used as queries against the predicted peptide sequences of *I. marina* (blastp). Of the 20 query proteins only 3 had hits with an e-value below 10^{-5}. It is therefore unlikely that a true mavirus integration has

Table 4.11 – top hit BLASTp results for querying the I. marina predicted peptide sequences with mavirus aa sequences

Query Name	Hit Name	% Identical Sites	Bit-Score	Description	InterPro ID	E Value
tr|F1DAS4|F1DAS4_9VIRU	g4015.t1	26.3%	159.84	DNA/RNA polymerases	IPR004868, IPR043502	1.47E-41
tr|F1DAS7|F1DAS7_9VIRU	g10380.t1	26.2%	49.29	GIY-YIG endonuclease	IPR035901, IPR000305, IPR035901	2.67E-7
tr|F1DAS3|F1DAS3_9VIRU	g17158.t1	35.4%	142.12	Chromo domain-like, Ribonuclease H-like	IPR000953, IPR016197, IPR012337, IPR001584, IPR036397	1.95E-39

occurred in the *I. marina* genome. These results can be seen in Table 4.11.

4.4 Discussion

Across the eukaryotic tree of life there is an enormous anthropogenic bias in genomic sequencing (del Campo *et al.,* 2014; Burki and Keeling 2014). This is acutely apparent in the stramenopile group, which displays an enormous range of habitat, trophic mode and cell structural diversity and yet there is a large bias towards the photosynthetic lineages, predominantly the bacillariophyceae, in genomic sequencing effort. The next largest number of genomes is from the terrestrial oomycetes that threaten human-orientated plant life via parasitism. In contrast the basal lineages of the stramenopile tree are poorly understood and understudied with just a few species investigated, such as *C. roenbergensis* and MAGs from metagenomics studies of environmental samples (Delmont *et al.,* 2022; Hackl *et al.,* 2020). The latter has demonstrated considerable diversity within the Bigyra and warrants greater focus and exploration. Here the genome of a species from within this diversity has been sequenced in an effort to shed some light on these organisms. Sequencing and subsequent genomic comparisons of *I. marina* to those well studied organisms reveals some surprising similarities in metabolism despite often very different lifestyles (e.g., heterotrophic vs phototrophic), as well as notable differences. The analysis presented here takes this concept one step further and also analyses the functionality of the microbial community within which this stramenopile is maintained, revealing potential metabolic complementation essential for the maintenance of the community structure.

General observations of the I. marina MAG – The *I. marina* genome size (~68 Mbp) is large in comparison to the haploid sizes of other stramenopiles. For example, the haploid size of *P. tricornutum* and *T. pseudonana* are 32.4 Mbp and 35.6 Mbp respectively (Bowler *et al.,* 2008; Armbrust *et al.,* 2004). The closely related bikosid *C. roenbergensis* has a haploid size of ~40 Mbp (Hackl *et al.,* 2020). The haploid genome size of the oomycetes are more comparable to that of *I. marina*. For example, *P. ramorum* is 65 Mbp (Lamour *et al.,* 2007). The exception to this is *P. infestans*, which has a haploid genome size of 240 Mbp. The comparisons are the same for the number of predicted genes in each of these species. *I. marina* is predicted to have ~20,000 genes, around twice as many predicted genes as are found in the two diatom species mentioned above. Compared to *P. ramorum*, *I. marina* has 25% more predicted genes.

The instances of polyploidization in the stramenopiles is also interesting. From Figure 4.13 *I. marina* is likely to be haploid (1n) whereas diatom species such as *T. pseudonana* and *F. cylindrus* have a diploid (2n) genome (Armbrust *et al.,* 2004; Mock *et al.,* 2017). Diatoms evolved after the plastid acquisition event in the stramenopile evolutionary history. One might conclude that the polyploidization event therefore also occurred after this point. However, other members of the Gyrista (see Chapter 3 Figure 3.1) such as the oomycetes, who diverged from the rest of the Gyrista prior to the plastid acquisition event, have also evolved a diploid genome. The close relative and member of the Bigyra with a sequenced genome, *C. roenbergensis*, is also diploid. The ploidy levels within the chyrsophytes is also inconsistent (Majda *et al.,* 2021). Some members were found to be diploid (e.g., *Spumella vulgaris*), others triploid (e.g., *Dinobryon pediforme*) and some even tetraploid, such as *Pseudospumella encystans*. The authors determined that the ploidy-level of the chysophytes analysed did not correlate with nutritional mode or taxonomy. The non-uniform spread of polyploidy across the supergroup warrants further investigation.

B vitamin biosynthesis – No single species of bacteria in the community was predicted to have a full suite of B vitamin biosynthesis capabilities. This suggests that these organisms may exchange these metabolites between themselves to provide each of the cofactors required for metabolic function. Indeed syntrophy for B vitamins is a trait that has been identified in other communities of microbes such as the insect gut (Sharma *et al.,* 2019), which suggests this may be commonplace for communities from different environments. Exchange of B vitamins for other metabolites between prokaryote and eukaryote in culture has been described multiple times (e.g. Cooper *et al.,* 2018; Amin *et al.,* 2015). As *I. marina* possesses only the B_{12}-dependent methionine synthase it requires

B_{12}, which can only be from an exogenous source. It is likely that *I. marina* consumes B_{12} and many of the other B vitamins for which it is auxotrophic (B_1, B_2, B_6 and B_7) via its presumed bacterial diet (Cavalier-Smith and Scoble, 2013). An exchange system as described for algal species may be less likely as the currency used by a heterotrophic eukaryote cannot be photosynthate. In the context of the growth of *I. marina* in culture it is necessary to include a wheat grain as a source of fixed carbon for the bacteria. Additionally, the medium contains an unknown quantity of vitamins as a result of the addition of the complex nutrient source, soil extract.

The demonstration of the B vitamin auxotrophies in *I. marina* will inform future efforts to culture this species in an axenic setting. Any effort to create a defined media to supply this species with its full nutritional requirement will need to include both vitamin B_6 and B_2, neither of which are traditional B vitamins reported as a requirement in algal research. Similar to many other stramenopiles including a wide variety of photosynthetic species, vitamins B_1, B_7 and B_{12} would also need to be added. However, what this analysis does not provide is the required concentrations of these vitamins to sustain growth of *I. marina*.

Urea cycle – The urea cycle is present in metazoans but not terrestrial plants or green algae. It is a mechanism that allows these organisms to excrete urea to avoid toxic build up and for a long time was thought to have evolved from within this group. However, the discovery of CPS and the urea cycle in the marine diatom *P. tricornutum* was significant as it was demonstrated that the role of the urea cycle in this organism was to package and distribute inorganic nitrogen and therefore contributes to the organism's ability to survive nitrogen-limited conditions (Allen *et al.,* 2011). This allows diatoms to mediate carbon fixation and nitrogen metabolism to aid growth and contribute to marine productivity. Here we have a strong indication that this mechanism is also present in the non-photosynthetic MAST-3 species *I. marina*. The evolution of the CPS enzyme sees two duplication events, the second of which is proposed to have occurred before the secondary endosymbiosis event (and therefore plastid acquisition) of diatoms (and haptophytes). The identification of the "novel-diatom" CPS in *I. marina* (g8254.t1) is further evidence that this enzyme evolved before diversification of the non- and photosynthetic stramenopiles.

Genome annotations – Although an annotation of taxonomy may not indicate origin it is possible that some of the *I. marina* genes with a highest hit to bacterial species may indicate instances of HGT. However, those sequences annotated with prokaryotic taxonomy may also demonstrate

instances of contamination in the genome assembly. Indeed, that *I. marina* lives in such close proximity to bacteria means either or both of these possibilities are likely.

The chromalveolate hypothesis predicted a single evolutionary origin for the plastids in the algal groups in the stramenopiles, haptophytes, alveolates and the cryptophytes. The proposal was for a single secondary-endosymbiosis event of a red-alga giving rise to photosynthesis in these groups (Keeling, 2009). If this were the case, one would expect to find traces of this plastid in the genomes of the non-photosynthetic members of these groups, which have then presumably lost their photosynthetic capacity. Evidence contradicting this hypothesis has come from phylogenomic analysis (Burki *et al.,* 2020) as well as the lack of identification of red-algal genes in non-photosynthetic "chromalveolates" (Stiller *et al.,* 2009; Wang, Sun, and Huang 2017). Indeed, the evidence reported here also supports the results of the later.

Bacterial assemblage and signalling – A number of the identified bacteria in the culture community are also reported to be commonly associated with diatoms in culture or in environmental samples: *Alteromonas*, *Pseudoalteromonas*, *Hyphomonas*, *Sulfitobacter* and *Winogradskyella* (Amin *et al.,* 2015). Given the close proximity of *I. marina* and the bacterial cells (as demonstrated by the SEM imaging Figures 4.4-5) in the culture, and the fact that it is likely to be nutritionally dependent on the bacterial assemblage, signalling pathways identified to mediate stramenopile-bacteria interactions were examined.

DMSP and DMS are highly abundant compounds in the ocean. These dominant sulphur containing molecules have been proposed as microbial signalling molecules. Tentative evidence that heterotrophic protists may also contribute to DMSP production was obtained by the presence of a putative gene encoding DSYB, which is g16254.t1 in the *I. marina* genome. Although Lyon et al. (2011) elucidated a putative biosynthetic pathway for DMSP in the diatom *F. cylindrus* by proteomics, DSYB is the only extensively characterised enzyme associated with DMSP biosynthesis and it catalyses the methylation of 4-methylthio-2-hydroxybutyrate to 4-dimethylsulfonio-2-hydroxybutyrate. If *I. marina* were to produce DMSP it may provide a potential signalling mechanism within the culture community.

Metabolic repertoire of the community – Evidenced by the frequency with which stramenopile species are found living in close association with bacteria (Amin *et al.,* 2015), it is clear that this relationship is important for stramenopiles in general, not least for *I. marina* as they are likely its key source of nutrition. This is demonstrated by the amino acid and B vitamin biosynthesis

pathway analysis performed here. As discussed above *I. marina* is predicted to be auxotrophic for a number of B vitamins but it is also predicted to only biosynthesise a few amino acids *de novo*. This is surprising given their importance in metabolism and the fact that many eukaryotes including humans can biosynthesise many more than the predicted capacity of *I. marina*. Equally, given the type of analysis performed to achieve these results, it may be the case that some of the pathways are present in reality, but homologues could not be identified in this search. These issues are discussed in greater detail in Section 3.3. Interestingly, many of the bacterial species were predicted to only biosynthesise a few amino acids too, suggesting scavenging from the community pool or direct exchange of metabolites. Across the community however the full complement of amino acids and B vitamins were predicted to be biosynthesised by at least one member. This societal approach to metabolite production corroborates the findings of Zelezniak et al. (2015) who determined that metabolite exchange promotes the survival of co-occurring species especially in communities experiencing nutrient stress. These types of interaction warrant further exploration.

4.4.1 Effect of the Coronavirus pandemic / future work

A consequence of the pandemic was severe restriction of access to the laboratory in 2020 and partially in 2021. This shifted the focus of some of the planned work away from being wet-lab orientated and more towards dry-lab, resulting in the metagenomic analysis presented above. However, the work that was not performed was a suite of growth experiments exploring the physiology of *I. marina* and its bacterial comrades. Prior to the pandemic I was in the early stages of developing a process to quantify the growth of the culture. This is not a simple task as the eukaryotic cells predominantly grow benthically on the surface of the growing vessel and only occasionally are found in the liquid-column. They are also very small in size ~2 µm in diameter (cell body) which makes them difficult to image and count with standard optical methods. The most promising avenue appeared to be using an image processing software to count cells automatically.

Other areas that would have been explored and would constitute future work centres around the medium on which this community is grown. The standard is the CCAP ASWP with the addition of a cereal grain (boiled). ASWP includes soil extract which makes the medium complex. Work should be done to develop a standard defined media on which this culture could be grown and subsequently more intricate experiments could be conducted. For example, varying the availability of exogenous micronutrients such as the B vitamins and assessing the effect on the growth of the culture as a whole and of *I. marina*. Ultimately, it would be best to have *I. marina* growing in culture both xenically and axenically on a defined medium. This would provide the greatest level of

manipulation of provided metabolites to explore the metabolic functions of the organism. This would allow experimental validation of the genomic results presented in this chapter.

As described above, initial attempts to isolate bacteria from the culture only yielded four different species. Further work could be done in attempting to isolate more of the bacteria species from the culture which computational evidence indicates are present. This could be done by experimenting with culturing methods (media changes) as well as plating density. Another interesting experiment would be to attempt to identify at what ratios these bacteria exist in the community and how this influences the community-driven metabolism. For example, the *Marinomonas* species (bin 27) is predicted to be the only species capable of biosynthesising the aromatic amino acids tyrosine and phenylalanine. Does this mean it exists at a higher level of abundance in comparison to other members of the community, in order to provide adequate amounts of these AAs?

One bioinformatic analysis that could build on the analysis performed here would be synteny analysis. Comparing the gene sets of a range of stramenopiles to identify how similar their genomic make up is would be extremely interesting and provide greater insight into the evolution of the group. In conclusion, although this work does achieve the aim of advancing our knowledge and increasing the available data for this enigmatic organism, it also opens up a wide range of possible research avenues that would further our understanding of the evolution of this organism and the stramenopile group as a whole.

5 – Summary and Discussion

There has been a historical bias in biological research on a small number of taxonomic groups. This bias is anthropogenic where, in general, research has been focused on charismatic megafauna, plants or humans. In the field of microbiology, the bias is towards photosynthetic species, those with important industrial or agricultural applications or species with implications on human health. As a result, a plethora of species have remained understudied.

In this thesis, a concerted effort was made to begin to address the bias by analysing data from organisms outside of this traditional research focus. This was done with the specific research angle of B vitamin metabolism and looked at one group of the eukaryotic tree of life in particular, the Stramenopila. Three experimental threads were followed in order to probe the question: How do B vitamins contribute to marine stramenopile communities?

5.1 Naturally occurring communities

The first of these approaches was environmental sampling, where vitamin amendments to naturally occurring microbial communities were carried out to assess the changes in community structure. In this experiment there was a particular focus on B vitamins that had historically been excluded from research in algal microbiology – B_2, B_3, B_5, B_6 and B_9 – to assess if these were limiting the growth of any members of the microbial community.

The intention for this part of the project was to run multiple repeats of this experiment with iterative improvements to the experimental design. Unfortunately, as a result of the COVID-19 pandemic, only the pilot experiment was completed. Despite this, the results showed that changes in the microbial community did occur upon the addition of these vitamins. Most notably changes in transcriptional activity were observed. Transcriptional activity was measured by calculating the ratio between the number of NGS reads for rRNA and those for rDNA for a given taxa. This gives a proxy measurement for growth and indicated that changes in vitamin availability had an effect on this community. However, as discussed in Chapter 2 (Section 2.4), the effect on transcriptional activity in this experiment was small and no changes were seen in the abundances of the taxa observed. This was most likely due to the short incubation period (24 h) used in the experiment. Stramenopiles were amongst the most abundant taxa in the experiment. However, the majority of these reads came from diatom species. MAST species were also detected which indicates that this approach can be used to assess MAST abundances or activities. This methodology could be developed in the future to study MASTs more closely, for example, by correlating abundance

measurements with the closely recorded environmental data that is taken at the sampling site E1, or indeed assessing co-occurrences of MASTs with bacteria of interest, such as those identified to live in close association with *I. marina* (Chapter 4).

As well as the limitations with the setup of the experiment previously discussed in Chapter 2, the experiment was also limited by the sequencing methodology employed. The primers used did not include specific primers for fungal rRNAs nor archaeal rRNA. Although the focus of the experiment was for the bacteria and protists in the community, the former two groups also play significant roles in the oceanic microbiome and may also be involved in the B vitamin cycling. Omitting them from the analysis results in an incomplete picture of the community. There is also the issue of incomplete capture of diversity in the two groups that were sequenced for (bacteria and eukaryotes). This occurred as a result of the high variability in the variable regions of the rRNA genes. Tragin et al. (2018) compared the richness of species identified by sequencing the V4 18S rRNA region versus the V9 region. They found that of all the species detected in the study, 22% were identified only by the V4 marker, 44% only by the V9 marker and the remaining 34% were detected by both. To overcome this issue one of two things could be done. Firstly, one could sequence both the V9 and V4 regions for full coverage of species. Alternatively, one could use long-read technology to sequence the entire length of the 18S rRNA gene, which would capture all of the variability.

5.2 B vitamin biosynthesis in stramenopiles

The second strand of work (Chapter 3) was analysis of gene complements in a large number of stramenopile species to identify patterns in the ability or inability to perform *de novo* biosynthesis of B vitamins. The result demonstrated that the metabolic functions of the stramenopiles is as diverse as the morphological features displayed across the group. At the scale of the whole group, evolutionary patterns emerged that indicated gene loss and gain events that have led to the acquisition or loss of different B vitamin biosynthesis pathways. However, more detailed analysis of individual species showed that despite these global patterns in the group, auxotrophy status for any given B vitamin could be different for one species to the next, even within genera. For example, in the in-depth analysis that was carried out, it was predicted that although *P. sojae* was unlikely to biosynthesise riboflavin *de novo, P. cinnamomi var cinnamomi* was able to and in general the large-scale analysis indicated that oomycetes do have the ability to biosynthesise B_2 (Figure 3.7a). Similarly, *I. marina* is predicted to lack the ability to biosynthesise pyridoxal despite this appearing as a commonly produced B vitamin in most other species analysed.

This species specificity highlights the difficulty posed by attempting to culture novel species. However, these results could be used to inform future culturing efforts. A workflow could be envisaged whereby:

- An environmental sample is collected, and a novel species of interest is identified by microscopy and/or flow cytometry for single cell sorting.
- Cells are grown on a complex media such as ASWP and DNA is extracted for PCR of the 18S rRNA gene. Alternatively, single-cell DNA amplification techniques are used directly on isolates.
- The 18S sequence is used to confirm a likely taxonomic position in the eukaryotic tree. The sequence turns out to be a novel stramenopile.
- The data presented in this section informs the user that this novel species is likely to be dependent on an exogenous source of B_7 and B_{12} because it belongs to the Chrysophyceae and this is commonly the case in this group.

A logical extension to the work presented here would be to look for B vitamin uptake and transport proteins. The cobalamin uptake protein CBA1 was identified in the diatoms *T. pseudonana* and *P. tricornutum* by differential expression in varying B_{12} concentrations (Bertrand *et al.,* 2012). This protein was 160-fold more abundant in B_{12} deplete conditions. Analysing the data set used in this section to identify homologues of CBA1 would provide further information on the strategies of B_{12} auxotrophic organisms. One would expect homologues to be identified in all species, but in particular in those with only the METH isoform of methionine synthase. Indeed, homologues for CBA1 have been identified in many photosynthetic stramenopiles (and a handful of Labyrinthulae) (Sayer, 2019) but not so far in basal lineages of the group.

The following proteins are reported in the literature for transport of the other B vitamins.

Thiamine: In mammals, there are two thiamine transporter proteins ThTr1 and ThTr2 (Landowski *et al.,* 2008). In *Salmonella typhimurium* there is a thiamine uptake system that involves a transmembrane binding protein ThiQ (Webb *et al.,* 1998). Similar proteins to this transmembrane protein have been identified by sequence similarity in stramenopile species but have not been functionally validated (McRose *et al.,* 2014).

Riboflavin: The transport system for riboflavin in microorganisms remains relatively elusive even in bacteria. The gene ribJ has been identified in two species of Trypanosoma (Balcazar *et al.,* 2017).

Metazoans, which are all B_2 auxotrophic, satisfy their requirement for the vitamin via their diet and it is taken up via the RFVT/SLC52 transporter system (Yonezawa and Inui, 2013).

Niacin: Jeanguenin et al. (2012) identified and characterised nicotinate transporters, responsible for niacin, B_3, uptake, in mammals and plants. They identified these from the ubiquitous transporter family NiaP. Bacterial NiaP proteins are also involved in nicotinate uptake in some species, whilst in others they are involved in thiamine uptake (Jeanguenin *et al.,* 2012). Given that NiaP proteins have been identified that are involved in the uptake of B_3 in bacteria, plants and animals, it would be reasonable to look to this family of transporters as the niacin transporters in stramenopiles too.

Pyridoxal: The first characterised transporter for B_6 was identified in *Saccharomyces cerevisiae* and is called TPN1 (Stolz and Vielreicher, 2003). It has also been proposed that the thiamine transporters ThTr1 and ThTr2 may also transport B6 across membranes (Yamashiro *et al.,* 2020).

Biotin: Biotin transporters are also poorly characterised across the eukaryotic tree. In mammalian cells the protein SMVT1 is used (Prasad *et al.,* 1998), in yeast the protein is named VHT1 and in bacteria two proteins of the ECF complex are required (Azhar *et al.,* 2015; Hebbeln *et al.,* 2007).

Folate: In humans PCFT is a proton coupled transporter of folate (Qiu *et al.,* 2006). de Crécy-Lagard et al. (2007) identified folate transporters in cyanobacteria and Arabidopsis (*Synechocystis slr0642* and At2g32040 respectively). However, they also demonstrate that the Arabidopsis protein is localised to the chloroplast which could indicate it is a phototroph specific mechanism (de Crécy-Lagard *et al.,* 2007).

Despite the relative scarcity of data on B vitamin transporters, particularly in specific groups such as the stramenopiles, it remains an intriguing line of enquiry. For example, for species that have independently developed auxotrophy (where the majority of their relatives are prototrophs), how have they evolved to supplement metabolism sufficiently where this was previously unnecessary (as they presumably were also prototrophic)?

5.3 Model stramenopile-bacteria community

Lastly, the community of *I. marina* and its bacterial co-isolates was analysed for evidence of interaction between species, in particular with regards to B vitamin exchange (Chapter 4). The analysis revealed a community-based approach to metabolism where no individual synthesises the full complement of B vitamins. Instead, these micronutrients are synthesised by certain members

of the community and shared amongst the other members. This is a phenomenon that has been gathering evidence, as the wealth of metagenomic studies of microbial communities increase in number. Although perhaps surprising that many of the species in this consortium have so few complete biosynthesis pathways for the B vitamins and amino acids, it has been reported previously in other communities, demonstrating the commonality of this type of shared metabolism (Anantharaman *et al.,* 2016; Brown *et al.,* 2015; Hug and Co, 2018).

The data presented here however only represents the metabolic potential of the community. What it cannot tell us is which of these metabolic pathways are being expressed or are functionally active (with the exception of the complementary *I. marina* transcriptomic data from Thakur et al. 2019). For a more complete understanding of what metabolites are being produced and to what extent the biosynthesis pathways are being expressed, proteomic, metabolomic and transcriptomic approaches would need to be explored (Kazamia *et al.,* 2016). A combined metatranscriptomic and genomic approach was used by Vorobev et al. (2020) to study marine protist communities, which demonstrated the ability of these techniques to identify environmental traits for different species such as contribution to DMSP metabolism and interactions, for example in haptophyte-cyanobacteria symbiosis. To test the true amino acid and B vitamin requirements of the bacterial members of the community, an assay could be conducted in which samples from the main culture are added to cultures containing defined basic media with the addition of varying combinations of B vitamins and amino acids. Subsequent identification of the species that grew in each growth condition would corroborate or refute the results in this section.

A further proviso of the work presented here is that it only represents this metabolic potential at a single time point (at the time of sequencing). Microbial communities in culture are highly sensitive to abiotic factors, which may affect the interactions observed (Burman and Bengtsson-Palme, 2021). This raises the question of how representative the community is of the naturally occurring assemblage in which these organisms reside. Since the isolation of this community, which members have since been lost? How has the metabolism of these organisms changed as a result of continual sub-culturing in ASWP media for 11 years? How have the relative abundances of each species in the community changed over this time period and does this make a difference to the community metabolism? Are all members of this culture necessary for a functioning community? Are some of the members redundant but just happened to be picked up in the initial environmental sample and have persisted? Given the time period over which this community has been co-

cultured, how much co-evolution has occurred and as a result to what extent are the interaction in this community a result of that co-evolution?

5.4 General discussions and future work

5.4.1 Potential future analysis concerning the diversity of the stramenopiles

Further group wide comparative analysis would also be interesting to carry out, for example, to investigate the polyploidization of stramenopile genomes across the group, as discussed in Section 4.4. Another trait that is equally inconsistent across the group is the presence or absence of flagella and the number of flagella a cell possesses. As previously stated, the group was founded, prior to the advent of genomic classification, on the basis that cells had two flagella of unequal length. This is certainly true for many members of the group, including members belonging to the basal lineages. For example, *P. vesiculosus* (MAST6) is bi-flagellate (Shiratori *et al.,* 2017) as is *C. ronebergensis* (Hackl *et al.,* 2020). Indeed, *P. tardus*, which is sister to all extant stramenopiles (Thakur *et al.,* 2019), also has two flagella (Shiratori *et al.,* 2015). However, as demonstrated in Section 4.3.1.2 (Figures 4.4-5) and previously reported by Cavalier-Smith and Scoble (2013), *I. marina* only possesses a single flagellum and diatoms are known not to have any flagella (Medlin *et al.,* 1996). Within other Ochrophyta groups the raphidophyte alga *Heterosigma akashiwo* possesses two flagella (Hara and Chihara, 1987). Many chrysophytes such as *Ochromonas* species have two flagella (Andersen, 2011). However, *Chromulina* species only have a single flagellum (Cambra-Sánchez, 2010). Tracking this trait in relation to the phylogenomic tree of the group might provide insight to the evolutionary divergence of this trait. A time correlated tree would add a time estimation to the loss and/or gain events of the stramenopile flagella.

5.4.2 Thoughts on B vitamin auxotrophy

It is evident from the results in Chapters 3 and 4 that auxotrophy for B vitamins is more common in the heterotrophic lineages of the stramenopiles. This is perhaps not surprising given the propensity to lose biosynthetic pathways for molecules abundant in the diet of an organism (e.g., the loss of vitamin C biosynthesis in select mammal species (Drouin *et al.,* 2011)). However, if vitamin availability in the diet of heterotrophs is a driving force for pathway deterioration and subsequent loss of biosynthesis, why would this have occurred for some of the B vitamins but not others? It would be of interest to investigate the relative flux through metabolic pathways dependent on a given B vitamin in an auxotroph and a prototroph. Given the potential limiting nature of auxotrophy (i.e., there could be a limited source of exogenous B vitamin) does the auxotrophic organism use those pathways that are dependent on that B vitamin less than its prototrophic

counterpart, which would be expected never to be limited due to its ability to biosynthesise the vitamin *de novo*? It would seem unlikely given the central role B vitamins play in metabolism. Instead, one would expect that auxotrophy would only develop in species where the exogenous source is plentiful. Indeed, given the importance of B vitamins to metabolism, it may be the case that the auxotrophic organisms have developed a system by which they are never limited by the B vitamin they are auxotrophic for. This could be achieved by a "farming" method by which the organism uses signalling or exchange molecules to cultivate a community which always provides the B vitamin at a sustainable concentration (Kazamia *et al.,* 2016).

5.4.3 Interactions between microbes

The analysis conducted on the TARA oceans interactome by Lima-Mendez et al. (2015) concluded that 72% of the detected associations were positive. These were predominantly between protists, but also included some interactions between protists and bacteria, which have been evidenced experimentally. However, Palmer and Foster (2022) argue that for bacteria, interactions are seldom positive and that mutualistic interactions where both species benefit from the interaction is rare. Kazamia et al. (2016) discuss the mechanisms to interactions between phytoplankton and bacteria, such as the foraging to farming hypothesis. This hypothesis is a possible explanation for the evolution of B_{12} auxotrophy in algae and the development of an interaction between the alga and B_{12} producing bacteria. That review also discusses the Black Queen Hypothesis, which is concerned with the development of stable bacterial communities. In order to develop hypotheses of this type, much more detail of the parties involved is required than is presented for the community in this thesis (Chapter 4). In order for this to be achieved, complementary approaches would be required such as proteomics and transcriptomics. These techniques would allow the unveiling of the dynamics of the interactions in the community. For the community to maintain itself with the number of species present (estimated at 24) at least some of the interactions must be mutualistic, or at least the positive and negative interactions must be in equilibrium.

5.4.4 Thoughts on bioinformatics for lesser studied taxa

Bioinformatic analyses in poorly characterised species, such as those presented in this thesis, are currently faced with a number of challenges. The most significant challenge is that many bioinformatic tools, in particular the KEGG ecosystem, are based on databases populated with data from well characterised species. This is a circular argument, as the data from previously unstudied species cannot be in a database before it is produced. Nonetheless, it does make the interpretation of the results from these kinds of analysis difficult, especially if the organism being analysed is

particularly distant from the species that are included in the databases. Exciting new initiatives such as The Earth BioGenome Project (EBP) may help. The EBP aims to sequence the genome of every extant species to reference quality, and will therefore help to address the imbalances in the availability of sequencing data for understudied branches of the tree of life (Scott-Somme *et al.,* 2022). Initiatives such as the Environmental Model Systems project (Faktorová *et al.,* 2020), which aims to develop genetic tools for a range of protists, represent concerted efforts to develop new genetic tools for emerging model systems. These projects working in conjunction with each other demonstrates how this field can advance in the years to come and how studying previously understudied taxa will become easier in the future.

Two sets of analyses would assist in the advancement of the results presented in this thesis. Firstly, analysis of the putative B vitamin biosynthesis pathway provides evidence of HGT. Recent publications have demonstrated just how widespread HGT has been in eukaryotes both from bacteria and other eukaryotes (Eme *et al.,* 2017; Fan *et al.,* 2020) including within stramenopiles (Dorrell *et al.,* 2021). These findings demonstrate how essential HGT is for developing novel traits in eukaryotes. Further detailed analysis of the sequences detected for B vitamin pathways in stramenopiles presented here could determine whether acquisition of pathways occurred as a result of HGT as previously proposed in some cases (e.g. for the niacin dichotomy (Ternes and Schönknecht, 2014)). This may be a mechanism that allows the shift for a species to go from an auxotroph to a prototroph, when all other close relatives remain auxotrophic.

Secondly, interrogation of sequences for evidence of pseudogenes would aid in the understanding of the evolution of these biosynthesis pathways. Although sequence comparisons can help in determining whether a putative homologue is a pseudogene, confirmation would take a great deal of physiological and genetic characterisation of individual pathway enzymes for each B vitamin pathway in many species to serve as a database of confirmed sequences. This would be required especially for those basal lineages of the stramenopiles who have little representation in current databases.

5.5 Concluding remarks

The work presented in this thesis demonstrates, through a number of lines of evidence, that B vitamins are important in the functioning of microbial communities. The experiment detailed in Chapter 2 shows that those B vitamins traditionally excluded from marine microbial research may have a role in the dynamics of the communities of the English Channel. The analysis presented in

Chapter 3 demonstrates the prevalence of B vitamin auxotrophy as well as the species-specific nature of this trait through examining a large library of sequencing data for diverse marine microbial eukaryotes. Finally, the analysis of a model-like marine microbial community (Chapter 4), isolated with one of the only culturable MAST species, shows that different members of the community are likely to rely on other members to satisfy the requirements of their metabolism. The mechanism by which this happens is not clear, however. This work adds to the growing appreciation for the importance of B vitamins in microbial communities, could inform future culturing efforts, and demonstrates that further research into the understudied stramenopiles and B vitamins is warranted.

6 - References

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

Albertin, W., Marullo, P., 2012. Polyploidy in fungi: evolution after whole-genome duplication. Proc Biol Sci 279, 2497–2509. https://doi.org/10.1098/rspb.2012.0434

Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D.S., Vardi, A., 2015. Identification of the algal dimethyl sulfide–releasing enzyme: A missing link in the marine sulfur cycle. Science 348, 1466–1469. https://doi.org/10.1126/science.aab1586

Allen, A.E., Dupont, C.L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J.P., Zheng, H., Johnson, D.A., Hu, H., Fernie, A.R., Bowler, C., 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473, 203–207. https://doi.org/10.1038/nature10074

Allen, R., Hoffmann, L.J., Law, C.S., Summerfield, T.C., 2020. Subtle bacterioplankton community responses to elevated CO2 and warming in the oligotrophic South Pacific gyre. Environmental Microbiology Reports 12, 377–386. https://doi.org/10.1111/1758-2229.12844

Altemose, N., Logsdon Glennis A., Bzikadze Andrey V., Sidhwani Pragya, Langley Sasha A., Caldas Gina V., Hoyt Savannah J., Uralsky Lev, Ryabov Fedor D., Shew Colin J., Sauria Michael E. G., Borchers Matthew, Gershman Ariel, Mikheenko Alla, Shepelev Valery A., Dvorkina Tatiana, Kunyavskaya Olga, Vollger Mitchell R., Rhie Arang, McCartney Ann M., Asri Mobin, Lorig-Roach Ryan, Shafin Kishwar, Lucas Julian K., Aganezov Sergey, Olson Daniel, de Lima Leonardo Gomes, Potapova Tamara, Hartley Gabrielle A., Haukness Marina, Kerpedjiev Peter, Gusev Fedor, Tigyi Kristof, Brooks Shelise, Young Alice, Nurk Sergey, Koren Sergey, Salama Sofie R., Paten Benedict, Rogaev Evgeny I., Streets Aaron, Karpen Gary H., Dernburg Abby F., Sullivan Beth A., Straight Aaron F., Wheeler Travis J., Gerton Jennifer L., Eichler Evan E., Phillippy Adam M., Timp Winston, Dennis Megan Y., O'Neill Rachel J., Zook Justin M., Schatz Michael C., Pevzner Pavel A., Diekhans Mark, Langley Charles H., Alexandrov Ivan A., Miga Karen H., 2022. Complete genomic and epigenetic maps of human centromeres. Science 376, eabl4178. https://doi.org/10.1126/science.abl4178

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389–402.

Amin, S.A., Green, D.H., Hart, M.C., Küpper, F.C., Sunda, W.G., Carrano, C.J., 2009. Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. Proceedings of the National

Academy of Sciences of the United States of America 106, 17071–6. https://doi.org/10.1073/pnas.0905512106

Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal, K.R., Morales, R.L., Berthiaume, C.T., Parker, M.S., Djunaedi, B., Ingalls, A.E., Parsek, M.R., Moran, M.A., Armbrust, E.V., 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature 522, 98–101. https://doi.org/10.1038/nature14488

Amin, S.A., Parker, M.S., Armbrust, E.V., 2012. Interactions between Diatoms and Bacteria. Microbiol Mol Biol Rev 76, 667–684. https://doi.org/10.1128/MMBR.00007-12

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., Brodie, E.L., Williams, K.H., Hubbard, S.S., Banfield, J.F., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun 7, 13219–13219. https://doi.org/10.1038/ncomms13219

Andersen, R.A., 2011. Ochromonas moestrupii sp. nov. (Chrysophyceae), a new golden flagellate from Australia. null 50, 600–607. https://doi.org/10.2216/11-13.1

Archibald, J.M., 2009. The puzzle of plastid evolution. Curr Biol 19, R81-88. https://doi.org/10.1016/j.cub.2008.11.067

Armbrust, E.V., 2009. The life of diatoms in the world's oceans. Nature 459, 185–192. https://doi.org/10.1038/nature08057

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., Bechner, M., Brzezinski, M.A., Chaal, B.K., Chiovitti, A., Davis, A.K., Demarest, M.S., Detter, J.C., Glavina, T., Goodstein, D., Hadi, M.Z., Hellsten, U., Hildebrand, M., Jenkins, B.D., Jurka, J., Kapitonov, V.V., Kröger, N., Lau, W.W.Y., Lane, T.W., Larimer, F.W., Lippmeier, J.C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M.S., Palenik, B., Pazour, G.J., Richardson, P.M., Rynearson, T.A., Saito, M.A., Schwartz, D.C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F.P., Rokhsar, D.S., 2004. The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science 306, 79–86. https://doi.org/10.1126/science.1101156

Arrigo, K.R., 2005. Marine microorganisms and global nutrient cycles. Nature 437, 349–355. https://doi.org/10.1038/nature04159

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

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., Thingstad, F., 1983. The Ecological Role of Water-Column Microbes in the Sea. Marine Ecology Progress Series 10, 257–263.

Azhar, A., Booker, G.W., Polyak, S.W., 2015. Mechanisms of Biotin Transport. Biochem Anal Biochem 04. https://doi.org/10.4172/2161-1009.1000210

Balcazar, D.E., Vanrell, M.C., Romano, P.S., Pereira, C.A., Goldbaum, F.A., Bonomi, H.R., Carrillo, C., 2017. The superfamily keeps growing: Identification in trypanosomatids of RibJ, the first riboflavin transporter family in protists. PLoS Negl Trop Dis 11, e0005513. https://doi.org/10.1371/journal.pntd.0005513

Bar-On, Y.M., Milo, R., 2019. The Biomass Composition of the Oceans: A Blueprint of Our Blue Planet. Cell 179, 1451–1454. https://doi.org/10.1016/j.cell.2019.11.018

Bar-On, Y.M., Phillips, R., Milo, R., 2018. The biomass distribution on Earth. PNAS 115, 6506–6511. https://doi.org/10.1073/pnas.1711842115

Beardall, J., 1988. Algal Cultures and Phytoplankton Ecology. null 27, 180–181. https://doi.org/10.2216/i0031-8884-27-1-180.1

Bertrand, E. M., Allen, A.E., Dupont, C.L., Norden-Krichmar, T.M., Bai, J., Valas, R.E., Saito, M.A., 2012. Influence of cobalamin scarcity on diatom molecular physiology and identification of a cobalamin acquisition protein. Proceedings of the National Academy of Sciences 109, E1762–E1771. https://doi.org/10.1073/pnas.1201731109

Bertrand, Erin M., Allen, A.E., Dupont, C.L., Norden-Krichmar, T.M., Bai, J., Valas, R.E., Saito, M.A., 2012. Influence of cobalamin scarcity on diatom molecular physiology and identification of a cobalamin acquisition protein. PNAS 109, E1762–E1771. https://doi.org/10.1073/pnas.1201731109

Bertrand, E.M., McCrow, J.P., Moustafa, A., Zheng, H., McQuaid, J.B., Delmont, T.O., Post, A.F., Sipler, R.E., Spackeen, J.L., Xu, K., Bronk, D.A., Hutchins, D.A., Allen, A.E., 2015. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. Proceedings of the National Academy of Sciences of the United States of America 112, 9938–43. https://doi.org/10.1073/pnas.1501615112

Bertrand, E.M., Saito, M.A., Lee, P.A., Dunbar, R.B., Sedwick, P.N., Ditullio, G.R., 2011. Iron limitation of a springtime bacterial and phytoplankton community in the ross sea: implications for vitamin b(12) nutrition. Front Microbiol 2, 160. https://doi.org/10.3389/fmicb.2011.00160

Bertrand, E.M., Saito, M.A., Rose, J.M., Riesselman, C.R., Lohan, M.C., Noble, A.E., Lee, P.A., DiTullio, G.R., 2007. Vitamin B12 and iron colimitation of phytoplankton growth in the Ross Sea. Limnology and Oceanography 52, 1079–1093. https://doi.org/10.4319/lo.2007.52.3.1079

Blazewicz, S.J., Barnard, R.L., Daly, R.A., Firestone, M.K., 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. The ISME Journal 7, 2061–2068. https://doi.org/10.1038/ismej.2013.102

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

Bonfante, P., Genre, A., 2010. Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. Nature Communications 1, 48. https://doi.org/10.1038/ncomms1046

Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E., Wincker, P., 2015. Tara Oceans. Tara Oceans studies plankton at planetary scale. Introduction. Science (New York, N.Y.) 348, 873–873. https://doi.org/10.1126/science.aac5605

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R.P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J.A., Brownlee, C., Cadoret, J.-P., Chiovitti, A., Choi, C.J., Coesel, S., De Martino, A., Detter, J.C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M.J.J., Jenkins, B.D., Jiroutova, K., Jorgensen, R.E., Joubert, Y., Kaplan, A., Kröger, N., Kroth, P.G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P.J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L.K., Montsant, A., Oudot-Le Secq, M.-P., Napoli, C., Obornik, M., Parker, M.S., Petit, J.-L., Porcel, B.M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T.A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M.R., Taylor, A.R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L.S., Rokhsar, D.S., Weissenbach, J., Armbrust, E.V., Green, B.R., Van de Peer, Y., Grigoriev, I.V., 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature 456, 239–244. https://doi.org/10.1038/nature07410

Bowsher, A.W., Kearns, P.J., Shade, A., 2019. 16S rRNA/rRNA Gene Ratios and Cell Activity Staining Reveal Consistent Patterns of Microbial Activity in Plant-Associated Soil. mSystems. https://doi.org/10.1128/mSystems.00003-19

Bray, J.R., Curtis, J.T., 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs 27, 325–349. https://doi.org/10.2307/1942268

Bristow, L.A., Mohr, W., Ahmerkamp, S., Kuypers, M.M.M., 2017b. Nutrients that limit growth in the ocean. Curr Biol 27, R474–R478. https://doi.org/10.1016/j.cub.2017.03.030

Brown, C.L., Keenum, I.M., Dai, D., Zhang, L., Vikesland, P.J., Pruden, A., 2021. Critical evaluation of short, long, and hybrid assembly for contextual analysis of antibiotic resistance genes in complex environmental metagenomes. Scientific Reports 11, 3753. https://doi.org/10.1038/s41598-021-83081-8

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., Banfield, J.F., 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523, 208–211. https://doi.org/10.1038/nature14486

Browning, T.J., Achterberg, E.P., Rapp, I., Engel, A., Bertrand, E.M., Tagliabue, A., Moore, C.M., 2017. Nutrient co-limitation at the boundary of an oceanic gyre. Nature 551, 242–246. https://doi.org/10.1038/nature24063

Burki, F., 2017. Chapter One - The Convoluted Evolution of Eukaryotes With Complex Plastids, in: Hirakawa, Y. (Ed.), Advances in Botanical Research. Academic Press, pp. 1–30. https://doi.org/10.1016/bs.abr.2017.06.001

Burki, F., Keeling, P.J., 2014. Rhizaria. Current Biology 24, R103–R107. https://doi.org/10.1016/J.CUB.2013.12.025

Burki, F., Roger, A.J., Brown, M.W., Simpson, A.G.B., 2020. The New Tree of Eukaryotes. Trends Ecol Evol 35, 43–55. https://doi.org/10.1016/j.tree.2019.08.008

Burman, E., Bengtsson-Palme, J., 2021. Microbial Community Interactions Are Sensitive to Small Changes in Temperature. Frontiers in Microbiology 12.

Callahan, Benjamin J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869

Callahan, Ben J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P., 2016. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. https://doi.org/10.12688/f1000research.8986.2

Cambra-Sánchez, J., 2010. Chrysophytes from some lakes and peat-bogs in the eastern Pyrenees, Catalonia (Spain). Biologia 65, 577–586. https://doi.org/10.2478/s11756-010-0057-z

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38, 5825–5829. https://doi.org/10.1093/molbev/msab293

Catlett, D., Matson, P.G., Carlson, C.A., Wilbanks, E.G., Siegel, D.A., Iglesias-Rodriguez, M.D., 2020. Evaluation of accuracy and precision in an amplicon sequencing workflow for marine protist communities. Limnology and Oceanography: Methods 18, 20–40. https://doi.org/10.1002/lom3.10343

Cavalier-Smith, T., 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Eukaryot Microbiol 46, 347–366. https://doi.org/10.1111/j.1550-7408.1999.tb04614.x

Cavalier-Smith, T., Scoble, J.M., 2013. Phylogeny of Heterokonta: Incisomonas marina, a uniciliate gliding opalozoan related to Solenicola (Nanomonadea), and evidence that Actinophryida evolved from raphidophytes. European Journal of Protistology 49, 328–353. https://doi.org/10.1016/J.EJOP.2012.09.002

Chen, Z., Erickson, D.L., Meng, J., 2020. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. BMC Genomics 21, 631. https://doi.org/10.1186/s12864-020-07041-8

Chernomor, O., von Haeseler, A., Minh, B.Q., 2016. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Systematic Biology 65, 997–1008. https://doi.org/10.1093/sysbio/syw037

Cirri, E., Pohnert, G., 2019. Algae−bacteria interactions that balance the planktonic microbiome - Cirri - 2019 - New Phytologist - Wiley Online Library [WWW Document]. URL https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.15765 (accessed 4.13.22).

Cooper, G.M., 2000. The Origin and Evolution of Cells. The Cell: A Molecular Approach. 2nd edition.

Cooper, M.B., Kazamia, E., Helliwell, K.E., Kudahl, U.J., Sayer, A., Wheeler, G.L., Smith, A.G., 2018. Cross-exchange of B-vitamins underpins a mutualistic interaction between Ostreococcus tauri and Dinoroseobacter shibae. The ISME Journal 1–1. https://doi.org/10.1038/s41396-018-0274-y

Coxon, K.M., Chakauya, E., Ottenhof, H.H., Whitney, H.M., Blundell, T.L., Abell, C., Smith, A.G., 2005. Pantothenate biosynthesis in higher plants. Biochemical Society Transactions 33, 743–746. https://doi.org/10.1042/BST0330743

Croft, M.T., Lawrence, A.D., Raux-Deery, E., Warren, M.J., Smith, A.G., 2005. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. Nature 438, 90–93. https://doi.org/10.1038/nature04056

Croft, M.T., Moulin, M., Webb, M.E., Smith, A.G., 2007. Thiamine biosynthesis in algae is regulated by riboswitches. Proceedings of the National Academy of Sciences 104, 20770–20775. https://doi.org/10.1073/pnas.0705786105

Croft, M.T., Warren, M.J., Smith, A.G., 2006. Algae Need Their Vitamins. Eukaryotic Cell 5, 1175–1183. https://doi.org/10.1128/EC.00097-06

Cruz-López, R., Maske, H., 2016. The Vitamin B1 and B12 Required by the Marine Dinoflagellate Lingulodinium polyedrum Can be Provided by its Associated Bacterial Community in Culture. Frontiers in Microbiology 7.

Curson, A.R.J., Williams, B.T., Pinchbeck, B.J., Sims, L.P., Martínez, A.B., Rivera, P.P.L., Kumaresan, D., Mercadé, E., Spurgin, L.G., Carrión, O., Moxon, S., Cattolico, R.A., Kuzhiumparambil, U., Guagliardo, P., Clode, P.L., Raina, J.-B., Todd, J.D., 2018. DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. Nature Microbiology 3, 430–439. https://doi.org/10.1038/s41564-018-0119-5

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., Van Broeckhoven, C., 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34, 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

de Crécy-Lagard, V., El Yacoubi, B., de la Garza, R.D., Noiriel, A., Hanson, A.D., 2007. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. BMC Genomics 8, 245. https://doi.org/10.1186/1471-2164-8-245

Deppeler, S.L., Davidson, A.T., 2017. Southern Ocean Phytoplankton in a Changing Climate. Frontiers in Marine Science 4.

De vargas, C., Aubry, M.-P., Probert, I., Young, J., 2007. CHAPTER 12 - Origin and Evolution of Coccolithophores: From Coastal Hunters to Oceanic Farmers, in: Falkowski, P.G., Knoll, A.H. (Eds.), Evolution of Primary Producers in the Sea. Academic Press, Burlington, pp. 251–285. https://doi.org/10.1016/B978-012370518-1/50013-8

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G., Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans Coordinators, T.O., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., Karsenti, E., 2015. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science (New York, N.Y.) 348, 1261605–1261605. https://doi.org/10.1126/science.1261605

del Campo, J., Massana, R., 2011. Emerging Diversity within Chrysophytes, Choanoflagellates and Bicosoecids Based on Molecular Surveys. Protist 162, 435–448. https://doi.org/10.1016/J.PROTIS.2010.10.003

del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., Ruiz-Trillo, I., 2014. The others: our biased perspective of eukaryotic genomes. Trends in ecology & evolution 29, 252–9. https://doi.org/10.1016/j.tree.2014.03.006

Delmont, T.O., Gaia, M., Hinsinger, D.D., Frémont, P., Vanni, C., Fernandez-Guerra, A., Eren, A.M., Kourlaiev, A., d'Agata, L., Clayssen, Q., Villar, E., Labadie, K., Cruaud, C., Poulain, J., Da Silva, C., Wessner, M., Noel, B., Aury, J.-M., Sunagawa, S., Acinas, S.G., Bork, P., Karsenti, E., Bowler, C., Sardet, C., Stemmann, L., de Vargas, C., Wincker, P., Lescot, M., Babin, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp, P., Jaillon, O., Kandels, S., Iudicone, D., Ogata, H., Pesant, S., Sullivan, M.B., Not, F., Lee, K.-B., Boss, E., Cochrane, G., Follows, M., Poulton, N., Raes, J., Sieracki, M., Speich, S., de Vargas, C., Bowler, C., Karsenti, E., Pelletier, E., Wincker, P., Jaillon, O., 2022. Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. Cell Genomics 100123. https://doi.org/10.1016/j.xgen.2022.100123

Delmont, T.O., Gaia, M., Hinsinger, D.D., Fremont, P., Vanni, C., Guerra, A.F., Eren, A.M., Kourlaiev, A., d'Agata, L., Clayssen, Q., Villar, E., Labadie, K., Cruaud, C., Poulain, J., Silva, C.D., Wessner, M., Noel, B., Aury, J.-M., Coordinators, T.O., Vargas, C. de, Bowler, C., Karsenti, E., Pelletier, E., Wincker, P., Jaillon, O., 2021. Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. bioRxiv 2020.10.15.341214. https://doi.org/10.1101/2020.10.15.341214

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics 6, 361–375. https://doi.org/10.1038/nrg1603

Derelle, R., López-García, P., Timpano, H., Moreira, D., 2016. A Phylogenomic Framework to Study the Diversity and Evolution of Stramenopiles (=Heterokonts). Molecular Biology and Evolution 33, 2890–2898. https://doi.org/10.1093/molbev/msw168

Dorrell, R.G., Smith, A.G., 2011. Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. Eukaryotic cell 10, 856–68. https://doi.org/10.1128/EC.00326-10

Dorrell, R.G., Villain Adrien, Perez-Lamarque Benoît, Audren de Kerdrel Guillemette, McCallum Giselle, Watson Andrew K., Ait-Mohamed Ouardia, Alberti Adriana, Corre Erwann, Frischkorn Kyle R., Pierella Karlusich Juan J., Pelletier Eric, Morlon Hélène, Bowler Chris, Blanc Guillaume, 2021. Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. Proceedings of the National Academy of Sciences 118, e2009974118. https://doi.org/10.1073/pnas.2009974118

Drebes, G., Kühn, S.F., Gmelch, A., Schnepf, E., 1996. Cryothecomonas aestivalis sp. nov., a colourless nanoflagellate feeding on the marine centric diatomGuinardia delicatula (Cleve) Hasle. Helgolander Meeresunters 50, 497–515. https://doi.org/10.1007/BF02367163

Drouin, G., Godin, J.-R., Page, B., 2011. The Genetics of Vitamin C Loss in Vertebrates. Current Genomics 12, 371–378. https://doi.org/10.2174/138920211796429736

Durham, B.P., Dearth, S.P., Sharma, S., Amin, S.A., Smith, C.B., Campagna, S.R., Armbrust, E.V., Moran, M.A., 2017. Recognition cascade and metabolite transfer in a marine bacteria-phytoplankton model system. Environ Microbiol 19, 3500–3513. https://doi.org/10.1111/1462-2920.13834

Eddy, S.R., 2011. Accelerated Profile HMM Searches. PLOS Computational Biology 7, e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Eme, L., Gentekaki, E., Curtis, B., Archibald, J.M., Roger, A.J., 2017. Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite Blastocystis to the Gut. Current Biology 27, 807–820. https://doi.org/10.1016/j.cub.2017.02.003

Emmett, A.D., Luros, G.O., 1920. WATER-SOLUBLE VITAMINES: I. ARE THE ANTINEURITIC AND THE GROWTH-PROMOTING WATER-SOLUBLE B VITAMINES THE SAME? Journal of Biological Chemistry 43, 265–286. https://doi.org/10.1016/S0021-9258(18)86329-3

Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T.O., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3, e1319. https://doi.org/10.7717/peerj.1319

Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C., Trigodet, F., Watson, A.R., Esen, Ö.C., Moore, R.M., Clayssen, Q., Lee, M.D., Kivenson, V., Graham, E.D., Merrill, B.D., Karkman, A., Blankenberg, D., Eppley, J.M., Sjödin, A., Scott, J.J., Vázquez-Campos, X., McKay, L.J., McDaniel, E.A., Stevens, S.L.R., Anderson, R.E., Fuessel, J., Fernandez-Guerra, A., Maignien, L., Delmont, T.O., Willis, A.D., 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nature Microbiology 6, 3–6. https://doi.org/10.1038/s41564-020-00834-3

Faktorová, D., Nisbet, R.E.R., Fernández Robledo, J.A., Casacuberta, E., Sudek, L., Allen, A.E., Ares, M., Aresté, C., Balestreri, C., Barbrook, A.C., Beardslee, P., Bender, S., Booth, D.S., Bouget, F.-Y., Bowler, C., Breglia, S.A., Brownlee, C., Burger, G., Cerutti, H., Cesaroni, R., Chiurillo, M.A., Clemente, T., Coles, D.B., Collier, J.L., Cooney, E.C., Coyne, K., Docampo, R., Dupont, C.L., Edgcomb, V., Einarsson, E., Elustondo, P.A., Federici, F., Freire-Beneitez, V., Freyria, N.J., Fukuda, K., García, P.A., Girguis, P.R., Gomaa, F., Gornik, S.G., Guo, J., Hampl, V., Hanawa, Y.,

Haro-Contreras, E.R., Hehenberger, E., Highfield, A., Hirakawa, Y., Hopes, A., Howe, C.J., Hu, I., Ibañez, J., Irwin, N.A.T., Ishii, Y., Janowicz, N.E., Jones, A.C., Kachale, A., Fujimura-Kamada, K., Kaur, B., Kaye, J.Z., Kazana, E., Keeling, P.J., King, N., Klobutcher, L.A., Lander, N., Lassadi, I., Li, Z., Lin, S., Lozano, J.-C., Luan, F., Maruyama, S., Matute, T., Miceli, C., Minagawa, J., Moosburner, M., Najle, S.R., Nanjappa, D., Nimmo, I.C., Noble, L., Novák Vanclová, A.M.G., Nowacki, M., Nuñez, I., Pain, A., Piersanti, A., Pucciarelli, S., Pyrih, J., Rest, J.S., Rius, M., Robertson, D., Ruaud, A., Ruiz-Trillo, I., Sigg, M.A., Silver, P.A., Slamovits, C.H., Jason Smith, G., Sprecher, B.N., Stern, R., Swart, E.C., Tsaousis, A.D., Tsypin, L., Turkewitz, A., Turnšek, J., Valach, M., Vergé, V., von Dassow, P., von der Haar, T., Waller, R.F., Wang, L., Wen, X., Wheeler, G., Woods, A., Zhang, H., Mock, T., Worden, A.Z., Lukeš, J., 2020. Genetic tool development in marine protists: emerging model organisms for experimental cell biology. Nature Methods 17, 481–494. https://doi.org/10.1038/s41592-020-0796-x

Falkowski, P.G., De Vargas, C., 2004. Shotgun Sequencing in the Sea: A Blast from the Past?, New Series.

Falkowski, P. G., Fenchel, T., Delong, E.F., 2008. The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science 320, 1034–1039. https://doi.org/10.1126/science.1153213

Falkowski, Paul G, Fenchel, T., Delong, E.F., 2008. The microbial engines that drive Earth's biogeochemical cycles. Science (New York, N.Y.) 320, 1034–9. https://doi.org/10.1126/science.1153213

Falkowski, P.G., Knoll, A.H., 2007. CHAPTER 1 - An Introduction to Primary Producers in the Sea: Who They Are, What They Do, and When They Evolved, in: Falkowski, Paul G., Knoll, Andrew H. (Eds.), Evolution of Primary Producers in the Sea. Academic Press, Burlington, pp. 1–6. https://doi.org/10.1016/B978-012370518-1/50002-3

Fan, X., Qiu Huan, Han Wentao, Wang Yitao, Xu Dong, Zhang Xiaowen, Bhattacharya Debashish, Ye Naihao, 2020. Phytoplankton pangenome reveals extensive prokaryotic horizontal gene transfer of diverse functions. Science Advances 6, eaba0111. https://doi.org/10.1126/sciadv.aba0111

Fattal-Valevski, A., 2011. Thiamine (Vitamin B1). J Evid Based Complementary Altern Med 16, 12–20. https://doi.org/10.1177/1533210110392941

Fenchel, T., 2008. The microbial loop – 25 years later. Journal of Experimental Marine Biology and Ecology 366, 99–103. https://doi.org/10.1016/j.jembe.2008.07.013

Ferrario, C., Duranti, S., Milani, C., Mancabelli, L., Lugli, G.A., Turroni, F., Mangifesta, M., Viappiani, A., Ossiprandi, M.C., van Sinderen, D., Ventura, M., 2015. Exploring Amino Acid Auxotrophy in Bifidobacterium bifidum PRL2010. Frontiers in Microbiology 6.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., Falkowski, P., 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science 281, 237–240. https://doi.org/10.1126/SCIENCE.281.5374.237

Fischer, M.G., Allen, M.J., Wilson, W.H., Suttle, C.A., 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc Natl Acad Sci U S A 107, 19508–19513. https://doi.org/10.1073/pnas.1007615107

Fischer, M.G., Hackl, T., 2016. Host genome integration and giant virus-induced reactivation of the virophage mavirus. Nature 540, 288–291. https://doi.org/10.1038/nature20593

Flynn, K.J., Butler, I., 1986. Nitrogen sources for the growth of marine microalgae: role of dissolved free amino acids. Marine Ecology Progress Series 34, 281–304.

Fukasawa, Y., Ermini, L., Wang, H., Carty, K., Cheung, M.-S., 2020. LongQC: A Quality Control Tool for Third Generation Sequencing Long Read Data. G3 Genes|Genomes|Genetics 10, 1193–1196. https://doi.org/10.1534/g3.119.400864

Gaertner, F.H., Subbayya Shetty, A., 1977. Kynureninase-type enzymes and the evolution of the aerobic tryptophan-to-nicotinamide adenine dinucleotide pathway. Biochimica et Biophysica Acta (BBA) - Enzymology 482, 453–460. https://doi.org/10.1016/0005-2744(77)90259-5

Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D., Koonin, E.V., 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res 49, D274–D281. https://doi.org/10.1093/nar/gkaa1018

Geneious Prime 2019.1.3, n.d. Geneious Prime 2019.1.3.

Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P.J., Huse, S., Joint, I., 2009. The seasonal structure of microbial communities in the Western English Channel. Environ Microbiol 11, 3132–3139. https://doi.org/10.1111/j.1462-2920.2009.02017.x

Giovannoni, S.J., 2017. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. Ann Rev Mar Sci 9, 231–255. https://doi.org/10.1146/annurev-marine-010814-015934

Gobler CJ, Norman C, Panzeca C, Taylor GT, SaÃ⊠Â±udo-Wilhelmy SA, 2007. Effect of B-vitamins (B1, B12) and inorganic nutrients on algal bloom dynamics in a coastal ecosystem. Aquat Microb Ecol 49, 181–194.

Godoi, P.H.C., Galhardo, R.S., Luche, D.D., Van Sluys, M.-A., Menck, C.F.M., Oliva, G., 2006. Structure of the Thiazole Biosynthetic Enzyme THI1 from Arabidopsis thaliana*. Journal of Biological Chemistry 281, 30957–30966. https://doi.org/10.1074/jbc.M604469200

Goldstein, S., Belsky, M., 1963. B12 and B1 auxotrophy of lower marine phycomycetes. Archiv f�r Mikrobiologie 47, 161–163. https://doi.org/10.1007/BF00422521

Goss, E.M., Tabima Javier F., Cooke David E. L., Restrepo Silvia, Fry William E., Forbes Gregory A., Fieland Valerie J., Cardenas Martha, Grünwald Niklaus J., 2014. The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. Proceedings of the National Academy of Sciences 111, 8791–8796. https://doi.org/10.1073/pnas.1401884111

Gould, S.B., Waller, R.F., McFadden, G.I., 2008. Plastid evolution. Annu Rev Plant Biol 59, 491–517. https://doi.org/10.1146/annurev.arplant.59.032607.092915

Gower, J.C., 1967. Multivariate Analysis and Multidimensional Geometry. Journal of the Royal Statistical Society: Series D (The Statistician) 17, 13–28. https://doi.org/10.2307/2987199

Grant, M.A.A., Kazamia, E., Cicuta, P., Smith, A.G., 2014. Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures. ISME J 8, 1418–1427. https://doi.org/10.1038/ismej.2014.9

Grattepanche, J.-D., Walker, L.M., Ott, B.M., Paim Pinto, D.L., Delwiche, C.F., Lane, C.E., Katz, L.A., 2018. Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data. BioEssays 40, 1700198–1700198. https://doi.org/10.1002/bies.201700198

Graupner, N., Jensen, M., Bock, C., Marks, S., Rahmann, S., Beisser, D., Boenigk, J., 2018. Evolution of heterotrophy in chrysophytes as reflected by comparative transcriptomics. FEMS Microbiology Ecology 94. https://doi.org/10.1093/femsec/fiy039

Graziano LM, Geider RJ, Li WKW, Olaizola M, 1996. Nitrogen limitation of North Atlantic phytoplankton: analysis of physiological condition in nutrient enrichment experiments. Aquat Microb Ecol 11, 53–64.

Green, B.R., 2011. After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl c. Photosynthesis Research 107, 103–115. https://doi.org/10.1007/s11120-010-9584-2

Green, J.C., , Leadbeater, Barry S. C., Diver, W. L., Systematics Association., 1989. The Chromophyte algae : problems and perspectives. Published for the Systematics Association by Clarendon Press ; Oxford University Press, Oxford; New York.

Griffith, G.R., Chandler, J.L., Gholson, R.K., 1975. Studies on the de novo biosynthesis of NAD in Escherichia coli. The separation of the nadB gene product from the nadA gene product and its purification. Eur J Biochem 54, 239–245. https://doi.org/10.1111/j.1432-1033.1975.tb04133.x

Guillard, R.R., Ryther, J.H., 1962. Studies of marine planktonic diatoms. I. Cyclotella nana Hustedt, and Detonula confervacea (cleve) Gran. Can J Microbiol 8, 229–239. https://doi.org/10.1139/m62-029

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hackl, T., Martin, R., Barenhoff, K., Duponchel, S., Heider, D., Fischer, M.G., 2020. Four high-quality draft genome assemblies of the marine heterotrophic nanoflagellate Cafeteria roenbergensis. Scientific Data 7, 29. https://doi.org/10.1038/s41597-020-0363-4

Hall, C., Dietrich, F.S., 2007. The Reacquisition of Biotin Prototrophy in Saccharomyces cerevisiae Involved Horizontal Gene Transfer, Gene Duplication and Gene Clustering. Genetics 177, 2293–2307. https://doi.org/10.1534/genetics.107.074963

Hara, Y., Chihara, M., 1987. Morphology, ultrastructure and taxonomy of the raphidophycean algaHeterosigma akashiwo. The botanical magazine = Shokubutsu-gaku-zasshi 100, 151–163. https://doi.org/10.1007/BF02488320

Havaux, M., Ksas, B., Szewczyk, A., Rumeau, D., Franck, F., Caffarri, S., Triantaphylidès, C., 2009. Vitamin B6 deficient plants display increased sensitivity to high light and photo-oxidative stress. BMC plant biology 9, 130–130. https://doi.org/10.1186/1471-2229-9-130

Hebbeln, P., Rodionov Dmitry A., Alfandega Anja, Eitinger Thomas, 2007. Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. Proceedings of the National Academy of Sciences 104, 2909–2914. https://doi.org/10.1073/pnas.0609905104

Heidelberg, K.B., Gilbert, J.A., Joint, I., 2010. Marine genomics: at the interface of marine microbial ecology and biodiscovery. Microbial Biotechnology 3, 531–543. https://doi.org/10.1111/j.1751-7915.2010.00193.x

Helliwell, K.E., Wheeler, G.L., Leptos, K.C., Goldstein, R.E., Smith, A.G., 2011. Insights into the Evolution of Vitamin B12 Auxotrophy from Sequenced Algal Genomes. Molecular Biology and Evolution 28, 2921–2933. https://doi.org/10.1093/molbev/msr124

Helliwell, K.E., Wheeler, G.L., Smith, A.G., 2013. Widespread decay of vitamin-related pathways: coincidence or consequence? Trends in Genetics 29, 469–478. https://doi.org/10.1016/J.TIG.2013.03.003

Henson, M.W., Pitre, D.M., Weckhorst, J.L., Lanclos, V.C., Webber, A.T., Thrash, J.C., 2016. Artificial Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico. mSphere 1, e00028-16. https://doi.org/10.1128/mSphere.00028-16

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35, 518–522. https://doi.org/10.1093/molbev/msx281

Horowitz, N.H., 1945. On the Evolution of Biochemical Syntheses. Proc Natl Acad Sci U S A 31, 153–157.

Hubalek Valerie, Buck Moritz, Tan BoonFei, Foght Julia, Wendeberg Annelie, Berry David, Bertilsson Stefan, Eiler Alexander, Lloyd Karen G., 2017. Vitamin and Amino Acid Auxotrophy in Anaerobic Consortia Operating under Methanogenic Conditions. mSystems 2, e00038-17. https://doi.org/10.1128/mSystems.00038-17

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on

5090 organisms and 2502 viruses. Nucleic Acids Res 47, D309–D314. https://doi.org/10.1093/nar/gky1085

Hug, L.A., Co, R., 2018. It Takes a Village: Microbial Communities Thrive through Interactions and Metabolic Handoffs. mSystems 3, e00152-17. https://doi.org/10.1128/mSystems.00152-17

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

Imanian, B., Pombert, J.-F., Keeling, P.J., 2010. The Complete Plastid Genomes of the Two 'Dinotoms' Durinskia baltica and Kryptoperidinium foliaceum. PLOS ONE 5, e10711. https://doi.org/10.1371/journal.pone.0010711

Jamali, A.A., Akbari, F., Ghorakhlu, M.M., de la Guardia, M., Yari Khosroushahi, A., 2012. Applications of diatoms as potential microalgae in nanobiotechnology. Bioimpacts 2, 83–89. https://doi.org/10.5681/bi.2012.012

Jeanguenin, L., Lara-Núñez, A., Rodionov, D.A., Osterman, A.L., Komarova, N.Y., Rentsch, D., Gregory, J.F., Hanson, A.D., 2012. Comparative genomics and functional analysis of the NiaP family uncover nicotinate transporters from bacteria, plants, and mammals. Functional & Integrative Genomics 12, 25–34. https://doi.org/10.1007/s10142-011-0255-y

Johnson, L.K., Alexander, H., Brown, C.T., 2019. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. GigaScience 8. https://doi.org/10.1093/gigascience/giy158

Johnson, W.M., Kido Soule, M.C., Kujawinski, E.B., 2016. Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP. ISME J 10, 2304–2316. https://doi.org/10.1038/ismej.2016.6

Jolley, E.T., Jones, A.K., 1977. The interaction between Navicula muralis grunow and an associated species of Flavobacterium. null 12, 315–328. https://doi.org/10.1080/00071617700650341

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

Jurgenson, C.T., Begley, T.P., Ealick, S.E., 2009. The structural and biochemical foundations of thiamin biosynthesis. Annu Rev Biochem 78, 569–603. https://doi.org/10.1146/annurev.biochem.78.072407.102340

Kaartokallio Hermanni, Laamanen Maria, Sivonen Kaarina, 2005. Responses of Baltic Sea Ice and Open-Water Natural Bacterial Communities to Salinity Change. Applied and Environmental Microbiology 71, 4364–4371. https://doi.org/10.1128/AEM.71.8.4364-4371.2005

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45, D353–D361. https://doi.org/10.1093/nar/gkw1092

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016a. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44, D457-462. https://doi.org/10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., Morishima, K., 2016b. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428, 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C.J., Fagerberg, B., Nielsen, J., Bäckhed, F., 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature 498, 99–103. https://doi.org/10.1038/nature12198

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066. https://doi.org/10.1093/nar/gkf436

Kazamia, E., Helliwell, K.E., Purton, S., Smith, A.G., 2016. How mutualisms arise in phytoplankton communities: building eco-evolutionary principles for aquatic microbes. Ecology Letters 19, 810–822. https://doi.org/10.1111/ele.12615

Keeling, P.J., 2009. Chromalveolates and the evolution of plastids by secondary endosymbiosis. J Eukaryot Microbiol 56, 1–8. https://doi.org/10.1111/j.1550-7408.2008.00371.x

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biology 12, e1001889–e1001889. https://doi.org/10.1371/journal.pbio.1001889

Keeling, P.J., Campo, J. del, 2017. Marine Protists Are Not Just Big Bacteria. Current Biology 27, R541–R549. https://doi.org/10.1016/j.cub.2017.03.075

Khan, S.T., Nakagawa, Y., Harayama, S., 2008. Fulvibacter tottoriensis gen. nov., sp. nov., a member of the family Flavobacteriaceae isolated from marine sediment. Int J Syst Evol Microbiol 58, 1670–1674. https://doi.org/10.1099/ijs.0.65554-0

Klein, A.M., Bohannan, B.J.M., Jaffe, D.A., Levin, D.A., Green, J.L., 2016. Molecular Evidence for Metabolically Active Bacteria in the Atmosphere. Frontiers in Microbiology 7.

Koch, F., Hattenrath-Lehmann, T.K., Goleski, J.A., Sañudo-Wilhelmy, S., Fisher, N.S., Gobler, C.J., 2012. Vitamin b(1) and b(12) uptake and cycling by plankton communities in coastal ecosystems. Frontiers in microbiology 3, 363–363. https://doi.org/10.3389/fmicb.2012.00363

Koch, F., Marcoval, M.A., Panzeca, C., Bruland, K.W., Sañudo-Wilhelmy, S.A., Gobler, C.J., 2011. The effect of vitamin B 12 on phytoplankton growth and community structure in the Gulf of Alaska. Limnology and Oceanography 56, 1023–1034. https://doi.org/10.4319/lo.2011.56.3.1023

Koch, F., Trimborn, S., 2019. Limitation by Fe, Zn, Co, and B12 Results in Similar Physiological Responses in Two Antarctic Phytoplankton Species. Frontiers in Marine Science 6, 514. https://doi.org/10.3389/fmars.2019.00514

Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., Pevzner, P.A., 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nature Methods 17, 1103–1110. https://doi.org/10.1038/s41592-020-00971-x

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research 27, 722–736. https://doi.org/10.1101/gr.215087.116

Kudahl, U.J., 2018. A computational biology approach to studying algae-bacterial interactions (Thesis). University of Cambridge. https://doi.org/10.17863/CAM.24234

Lamour, K.H., Win, J., Kamoun, S., 2007. Oomycete genomics: new insights and future directions. FEMS Microbiology Letters 274, 1–8. https://doi.org/10.1111/j.1574-6968.2007.00786.x

Landowski, C.P., Suzuki, Y., Hediger, M.A., 2008. CHAPTER 4 - The Mammalian Transporter Families, in: ALPERN, R.J., HEBERT, S.C. (Eds.), Seldin and Giebisch's The Kidney (Fourth Edition). Academic Press, San Diego, pp. 91–146. https://doi.org/10.1016/B978-012088488-9.50007-3

Latijnhouwers, M., de Wit, P.J.G.M., Govers, F., 2003. Oomycetes and fungi: similar weaponry to attack plants. Trends in Microbiology 11, 462–469. https://doi.org/10.1016/J.TIM.2003.08.002

Latorre-Pérez, A., Villalba-Bermell, P., Pascual, J., Vilanova, C., 2020. Assembly methods for nanopore-based metagenomic sequencing: a comparative study. Scientific Reports 10, 13588. https://doi.org/10.1038/s41598-020-70491-3

Lavrinienko, A., Jernfors, T., Koskimäki, J.J., Pirttilä, A.M., Watts, P.C., 2021. Does Intraspecific Variation in rDNA Copy Number Affect Analysis of Microbial Communities? Trends in Microbiology 29, 19–27. https://doi.org/10.1016/j.tim.2020.05.019

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jørgensen, T., Brandslund, I., Nielsen, H.B., Juncker, A.S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E.G., Brunak, S., Clément, K., Doré, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W.M., Zucker, J.-D., Raes, J., Hansen, T., Guedon, E., Delorme, C., Layec, S., Khaci, G., van de Guchte, M., Vandemeulebrouck, G., Jamet, A., Dervyn, R., Sanchez, N., Maguin, E., Haimet, F., Winogradski, Y., Cultrone, A., Leclerc, M., Juste, C., Blottière, H., Pelletier, E., LePaslier, D., Artiguenave, F., Bruls, T., Weissenbach, J., Turner, K., Parkhill, J., Antolin, M., Manichanh, C., Casellas, F., Boruel, N., Varela, E., Torrejon, A., Guarner, F., Denariaz, G., Derrien, M., van Hylckama Vlieg, J.E.T., Veiga, P., Oozeer, R., Knol, J., Rescigno, M., Brechot, C., M'Rini, C., Mérieux, A., Yamada, T., Bork, P., Wang, J., Ehrlich, S.D., Pedersen, O., MetaHIT consortium, 2013. Richness of human gut microbiome correlates with metabolic markers. Nature 500, 541–546. https://doi.org/10.1038/nature12506

Lee, Z.M.-P., Bussema, C., III, Schmidt, T.M., 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Research 37, D489–D493. https://doi.org/10.1093/nar/gkn689

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lie, A.A.Y., Liu, Z., Terrado, R., Tatters, A.O., Heidelberg, K.B., Caron, D.A., 2018. A tale of two mixotrophic chrysophytes: Insights into the metabolisms of two Ochromonas species (Chrysophyceae) through a comparison of gene expression. PLOS ONE 13, e0192439–e0192439. https://doi.org/10.1371/journal.pone.0192439

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A.M., Coppola, L., Cornejo-Castillo, F.M., d'Ovidio, F., De Meester, L., Ferrera, I., Garet-Delmas, M.-J., Guidi, L., Lara, E., Pesant, S., Royo-Llonch, M., Salazar, G., Sánchez, P., Sebastian, M., Souffreau, C., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans coordinators, T.O., Gorsky, G., Not, F., Ogata, H., Speich, S., Stemmann, L., Weissenbach, J., Wincker, P., Acinas, S.G., Sunagawa, S., Bork, P., Sullivan, M.B., Karsenti, E., Bowler, C., de Vargas, C., Raes, J., 2015. Ocean plankton. Determinants of community structure in the global plankton interactome. Science (New York, N.Y.) 348, 1262073–1262073. https://doi.org/10.1126/science.1262073

Liu, P., Ding, H., Christensen, B.M., Li, J., 2012. Cysteine sulfinic acid decarboxylase activity of Aedes aegypti aspartate 1-decarboxylase: The structural basis of its substrate selectivity. Insect Biochemistry and Molecular Biology 42, 396–403. https://doi.org/10.1016/j.ibmb.2012.02.001

Lobley, C.M.C., Schmitzberger, F., Kilkenny, M.L., Whitney, H., Ottenhof, H.H., Chakauya, E., Webb, M.E., Birch, L.M., Tuck, K.L., Abell, C., Smith, A.G., Blundell, T.L., 2003. Structural insights

into the evolution of the pantothenate-biosynthesis pathway. Biochem Soc Trans 31, 563–571. https://doi.org/10.1042/bst0310563

Louca, S., Doebeli, M., Parfrey, L.W., 2018. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome 6, 41. https://doi.org/10.1186/s40168-018-0420-9

Mackey, K.R.M., Labiosa, R.G., Calhoun, M., Street, J.H., Post, A.F., Paytan, A., 2007. Phosphorus availability, phytoplankton community dynamics, and taxon-specific phosphorus status in the Gulf of Aqaba, Red Sea. Limnology and Oceanography 52, 873–885. https://doi.org/10.4319/lo.2007.52.2.0873

Majda, S., Beisser, D., Boenigk, J., 2021. Nutrient-driven genome evolution revealed by comparative genomics of chrysomonad flagellates. Communications Biology 4, 328. https://doi.org/10.1038/s42003-021-01781-3

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution 38, 4647–4654. https://doi.org/10.1093/molbev/msab199

Marin, B., M. Nowack, E.C., Melkonian, M., 2005. A Plastid in the Making: Evidence for a Second Primary Endosymbiosis. Protist 156, 425–432. https://doi.org/10.1016/j.protis.2005.09.001

Martens, J.H., Barg, H., Warren, M.J., Jahn, D., 2002. Microbial production of vitamin B12. Appl Microbiol Biotechnol 58, 275–285. https://doi.org/10.1007/s00253-001-0902-7

Martin, D.L., Chiari, Y., Boone, E., Sherman, T.D., Ross, C., Wyllie-Echeverria, S., Gaydos, J.K., Boettcher, A.A., 2016. Functional, Phylogenetic and Host-Geographic Signatures of Labyrinthula spp. Provide for Putative Species Delimitation and a Global-Scale View of Seagrass Wasting Disease. Estuaries and Coasts 39, 1403–1421. https://doi.org/10.1007/s12237-016-0087-z

Martin, K., Schmidt, K., Toseland, A., Boulton, C.A., Barry, K., Beszteri, B., Brussaard, C.P.D., Clum, A., Daum, C.G., Eloe-Fadrosh, E., Fong, A., Foster, Brian, Foster, Bryce, Ginzburg, M., Huntemann, M., Ivanova, N.N., Kyrpides, N.C., Lindquist, E., Mukherjee, S., Palaniappan, K., Reddy, T.B.K., Rizkallah, M.R., Roux, S., Timmermans, K., Tringe, S.G., van de Poll, W.H., Varghese, N., Valentin, K.U., Lenton, T.M., Grigoriev, I.V., Leggett, R.M., Moulton, V., Mock, T., 2021. The biogeographic differentiation of algal microbiomes in the upper ocean from pole to pole. Nature Communications 12, 5483. https://doi.org/10.1038/s41467-021-25646-9

Martinez-Garcia, M., Brazel, D., Poulton, N.J., Swan, B.K., Gomez, M.L., Masland, D., Sieracki, M.E., Stepanauskas, R., 2012. Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. The ISME Journal 6, 703–707. https://doi.org/10.1038/ismej.2011.126

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K., Pedrós-Alió, C., 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. Applied

and environmental microbiology 70, 3528–34. https://doi.org/10.1128/AEM.70.6.3528-3534.2004

Massana, R., del Campo, J., Sieracki, M.E., Audic, S., Logares, R., 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. The ISME Journal 8, 854–866. https://doi.org/10.1038/ismej.2013.204

Massana, R., Guillou, L., Díez, B., Pedrós-Alió, C., 2002. Unveiling the Organisms behind Novel Eukaryotic Ribosomal DNA Sequences from the Ocean. Appl. Environ. Microbiol. 68, 4554–4558. https://doi.org/10.1128/AEM.68.9.4554-4558.2002

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

McRose, D., Guo, J., Monier, A., Sudek, S., Wilken, S., Yan, S., Mock, T., Archibald, J.M., Begley, T.P., Reyes-Prieto, A., Worden, A.Z., 2014. Alternatives to vitamin B1 uptake revealed with discovery of riboswitches in multiple marine eukaryotic lineages. ISME J 8, 2517–2529. https://doi.org/10.1038/ismej.2014.146

Medlin, L.K., Kooistra, W.H., Gersonde, R., Wellbrock, U., 1996. Evolution of the diatoms (Bacillariophyta). II. Nuclear-encoded small-subunit rRNA sequence comparisons confirm a paraphyletic origin for the centric diatoms. Mol Biol Evol 13, 67–75. https://doi.org/10.1093/oxfordjournals.molbev.a025571

Meyer-Ficca, M., Kirkland, J.B., 2016. Niacin. Advances in nutrition (Bethesda, Md.) 7, 556–8. https://doi.org/10.3945/an.115.011239

Miflin, B.J., Lea, P.J., 1980. 4 - Ammonia Assimilation, in: Miflin, B.J. (Ed.), Amino Acids and Derivatives. Academic Press, pp. 169–202. https://doi.org/10.1016/B978-0-12-675405-6.50010-3

Miller, J.R., Zhou, P., Mudge, J., Gurtowski, J., Lee, H., Ramaraj, T., Walenz, B.P., Liu, J., Stupar, R.M., Denny, R., Song, L., Singh, N., Maron, L.G., McCouch, S.R., McCombie, W.R., Schatz, M.C., Tiffin, P., Young, N.D., Silverstein, K.A.T., 2017. Hybrid assembly with long and short reads improves discovery of gene family expansions. BMC Genomics 18, 541. https://doi.org/10.1186/s12864-017-3927-8

Miller, M.B., Bassler, B.L., 2001. Quorum sensing in bacteria. Annu Rev Microbiol 55, 165–199. https://doi.org/10.1146/annurev.micro.55.1.165

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: The protein families database in 2021. Nucleic Acids Research 49, D412–D419. https://doi.org/10.1093/nar/gkaa913

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C.,

Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.-Y., Finn, R.D., 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Research 47, D351–D360. https://doi.org/10.1093/nar/gky1100

Mock, T., Otillar, R.P., Strauss, J., McMullan, M., Paajanen, P., Schmutz, J., Salamov, A., Sanges, R., Toseland, A., Ward, B.J., Allen, A.E., Dupont, C.L., Frickenhaus, S., Maumus, F., Veluchamy, A., Wu, T., Barry, K.W., Falciatore, A., Ferrante, M.I., Fortunato, A.E., Glöckner, G., Gruber, A., Hipkin, R., Janech, M.G., Kroth, P.G., Leese, F., Lindquist, E.A., Lyon, B.R., Martin, J., Mayer, C., Parker, M., Quesneville, H., Raymond, J.A., Uhlig, C., Valas, R.E., Valentin, K.U., Worden, A.Z., Armbrust, E.V., Clark, M.D., Bowler, C., Green, B.R., Moulton, V., van Oosterhout, C., Grigoriev, I.V., 2017. Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus. Nature 541, 536–540. https://doi.org/10.1038/nature20803

Monteverde, D.R., Gómez-Consarnau, L., Suffridge, C., Sañudo-Wilhelmy, S.A., 2017. Life's utilization of B vitamins on early Earth. Geobiology 15, 3–18. https://doi.org/10.1111/gbi.12202

Moore, C.M., 2016. Diagnosing oceanic nutrient deficiency. Philos Trans A Math Phys Eng Sci 374, 20150290. https://doi.org/10.1098/rsta.2015.0290

Moore, C.M., Mills, M.M., Arrigo, K.R., Berman-Frank, I., Bopp, L., Boyd, P.W., Galbraith, E.D., Geider, R.J., Guieu, C., Jaccard, S.L., Jickells, T.D., La Roche, J., Lenton, T.M., Mahowald, N.M., Marañón, E., Marinov, I., Moore, J.K., Nakatsuka, T., Oschlies, A., Saito, M.A., Thingstad, T.F., Tsuda, A., Ulloa, O., 2013. Processes and patterns of oceanic nutrient limitation. Nature Geoscience 6, 701–710. https://doi.org/10.1038/ngeo1765

Moore, J.K., Doney, S.C., Glover, D.M., Fung, I.Y., 2001. Iron cycling and nutrient-limitation patterns in surface waters of the World Ocean. Deep Sea Research Part II: Topical Studies in Oceanography, The US JGOFS Synthesis and Modeling Project: Phase 1 49, 463–507. https://doi.org/10.1016/S0967-0645(01)00109-6

Moustafa, A., Beszteri, B., Maier, U.G., Bowler, C., Valentin, K., Bhattacharya, D., 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science 324, 1724–1726. https://doi.org/10.1126/science.1172983

Naish, M., Alonge Michael, Wlodzimierz Piotr, Tock Andrew J., Abramson Bradley W., Schmücker Anna, Mandáková Terezie, Jamge Bhagyshree, Lambing Christophe, Kuo Pallas, Yelina Natasha, Hartwick Nolan, Colt Kelly, Smith Lisa M., Ton Jurriaan, Kakutani Tetsuji, Martienssen Robert A., Schneeberger Korbinian, Lysak Martin A., Berger Frédéric, Bousios Alexandros, Michael Todd P., Schatz Michael C., Henderson Ian R., 2021. The genetic and epigenetic landscape of the Arabidopsis centromeres. Science 374, eabi7489. https://doi.org/10.1126/science.abi7489

Nealson, K.H., Hastings, J.W., 1979. Bacterial bioluminescence: its control and ecological significance. Microbiol Rev 43, 496–518.

Nelson, D.M., Treguer, P., Brezezinski, M.A., Leynaert, A., Queguiner, B., 1995. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation - Nelson - 1995 - Global Biogeochemical Cycles - Wiley Online Library [WWW Document]. URL https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/95GB01070 (accessed 5.10.22).

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V., Dubchak, I., 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res 42, D26-31. https://doi.org/10.1093/nar/gkt1069

Opal, S.M., 2009. A Brief History of Microbiology and Immunology. Vaccines: A Biography 31–56. https://doi.org/10.1007/978-1-4419-1108-7_3

Ortiz-Burgos, S., 2016. Shannon-Weaver Diversity Index, in: Kennish, M.J. (Ed.), Encyclopedia of Estuaries. Springer Netherlands, Dordrecht, pp. 572–573. https://doi.org/10.1007/978-94-017-8801-4_233

Paerl, R.W., Sundh, J., Tan, D., Svenningsen, S.L., Hylander, S., Pinhassi, J., Andersson, A.F., Riemann, L., 2018. Prevalent reliance of bacterioplankton on exogenous vitamin B1 and precursor availability. Proceedings of the National Academy of Sciences of the United States of America 201806425–201806425. https://doi.org/10.1073/pnas.1806425115

Pajares, S., Ramos, R., 2019. Processes and Microorganisms Involved in the Marine Nitrogen Cycle: Knowledge and Gaps. Frontiers in Marine Science 6.

Powers, H.J., 2003. Riboflavin (vitamin B-2) and health. The American Journal of Clinical Nutrition 77, 1352–1360. https://doi.org/10.1093/ajcn/77.6.1352

Praekelt, U.M., Meacock, P.A., 1992. MOL1, a Saccharomyces cerevisiae gene that is highly expressed in early stationary phase during growth on molasses. Yeast 8, 699–710. https://doi.org/10.1002/yea.320080903

Prasad, P.D., Wang, H., Kekuda, R., Fujita, T., Fei, Y.J., Devoe, L.D., Leibach, F.H., Ganapathy, V., 1998. Cloning and functional expression of a cDNA encoding a mammalian sodium-dependent vitamin transporter mediating the uptake of pantothenate, biotin, and lipoate. J Biol Chem 273, 7501–7506. https://doi.org/10.1074/jbc.273.13.7501

Preisig, H.R., Andersen, R.A., 2005. Historical review of algal culturing techniques. Algal culturing techniques 65, 79–82.

Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution 26, 1641–1650. https://doi.org/10.1093/molbev/msp077

Pruesse, E., Peplies, J., Glöckner, F.O., 2012. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28, 1823–1829. https://doi.org/10.1093/bioinformatics/bts252

Qiu, A., Jansen, M., Sakaris, A., Min, S.H., Chattopadhyay, S., Tsai, E., Sandoval, C., Zhao, R., Akabas, M.H., Goldman, I.D., 2006. Identification of an Intestinal Folate Transporter and the Molecular Basis for Hereditary Folate Malabsorption. Cell 127, 917–928. https://doi.org/10.1016/j.cell.2006.09.041

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team, 2021. R: A language and environment for statistical computing.

Raghukumar, S., 2002. Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). European Journal of Protistology 38, 127–145. https://doi.org/10.1078/0932-4739-00832

Raghukumar, S., Damare, V.S., 2011. Increasing evidence for the important role of Labyrinthulomycetes in marine ecosystems 54, 3–11. https://doi.org/10.1515/bot.2011.008

Ranallo-Benavidez, T.R., Jaron, K.S., Schatz, M.C., 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature Communications 11, 1432. https://doi.org/10.1038/s41467-020-14998-3

Ren, J., Chaisson, M.J.P., 2021. lra: A long read aligner for sequences and contigs. PLOS Computational Biology 17, e1009078. https://doi.org/10.1371/journal.pcbi.1009078

Revuelta, J.L., Serrano-Amatriain, C., Ledesma-Amaro, R., Jiménez, A., 2018. Formation of folates by microorganisms: towards the biotechnological production of this vitamin. Applied microbiology and biotechnology 102, 8613–8620. https://doi.org/10.1007/s00253-018-9266-0

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics 13, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Robbins, W.J., 1938. Thiamin and Growth of Species of Phytophthora. Bulletin of the Torrey Botanical Club 65, 267–267. https://doi.org/10.2307/2481231

Robinson. F. A., 1951. The vitamin B complex. Journal of the American Pharmaceutical Association 40.

Roje, S., 2007. Vitamin B biosynthesis in plants. Phytochemistry 68, 1904–1921. https://doi.org/10.1016/J.PHYTOCHEM.2007.03.038

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.-H., Falcón, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M.,

Venter, J.C., 2007. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biology 5, e77–e77. https://doi.org/10.1371/journal.pbio.0050077

Sañudo-Wilhelmy, S.A., Cutter, L.S., Durazo, R., Smail, E.A., Gómez-Consarnau, L., Webb, E.A., Prokopenko, M.G., Berelson, W.M., Karl, D.M., 2012. Multiple B-vitamin depletion in large areas of the coastal ocean. Proc Natl Acad Sci U S A 109, 14041–14045. https://doi.org/10.1073/pnas.1208755109

Sañudo-Wilhelmy, S.A., Gobler, C.J., Okbamichael, M., Taylor, G.T., 2006. Regulation of phytoplankton dynamics by vitamin B 12. Geophysical Research Letters 33, L04604–L04604. https://doi.org/10.1029/2005GL025046

Sañudo-Wilhelmy, S.A., Gómez-Consarnau, L., Suffridge, C., Webb, E.A., 2014. The Role of B Vitamins in Marine Biogeochemistry. Annu. Rev. Mar. Sci. 6, 339–367. https://doi.org/10.1146/annurev-marine-120710-100912

Salgar-Chaparro, S.J., Machuca, L.L., 2019. Complementary DNA/RNA-Based Profiling: Characterization of Corrosive Microbial Communities and Their Functional Profiles in an Oil Production Facility. Frontiers in Microbiology 10.

Sayer, A.P., 2020. Investigating the molecular basis of vitamin B12 uptake in Chlamydomonas reinhardtii (Thesis). University of Cambridge. https://doi.org/10.17863/CAM.49830

Schwartz, C.J., Djaman Ouliana, Imlay James A., Kiley Patricia J., 2000. The cysteine desulfurase, IscS, has a major role in in vivo Fe-S cluster formation in Escherichia coli. Proceedings of the National Academy of Sciences 97, 9009–9014. https://doi.org/10.1073/pnas.160261497

Scott-Somme, K., Harley, J., Adkins, P., 2022. Sequencing our seas. The Biochemist 44, 2–7. https://doi.org/10.1042/bio_2022_106

Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., Leconte, J., Mangot, J.-F., Poulain, J., Labadie, K., Logares, R., Sunagawa, S., de Berardinis, V., Salanoubat, M., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Pesant, S., Poulton, N., Stepanauskas, R., Bork, P., Bowler, C., Hingamp, P., Sullivan, M.B., Iudicone, D., Massana, R., Aury, J.-M., Henrissat, B., Karsenti, E., Jaillon, O., Sieracki, M., de Vargas, C., Wincker, P., 2018. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. Nature Communications 9, 310–310. https://doi.org/10.1038/s41467-017-02235-3

Seif Yara, Choudhary Kumari Sonal, Hefner Ying, Anand Amitesh, Yang Laurence, Palsson Bernhard O., 2020. Metabolic and genetic basis for auxotrophies in Gram-negative species. Proceedings of the National Academy of Sciences 117, 6264–6273. https://doi.org/10.1073/pnas.1910499117

Seppey, M., Manni, M., Zdobnov, E.M., 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol 1962, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

Seymour, J.R., Simó, R., Ahmed, T., Stocker, R., 2010. Chemoattraction to Dimethylsulfoniopropionate Throughout the Marine Microbial Food Web. Science 329, 342–345. https://doi.org/10.1126/science.1188418

Sharma, N., Simon, D.P., Diaz-Garza, A.M., Fantino, E., Messaabi, A., Meddeb-Mouelhi, F., Germain, H., Desgagné-Penix, I., 2021. Diatoms Biotechnology: Various Industrial Applications for a Greener Tomorrow. Frontiers in Marine Science 8.

Sharma, V., Rodionov, D.A., Leyn, S.A., Tran, D., Iablokov, S.N., Ding, H., Peterson, D.A., Osterman, A.L., Peterson, S.N., 2019. B-Vitamin Sharing Promotes Stability of Gut Microbial Communities. Frontiers in Microbiology 10, 1485–1485. https://doi.org/10.3389/fmicb.2019.01485

Shaw, A.K., Halpern, A.L., Beeson, K., Tran, B., Venter, J.C., Martiny, J.B.H., 2008. It's all relative: ranking the diversity of aquatic bacterial communities. Environ Microbiol 10, 2200–2210. https://doi.org/10.1111/j.1462-2920.2008.01626.x

Shaw, D.K., Sekar, J., Ramalingam, P.V., 2022. Recent insights into oceanic dimethylsulfoniopropionate biosynthesis and catabolism. Environmental Microbiology n/a. https://doi.org/10.1111/1462-2920.16045

Shelton, A.N., Seth, E.C., Mok, K.C., Han, A.W., Jackson, S.N., Haft, D.R., Taga, M.E., 2019. Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. ISME J 13, 789–804. https://doi.org/10.1038/s41396-018-0304-9

Shibl, A.A., Isaac, A., Ochsenkühn, M.A., Cárdenas, A., Fei, C., Behringer, G., Arnoux, M., Drou, N., Santos, M.P., Gunsalus, K.C., Voolstra, C.R., Amin, S.A., 2020. Diatom modulation of select bacteria through use of two unique secondary metabolites. Proc Natl Acad Sci U S A 117, 27445–27455. https://doi.org/10.1073/pnas.2012088117

Shiratori, T., Nakayama, T., Ishida, K., 2015. A New Deep-branching Stramenopile, Platysulcus tardus gen. nov., sp. nov. Protist 166, 337–348. https://doi.org/10.1016/j.protis.2015.05.001

Shiratori, T., Thakur, R., Ishida, K.-I., 2017. Pseudophyllomitus vesiculosus (Larsen and Patterson 1990) Lee, 2002, a Poorly Studied Phagotrophic Biflagellate is the First Characterized Member of Stramenopile Environmental Clade MAST-6. Protist 168, 439–451. https://doi.org/10.1016/j.protis.2017.06.004

Shreiner, A.B., Kao, J.Y., Young, V.B., 2015. The gut microbiome in health and in disease. Curr Opin Gastroenterol 31, 69–75. https://doi.org/10.1097/MOG.0000000000000139

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Solomon Goldstein, 1973. ZO OSPORIC MARINE FUNGI (THRAUSTO CHYTRIACEAE AND DERMOCYSTIDIACEAE). Annu. Rev. Microbiol. 13–26.

Song, J., Choi, A., Im, M., Joung, Y., Yoshizawa, S., Cho, J.-C., Kogure, K. 2015, 2015. Aurantivirga profunda gen. nov., sp. nov., isolated from deep-seawater, a novel member of the family Flavobacteriaceae. International Journal of Systematic and Evolutionary Microbiology 65, 4850–4856. https://doi.org/10.1099/ijsem.0.000662

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33, W465–W467. https://doi.org/10.1093/nar/gki458

Steele, J.H., 2013. Marine Mesocosms☆, in: Reference Module in Earth Systems and Environmental Sciences. Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.04207-X

Stephens, T.G., Gabr, A., Calatrava, V., Grossman, A.R., Bhattacharya, D., 2021. Why is primary endosymbiosis so rare? New Phytologist 231, 1693–1699. https://doi.org/10.1111/nph.17478

Stiller, J.W., Huang, J., Ding, Q., Tian, J., Goodwillie, C., 2009. Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? BMC Genomics 10, 484. https://doi.org/10.1186/1471-2164-10-484

Stiller, J.W., Schreiber, J., Yue, J., Guo, H., Ding, Q., Huang, J., 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. Nature Communications 5, 5764. https://doi.org/10.1038/ncomms6764

Stolz, J., Vielreicher, M., 2003. Tpn1p, the plasma membrane vitamin B6 transporter of Saccharomyces cerevisiae. J Biol Chem 278, 18990–18996. https://doi.org/10.1074/jbc.M300949200

Strauss, A. 1979, 1979. Isolation and Genetical Classification of Aromatic Amino Acid Auxotrophic Mutants in Schizosaccharomyces pombe. Microbiology 113, 173–176. https://doi.org/10.1099/00221287-113-1-173

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., Cornejo-Castillo, F.M., Costea, P.I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J.M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B.T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans coordinators, T.O., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M.B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S.G., Bork, P., 2015. Ocean plankton. Structure and function of the global ocean microbiome. Science (New York, N.Y.) 348, 1261359–1261359. https://doi.org/10.1126/science.1261359

Tambasco-Studart, M., Titiz, O., Raschle, T., Forster, G., Amrhein, N., Fitzpatrick, T.B., 2005. Vitamin B6 biosynthesis in higher plants. Proc Natl Acad Sci U S A 102, 13687–13692. https://doi.org/10.1073/pnas.0506228102

Tang, Y.Z., Koch, F., Gobler, C.J., 2010. Most harmful algal bloom species are vitamin B1 and B12 auxotrophs. Proceedings of the National Academy of Sciences 107, 20756–20761. https://doi.org/10.1073/PNAS.1009566107

Tani, N., Yoneda, K., Suzuki, I., 2018. The effect of thiamine on the growth and fatty acid content of Aurantiochytrium sp. Algal Research 36, 57–66. https://doi.org/10.1016/J.ALGAL.2018.10.012

Taylor, J.D., Cunliffe, M., 2014. High-throughput sequencing reveals neustonic and planktonic microbial eukaryote diversity in coastal waters. Journal of Phycology 50, 960–965. https://doi.org/10.1111/jpy.12228

Ternes, C.M., Schönknecht, G., 2014. Gene Transfers Shaped the Evolution of De Novo NAD + Biosynthesis in Eukaryotes. Genome Biology and Evolution 6, 2335–2349. https://doi.org/10.1093/gbe/evu185

Thakur, R., Shiratori, T., Ishida, K., 2019. Taxon-rich Multigene Phylogenetic Analyses Resolve the Phylogenetic Relationship Among Deep-branching Stramenopiles. Protist 170, 125682–125682. https://doi.org/10.1016/J.PROTIS.2019.125682

Tyler, B.M., 2007. Phytophthora sojae: root rot pathogen of soybean and model oomycete. Mol Plant Pathol 8, 1–8. https://doi.org/10.1111/j.1364-3703.2006.00373.x

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., Thermes, C., 2018. The Third Revolution in Sequencing Technology. Trends in Genetics 34, 666–681. https://doi.org/10.1016/j.tig.2018.05.008

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., Smith, H.O., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74. https://doi.org/10.1126/science.1093857

Viana, T.V., Fistarol, G.O., Amario, M., Menezes, R.B., Carneiro, B.L.R., Chaves, D.M., Hargreaves, P.I., Silva-Lima, A.W., Valentin, J.L., Tenenbaum, D.R., Arruda, E.F., Paranhos, R., Salomon, P.S., 2019. Massive Blooms of Chattonella subsalsa Biecheler (Raphidophyceae) in a Hypereutrophic, Tropical Estuary—Guanabara Bay, Brazil. Frontiers in Marine Science 6.

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., Schatz, M.C., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE 9, e112963. https://doi.org/10.1371/journal.pone.0112963

Wang, Q., Sun, H., Huang, J., 2017. Re-analyses of “Algal” Genes Suggest a Complex Evolutionary History of Oomycetes. Frontiers in Plant Science 8.

Wang, Yunhao, Zhao, Y., Bollas, A., Wang, Yuru, Au, K.F., 2021. Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology 39, 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Webb, E., Claas, K., Downs, D., 1998. thiBPQ Encodes an ABC Transporter Required for Transport of Thiamine and Thiamine Pyrophosphate inSalmonella typhimurium *. Journal of Biological Chemistry 273, 8946–8950. https://doi.org/10.1074/jbc.273.15.8946

Webb, M.E., Marquet, A., Mendel, R.R., Rébeillé, F., Smith, A.G., 2007. Elucidating biosynthetic pathways for vitamins and cofactors. Natural Product Reports 24, 988–988. https://doi.org/10.1039/b703105j

Webster, J., Weber, R., 2007. Introduction to Fungi, 3rd ed. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511809026

Weisse, T., 2017. Functional diversity of aquatic ciliates. European Journal of Protistology, Integrating the three dimensions of ciliate diversity: function, taxonomy, and genetics 61, 331–358. https://doi.org/10.1016/j.ejop.2017.04.001

Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Prokaryotes: The unseen majority. PNAS 95, 6578–6583. https://doi.org/10.1073/pnas.95.12.6578

Whittaker, R.H., 1972. Evolution and Measurement of Species Diversity. TAXON 21, 213–251. https://doi.org/10.2307/1218190

Wightman, R., Meacock, P.A.Y. 2003, 2003. The THI5 gene family of Saccharomyces cerevisiae: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine. Microbiology 149, 1447–1460. https://doi.org/10.1099/mic.0.26194-0

Woodward, E.M.S., Harris, C., 2014. Nutrient concentration profiles from long term time series at Station L4 in the Western English Channel from 2000 to 2013. https://doi.org/10.5285/F1968A39-26BF-55FE-E044-000B5DE50F38

Yamashiro, T., Yasujima, T., Said, H.M., Yuasa, H., 2020. pH-dependent pyridoxine transport by SLC19A2 and SLC19A3: Implications for absorption in acidic microclimates. Journal of Biological Chemistry 295, 16998–17008. https://doi.org/10.1074/jbc.RA120.013610

Yan, S., Fuchs, B.M., Lenk, S., Harder, J., Wulf, J., Jiao, N.-Z., Amann, R., 2009. Biogeography and phylogeny of the NOR5/OM60 clade of Gammaproteobacteria. Systematic and Applied Microbiology 32, 124–139. https://doi.org/10.1016/j.syapm.2008.12.001

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Research 42, D643–D648. https://doi.org/10.1093/nar/gkt1209

Yonezawa, A., Inui, K., 2013. Novel riboflavin transporter family RFVT/SLC52: identification, nomenclature, functional characterization and genetic diseases of RFVT/SLC52. Mol Aspects Med 34, 693–701. https://doi.org/10.1016/j.mam.2012.07.014

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., Bhattacharya, D., 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21, 809–818. https://doi.org/10.1093/molbev/msh075

Zempleni, J., Wijeratne, S.S.K., Hassan, Y.I., 2009. Biotin. BioFactors (Oxford, England) 35, 36–46. https://doi.org/10.1002/biof.8

Zengler, K., Zaramela, L.S., 2018. The social network of microorganisms — how auxotrophies shape complex communities. Nat Rev Microbiol 16, 383–390. https://doi.org/10.1038/s41579-018-0004-5

Zhang, X.-H., Liu, Ji, Liu, Jingli, Yang, G., Xue, C.-X., Curson, A.R.J., Todd, J.D., 2019. Biogenic production of DMSP and its degradation to DMS-their roles in the global sulfur cycle. Sci China Life Sci 62, 1296–1319. https://doi.org/10.1007/s11427-018-9524-y

Zimin, A.V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A., Dvořák, J., Salzberg, S.L., 2017. Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res 27, 787–792. https://doi.org/10.1101/gr.213405.116

Appendix 1 – Chapter 2 - Assessing the effect of B vitamin amendments to natural assemblages of oceanic microbes

1.1 Amplicon workflow used for Chapter 2 analysis

Load required packages

```
# use the library() function to load the packages you would like to work with

install.packages("dada2")
install.packages("phyloseq")
install.packages("ggplot2")
install.packages("cowplot")
install.packages("knitr")
install.packages("zoo")
install.packages("tibble")
install.packages("vegan")
install.packages("microbiome")

install.packages("devtools")
library("devtools")
devtools::install_github("benjjneb/dada2", ref="v1.16") # change the ref argument to get ot
her versions

library(dada2)
library(phyloseq)
library(ggplot2)
library(cowplot)
library(knitr)
library(zoo)
library(tibble)
library(vegan)
#library(microbiome)
library(ade4)
library(ape)
library(tidyverse)
library(reshape2)
library(egg)
library(gghighlight)
library(ggrepel)
library(ape)
library(dabestr)
library(ade4)
library(mgcv)
library(reshape)
library(dplyr)
library(plyr)
library(scales)
library(seqinr)
#library(ggtree)
library(colorspace)
library(pals)
#library(DESeq2)
library(MBA)
```

Programme custom functions

'ggrare' generates a rarefaction curve at the end of this processing pipeline.

```r
# ggrare plots a rarefaction curve from a phyloseq object
ggrare <- function(physeq_object, step = 10, label = NULL, color = NULL, plot = TRUE, paral
lel = FALSE, se = TRUE) {

  x <- methods::as(phyloseq::otu_table(physeq_object), "matrix")
  if (phyloseq::taxa_are_rows(physeq_object)) { x <- t(x) }

  ## This script is adapted from vegan `rarecurve` function
  tot <- rowSums(x)
  S <- rowSums(x > 0)
  nr <- nrow(x)

  rarefun <- function(i) {
    cat(paste("rarefying sample", rownames(x)[i]), sep = "\n")
    n <- seq(1, tot[i], by = step)
    if (n[length(n)] != tot[i]) {
      n <- c(n, tot[i])
    }
    y <- vegan::rarefy(x[i, ,drop = FALSE], n, se = se)
    if (nrow(y) != 1) {
      rownames(y) <- c(".S", ".se")
      return(data.frame(t(y), Size = n, Sample = rownames(x)[i]))
    } else {
      return(data.frame(.S = y[1, ], Size = n, Sample = rownames(x)[i]))
    }
  }
  if (parallel) {
    out <- parallel::mclapply(seq_len(nr), rarefun, mc.preschedule = FALSE)
  } else {
    out <- lapply(seq_len(nr), rarefun)
  }
  df <- do.call(rbind, out)

  # Get sample data
  if (!is.null(phyloseq::sample_data(physeq_object, FALSE))) {
    sdf <- methods::as(phyloseq::sample_data(physeq_object), "data.frame")
    sdf$Sample <- rownames(sdf)
    data <- merge(df, sdf, by = "Sample")
    labels <- data.frame(x = tot, y = S, Sample = rownames(x))
    labels <- merge(labels, sdf, by = "Sample")
  }

  # Add, any custom-supplied plot-mapped variables
  if ( length(color) > 1 ) {
    data$color <- color
    names(data)[names(data) == "color"] <- deparse(substitute(color))
    color <- deparse(substitute(color))
  }

  if ( length(label) > 1 ) {
    labels$label <- label
    names(labels)[names(labels) == "label"] <- deparse(substitute(label))
    label <- deparse(substitute(label))
  }

  p <- ggplot2::ggplot(data = data,
                       ggplot2::aes_string(x = "Size",
                                           y = ".S",
                                           group = "Sample",
                                           color = color))

  p <- p + ggplot2::labs(x = "Sequence Sample Size", y = "Species Richness")

  if (!is.null(label)) {
    p <- p + ggplot2::geom_text(data = labels,
                                ggplot2::aes_string(x = "x",
```

```
                                                y = "y",
                                                label = label,
                                                color = color),
                                size = 4, hjust = 0)
  }

  p <- p + ggplot2::geom_line()
  if (se) { ## add standard error if available
    p <- p +
      ggplot2::geom_ribbon(ggplot2::aes_string(ymin = ".S - .se",
                                               ymax = ".S + .se",
                                               color = NULL,
                                               fill = color),
                           alpha = 0.2)
  }
  if (plot) {
    plot(p)
  }
  invisible(p)
}
```

Set working directory and file paths

Tell R where to look for the files we want to upload, and where we want R to save any output files.

```
# path to location of raw sequencing files
path <- "D:/Amplicon seq/Absolon16S/Absolon16S"

# use the list.files() function to print all files in the path directory
list.files(path)
```

Identify forward and reverse paired-end read files

Pattern parameter is used to identify consistent suffixes of file names, which indicate whether each file is a forward or reverse read file.

```
# extract first part of file names for future reference
f.names <- as.vector(list.files(path, pattern = "_R1_001.fastq", full.names = FALSE))
r.names <- as.vector(list.files(path, pattern = "_R2_001.fastq", full.names = FALSE))

# identify forward and reverse read files
fnFs <- sort(list.files(path, pattern = "_R1_001.fastq", full.names = TRUE))
fnRs <- sort(list.files(path, pattern = "_R2_001.fastq", full.names = TRUE))
```

Plot sequence quality profiles

Sequence read quality information is contained within each fastq file.

```
# plot quality profiles for forward and reverse reads, only the first 9 samples are selecte
d
qpf <- plotQualityProfile(fnFs[1:9]) #note we only choose to visualise 9 samples
qpr <- plotQualityProfile(fnRs[1:9])

# save quality profiles (optional)
ggsave("20211008_16S_quality_F.jpeg", qpf, width = 20, height = 16, units = "cm", device =
"jpeg")
```

```
ggsave("20211008_16S_quality_R.jpeg", qpr, width = 20, height = 16, units = "cm", device =
"jpeg")

# display quality profiles
qpf
qpr
```

Setup directory for filtered and trimmed sequences

Create a new directory location where filtered and trimmed sequence files can be saved after the following quality control steps.

```
# sets up new directory titled 'filtered' within the current file path
filt_path <- file.path(path, "20211008_filtered")

# sets the way we want R to save new files into the new directory
filtFs <- file.path(filt_path, paste0(f.names, "_F_filt.fastq.gz"))
filtRs <- file.path(filt_path, paste0(r.names, "_R_filt.fastq.gz"))
```

Filter and trim sequences

At this stage, remove primers from forward and reverse reads respectively, and truncate reads based on sequence quality drop off. An error estimating parameter is then used to throw out sequences which are below a certain overall quality. Note that the trim and truncate parameters *must* be adjusted manually here based on primer length and inspect of quality profiles. – Filtering using adapter length of 17 for 16Sused by IMR https://imr.bio/protocols.html

```
start_time_1 <- Sys.time()
out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, truncLen=c(270,220), trimLeft=c(17,17), ma
xN=0, truncQ=2, rm.phix=TRUE, compress=TRUE)
end_time_1 <- Sys.time()
head(out)
end_time_1 - start_time_1 # 18 minutes
```

Learn error rates

DADA2 relies of error estimation to resolve amplicon sequence variants. Here, the predicted error rate for each potential combination of bases is calculated and regressed against the quality score at the read position. Error rates should decrease as a function of quality score, and these plots should be investigated for major deviations. These error rates are then used downstream.

```
# calculate the error model
start_time_2 <- Sys.time()
errF <- learnErrors(filtFs, multithread = TRUE)
errR <- learnErrors(filtRs, multithread = TRUE)
end_time_2 <- Sys.time()
end_time_2 - start_time_2 # 1.26 Hours


# plot errors
plot.errF <- plotErrors(errF, nominalQ = TRUE) + theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
plot.errR <- plotErrors(errR, nominalQ = TRUE) + theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
```

```
ggsave("20211008_16S_cunlif_error_F.jpeg", qpf, width = 20, height = 16, units = "cm", devi
ce = "jpeg")
ggsave("20211008_16S_cunlif_error_R.jpeg", qpr, width = 20, height = 16, units = "cm", devi
ce = "jpeg")

# display error plots
plot.errF
plot.errR
```

Dereplicate sequences

Sequences are dereplicated, meaning that identical sequences are removed (temporarily). This reduces the computational demand when resolving amplicon sequence variants by eliminating redundant comparisons.

```
# dereplicate sequences
start_time_3 <- Sys.time()
derepFs <- derepFastq(filtFs, verbose = TRUE)
derepRs <- derepFastq(filtRs, verbose = TRUE)
end_time_3 <- Sys.time()
end_time_3 - start_time_3 # 3 minutes

# match names for clarity
names(derepFs) <- f.names
names(derepRs) <- r.names
```

Resolve amplicon sequence variants (ASVs)

The DADA2 algorithm is then applied to dereplicated sequences, resolving amplicon sequence variants at single nucleotide resolution for forward and reverse reads independently.

```
start_time_4 <- Sys.time()
dadaFs <- dada(derepFs, err = errF, multithread = TRUE)
dadaRs <- dada(derepRs, err = errR, multithread = TRUE)
end_time_4 <- Sys.time()
end_time_4 - start_time_4 # 2 hours
```

Merge paired reads

Merge forward and reverse reads, yielding full length sequences for the targetted amplicon region.

```
start_time_5 <- Sys.time()
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose = TRUE)
end_time_5 <- Sys.time()
end_time_5 - start_time_5 # 8 minutes
```

Generate sequence table

Generate a sequencing table to see the number of reads assigned to each ASV across all samples.

```
seqtab <- makeSequenceTable(mergers)
table(nchar(getSequences(seqtab)))
```

Remove chimeras

Chimeras are removed by searching for bimeric sequences. This is an abundance based algorithm, which searches for ASVs where each half of the sequence could be assigned to a different *more abundant* ASV. These sequences are assumed to be chimeric, as the abundance of chimeras should always be lower than the abundance of true reads.

```
start_time_6 <- Sys.time()
seqtab.chi <- removeBimeraDenovo(seqtab, method = "consensus", multithread = TRUE, verbose
= TRUE)
end_time_6 <- Sys.time()
end_time_6 - start_time_6 # 2.6 hours
```

Track sequence loss

Generate a table to track the loss of sequences at different stages of the pipeline.

```
# function for sum of unique sequences in a dataframe
getN <- function(x) sum(getUniques(x))

# bind columns: sequence; unique sequences after dada2 (forward only); unique sequences aft
er merging forward and reverse reads; unique sequences after chimera removal
track <- cbind(out, sapply(dadaFs, getN), sapply(mergers, getN), rowSums(seqtab.chi))
colnames(track) <- c("input", "filtered", "denoised", "merged", "no chim")
rownames(track) <- f.names

# print table
print(track)
```

Assign taxonomy

Assign the taxonomy to ASV sequences according to your chosen reference database. SILVA release 132 is preferable for prokaryotic 16S, whilst PR2 is preferable for eukaryotic 18S.

```
# assign main taxonomy against SILVA database
start_time_7 <- Sys.time()
silva.taxa <- assignTaxonomy(seqtab.chi,"C:/Users/domin/Documents/PhD/Microbial_ecology_res
ources/silva_nr_v132_train_set.fa" , multithread = TRUE)
end_time_7 <- Sys.time()
end_time_7 - start_time_7 # 11 minutes

## assign species level taxonomy against SILVA species database (optional)
#silva.taxa <- addSpecies(silva.taxa, "silva_species_assignment_v132.fa")
```

Upload metadata

Metadata, which can now be associated with sequencing data.

```
meta.16.20211008 <- read.csv("D:/Amplicon seq/Absolon16S/Absolon16S/20210902_16S_metadata.csv", header = TRUE)
head(meta.16.20211008)
```

Format metadata

Formatting metadata to fit the necessary structure of the phyloseq object is performed at this stage. This ensures that the row names of the metadata table and the sequencing table are identical.

```
# create a vector for sample names
s.vec <- as.vector(1:58)  #number should reflect your total number of samples
s.nam <- cbind("sample_", s.vec)
s.nam <- as.data.frame(s.nam)
s.names <- paste0(s.nam$V1, s.nam$s.vec)
s.names <- as.data.frame(s.names)

# apply sample names to metadata
row.names(meta.16.20211008) <- s.names$s.names
meta.16.20211008 <- as.data.frame(meta.16.20211008)

# apply sample names to sequence table
row.names(seqtab.chi) <- s.names$s.names
```

Construct phyloseq object

Construct a phyloseq object which is the main platform for downstream processing of sequencing data.

```
phy.20211008.16S <- phyloseq(otu_table(seqtab.chi, taxa_are_rows = FALSE), tax_table(silva.taxa), sample_data(meta.16.20211008))
```

Format ASV names

The sequence table currently displays the name of each ASV as a full sequence, making these elements messy and difficult to inspect. Full sequences can be replaced by simple numbered ASV names (e.g. “asv_1”).

```
# create vector for ASV names
dim(seqtab.chi)
a.vec <- as.vector(1:2931)  #number should reflect your total ASVs
a.nam <- cbind("asv_", a.vec)
a.nam <- as.data.frame(a.nam)
asv.names <- paste0(a.nam$V1, a.nam$a.vec)
asv.names <- as.data.frame(asv.names)

taxa_names(phy.20211008.16S) <- asv.names$asv.names
```

Reformat taxonomic data

This is an optional step, which reformats taxonomic data in a way that is easier to inspect and plot. This script adds a new column to the taxonomy table, which consists of the highest taxonomic classification in addition to the asv number (for example: *Prochlorococcus ASV_1*).

```
bc.t = t(as.data.frame(tax_table(phy.20211008.16S)))
bc.fill = na.locf(bc.t, na.rm = TRUE)
t.bc.fill = as.data.frame(t(bc.fill))
head(t.bc.fill)
rnc.bc = rownames_to_column(t.bc.fill, "ASV")

## Creates a column with the best classification and the ASV
rnc.bc$taxa_ASV = paste(rnc.bc$Genus,rnc.bc$ASV)

## Bind this column back onto the original tax_table
safe.bc = as.data.frame(tax_table(phy.20211008.16S))
safe.bc$taxa_ASV = paste(rnc.bc$taxa_ASV)
View(safe.bc)

# Setup object as tax_table
bc.tax = tax_table(safe.bc)
colnames(bc.tax) = colnames(safe.bc)
rownames(bc.tax) = rownames(safe.bc)
View(bc.tax)

## Update phyloseq object with new table
identical(bc.tax[1:2931,1:6], tax_table(phy.20211008.16S)) #should be true
tax_table(phy.20211008.16S) = bc.tax
head(tax_table(phy.20211008.16S))
```

Calculate the number of reads remaining in each sample

```
rowSums(otu_table(phy.20211008.16S))
mean(rowSums(otu_table(phy.20211008.16S)))
min(rowSums(otu_table(phy.20211008.16S)))
max(rowSums(otu_table(phy.20211008.16S)))
```

Exclude sequences

Sequences which are poorly assigned, or not of interest to the study can be removed at this stage. Here, we wish to retain only sequences of bacterial of archael origin, which are not classified as chloroplasts.

```
dim(tax_table(phy.20211008.16S)) #original number of ASV
phy.20211008.16S = subset_taxa(phy.20211008.16S, Kingdom=="Bacteria" | Kingdom=="Archaea")
dim(tax_table(phy.20211008.16S)) #number of ASVs after non-bacteria/archaea removed
phy.20211008.16S = subset_taxa(phy.20211008.16S, Order!="Chloroplast")
dim(tax_table(phy.20211008.16S)) #number of ASVs after chloroplasts removed
```

###Plot rarefaction curve Generate a rarefaction curve to ensure that sequence coverage is sufficient to represent the microbial diversity present in the sample. Prior to generating the curve we must definte the ggrare function.

```
rare.curve.16.20211008 = ggrare(phy.20211008.16S, step = 1000, se = FALSE) + theme_bw()
ggsave("D:/Amplicon seq/Absolon16S/Absolon16S/20211008_16S_rarefaction_curve.jpeg", rare.cu
rve.16.20211008, width = 15, height = 7.5, units = "cm", device = "jpeg")
```

Perform rarefaction

Subsample sequences to an even depth, which will be the minimum sequencing depth observed in retained samples; samples with very low sequencing depth may be discarded by choice, prior to this step. It is also important to ensure that the subsampling depth provides sufficient coverage of diversity, based on inspection of the previously generated rarefaction curves.

```
min(rowSums(otu_table(phy.20211008.16S))) #13574
set.seed(711) # reporducibility seed
phy.20211008.16S.rarefied = rarefy_even_depth(phy.20211008.16S, sample.size = 13574, trimOT
Us = TRUE)
dim(tax_table(phy.20211008.16S.rarefied))

saveRDS(phy.20211008.16S.rarefied, file = "D:/Amplicon seq/Absolon16S/Absolon16S/20211008_1
6S_rarefied_phyloseq_obj.RDS")
```

1.2 Downstream analysis of amplicon data

Stats and downstream analysis

function definitions

```
theme_ro <- function(){
  theme_bw()+
    theme(legend.position = "none",
          panel.grid.major = element_line(colour = "grey90", linetype = "dotted"),
          panel.grid.minor = element_blank(),
          strip.background = element_blank(),
          axis.text.x = element_text(size = 9),
          axis.text.y = element_text(size = 9),
          axis.title.x = element_text(size = 10),
          axis.title.y = element_text(size = 10))
}

theme_ro_legend <- function(){
  theme_bw()+
    theme(panel.grid.major = element_line(colour = "grey90", linetype = "dotted"),
          panel.grid.minor = element_blank(),
          strip.background = element_blank(),
          axis.text.x = element_text(size = 9),
          axis.text.y = element_text(size = 9),
          axis.title.x = element_text(size = 10),
          axis.title.y = element_text(size = 10))
}

split_proteo_silva <- function(physeq) {
  one = as.data.frame(tax_table(physeq))
  two = as.data.frame(lapply(one, as.character), stringsAsFactors = F)
  two$Phylum[two$Phylum=="Proteobacteria"] = two$Class[two$Phylum=="Proteobacteria"]
  two[] = lapply(two, factor)
  three = tax_table(two)
  rownames(three) = rownames(tax_table(physeq))
  colnames(three) = colnames(tax_table(physeq))
  tax_table(physeq) = three
  return(physeq)
}
```

```
## Preserve and rename the phyloseq object
phy.16S.rare <- phy.20211008.16S.rarefied
phy.16S.rare <- `20211008_16S_rarefied_phyloseq_obj`

## Drop the T0 samples
phy.16S.rare_sansT0 <- subset_samples(phy.16S.rare, condition != "T0")

## Re-Load metadata
meta.16 <- read.csv("D:/Amplicon seq/Absolon16S/Absolon16S/20210902_16S_metadata.csv", head
er = TRUE)
```

Alpha diversity plot and statistics including two-way ANOVA and Tukeys significance test

```
rich <- estimate_richness(phy.16S.rare_sansT0, split = TRUE, measures = "Shannon")

alpha <- cbind(rich, sample_data(phy.16S.rare_sansT0))
head(alpha)

## Two-way ANOVA to test significance between groups by type and condition

twowayanova <- aov(Shannon ~ type + condition, data = alpha)
summary(twowayanova)

## Tukey test to investigate the pairwise differences between all conditions

tukey <- TukeyHSD(x = twowayanova)
tukey
plot(tukey)

alpha_pot <- plot_richness(phy.16S.rare_sansT0, x="condition", measures = "Shannon", color
= "type")

ggsave("D:/Amplicon seq/Absolon16S/Absolon16S/20211008_16S_alpha_diversity_plot.jpeg", alph
a_pot, width = 15, height = 10, units = "cm", device = "jpeg")

svglite("F:/Amplicon seq/Absolon16S/Absolon16S/20220527_16S_alpha_diversity_plot-replot.svg
", width = 6, height = 3)
plot_richness(phy.16S.rare_sansT0, x="condition", measures = "Shannon", color = "type")+
  theme_ro()+
  theme_ro_legend()
dev.off()
```

PCoA For Beta diversity

```
bray.16.sansT0 <- vegdist(data.frame(otu_table(phy.16S.rare_sansT0)), method = "bray")
sample.data.pcoa_sansT0 <- sample_data(phy.16S.rare_sansT0)

add.16.sansT0 <-  !(is.euclid(bray.16.sansT0))

#use this to get the axis scores (Rel_corr_eig)
pcoa(bray.16.sansT0)$values

pcoa.16.sansT0 <- cmdscale(bray.16.sansT0, k = nrow(data.frame(otu_table(phy.16S.rare_sansT
0)))-1, eig = T, add = add.16.sansT0)

plot(pcoa.16.sansT0)

pcoa.bray.df_sansT0 <- as.data.frame(pcoa.16.sansT0$points[,1:44])
pcoa.bray.df_sansT0 <- cbind(pcoa.bray.df_sansT0, sample.data.pcoa_sansT0)
```

```
bray.pcoa.plot_sansT0 <- ggplot(pcoa.bray.df_sansT0, aes(x=V1, y=V2, shape = type, fill = c
ondition)) +
  geom_point(size = 2.5, alpha = 1, colour = "black", stroke = 0.2) +
  scale_shape_manual(values = c(21,23)) +
  scale_fill_brewer(palette = "Spectral") +
  xlab("PCoA1 (61.1%)") +
  ylab("PCoA2 (6.7%)") +
  theme_ro_legend() +
  guides(fill = guide_legend(override.aes = list(shape = 21)))

bray.pcoa.plot_sansT0

svglite("D:/Amplicon seq/Absolon16S/Absolon16S/20211213_16S_bray.pcoa.plot_sansT0.svg", wid
th = 6, height = 3)
bray.pcoa.plot_sansT0
dev.off()

#save_plot("D:/Amplicon seq/Absolon16S/Absolon16S/20211008_bray.pcoa.plot_sansT0.png", bray
.pcoa.plot_sansT0, base_height = 3, base_width = 6)
```

PCoA For Beta diversity - by type

```
forbray.DNA <- subset_samples(phy.16S.rare_sansT0, type == "DNA")
forbray.RNA <- subset_samples(phy.16S.rare_sansT0, type == "cDNA")

bray.16.sansT0.DNA <- vegdist(data.frame(otu_table(forbray.DNA)), method = "bray")
sample.data.pcoa_forbray.DNA <- sample_data(forbray.DNA)

add.16.forbray.DNA <-  !(is.euclid(bray.16.sansT0.DNA))

#use this to get the axis scores (Rel_corr_eig)
pcoa(bray.16.sansT0.DNA)$values

pcoa.16.sansT0.DNA <- cmdscale(bray.16.sansT0.DNA, k = nrow(data.frame(otu_table(forbray.DN
A)))-1, eig = T, add = add.16.forbray.DNA)

#plot(pcoa.16.sansT0)

pcoa.bray.df_sansT0.DNA <- as.data.frame(pcoa.16.sansT0.DNA$points[,1:20])
pcoa.bray.df_sansT0.DNA <- cbind(pcoa.bray.df_sansT0.DNA, sample.data.pcoa_forbray.DNA)

bray.pcoa.plot_sansT0.DNA <- ggplot(pcoa.bray.df_sansT0.DNA, aes(x=V1, y=V2, shape = type,
fill = condition)) +
  geom_point(size = 2.5, alpha = 1, colour = "black", stroke = 0.2) +
  scale_shape_manual(values = c(21,23)) +
  scale_fill_brewer(palette = "Spectral") +
  xlab("PCoA1 (48.5%)") +
  ylab("PCoA2 (10.9%)") +
  theme_ro_legend() +
  guides(fill = guide_legend(override.aes = list(shape = 21)))

bray.pcoa.plot_sansT0.DNA

svglite("D:/Amplicon seq/Absolon16S/Absolon16S/20211213_16S_bray.pcoa.plot_sansT0-DNA.svg",
width = 6, height = 3)
bray.pcoa.plot_sansT0.DNA
dev.off()

############## RNA

bray.16.sansT0.RNA <- vegdist(data.frame(otu_table(forbray.RNA)), method = "bray")
sample.data.pcoa_forbray.RNA <- sample_data(forbray.RNA)
```

```
add.16.forbray.RNA <-  !(is.euclid(bray.16.sansT0.RNA))

#use this to get the axis scores (Rel_corr_eig)
pcoa(bray.16.sansT0.RNA)$values

pcoa.16.sansT0.RNA <- cmdscale(bray.16.sansT0.RNA, k = nrow(data.frame(otu_table(forbray.RN
A)))-1, eig = T, add = add.16.forbray.RNA)

#plot(pcoa.16.sansT0)

pcoa.bray.df_sansT0.RNA <- as.data.frame(pcoa.16.sansT0.RNA$points[,1:20])
pcoa.bray.df_sansT0.RNA <- cbind(pcoa.bray.df_sansT0.RNA, sample.data.pcoa_forbray.RNA)

bray.pcoa.plot_sansT0.RNA <- ggplot(pcoa.bray.df_sansT0.RNA, aes(x=V1, y=V2, shape = type,
fill = condition)) +
  geom_point(size = 2.5, alpha = 1, colour = "black", stroke = 0.2) +
  scale_shape_manual(values = c(21,23)) +
  scale_fill_brewer(palette = "Spectral") +
  xlab("PCoA1 (28.5%)") +
  ylab("PCoA2 (16.6%)") +
  theme_ro_legend() +
  guides(fill = guide_legend(override.aes = list(shape = 21)))

bray.pcoa.plot_sansT0.RNA

svglite("D:/Amplicon seq/Absolon16S/Absolon16S/20211213_16S_bray.pcoa.plot_sansT0-RNA.svg",
width = 6, height = 3)
bray.pcoa.plot_sansT0.RNA
dev.off()

save_plot("D:/Amplicon seq/Absolon16S/Absolon16S/20211008_bray.pcoa.plot_sansT0-DNA.png", b
ray.pcoa.plot_sansT0.DNA, base_height = 3, base_width = 6)
save_plot("D:/Amplicon seq/Absolon16S/Absolon16S/20211008_bray.pcoa.plot_sansT0-RNA.png", b
ray.pcoa.plot_sansT0.RNA, base_height = 3, base_width = 6)
```

PERMANOVA

```
# PERMANOVA ------------------------------------------------------------


metadata.16 <- as(sample_data(phy.16S.rare_sansT0), "data.frame")
bray.16.sansT0

adonis(bray.16.sansT0 ~ condition*type, data = metadata.16, permutations = 999)

# Homogeneity of multivariate dispersions ------------------------------
beta.16.sansT0 <- betadisper(bray.16.sansT0, metadata.16$condition, type = "centroid", sqrt
.dist = FALSE)
permutest(beta.16.sansT0, pairwise = T, permutations = 999)
```

ad-hoc - pairwise beta diversity permanovas

```
# PERMANOVA ------------------------------------------------------------

cond.1.1_1.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "1.2"))
cond.1.1_1.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "1.3"))
cond.1.1_1.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "1.4"))
cond.1.1_1.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "1.5"))
cond.1.1_1.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "1.6"))
cond.1.1_2.1 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "2.1"))
```

```
cond.1.1_2.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "2.2"))
cond.1.1_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "2.3"))
cond.1.1_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "2.4"))
cond.1.1_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "2.5"))
cond.1.1_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.1", "2.6"))

cond.1.2_1.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "1.3"))
cond.1.2_1.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "1.4"))
cond.1.2_1.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "1.5"))
cond.1.2_1.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "1.6"))
cond.1.2_2.1 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "2.1"))
cond.1.2_2.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "2.2"))
cond.1.2_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "2.3"))
cond.1.2_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "2.4"))
cond.1.2_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "2.5"))
cond.1.2_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.2", "2.6"))

cond.1.3_1.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "1.4"))
cond.1.3_1.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "1.5"))
cond.1.3_1.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "1.6"))
cond.1.3_2.1 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "2.1"))
cond.1.3_2.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "2.2"))
cond.1.3_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "2.3"))
cond.1.3_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "2.4"))
cond.1.3_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "2.5"))
cond.1.3_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.3", "2.6"))

cond.1.4_1.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "1.5"))
cond.1.4_1.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "1.6"))
cond.1.4_2.1 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "2.1"))
cond.1.4_2.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "2.2"))
cond.1.4_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "2.3"))
cond.1.4_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "2.4"))
cond.1.4_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "2.5"))
cond.1.4_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.4", "2.6"))

cond.1.5_1.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "1.6"))
cond.1.5_2.1 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.1"))
cond.1.5_2.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.2"))
cond.1.5_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.3"))
cond.1.5_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.4"))
cond.1.5_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.5"))
cond.1.5_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.6"))

cond.1.6_2.1 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.1"))
cond.1.6_2.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.2"))
cond.1.6_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.3"))
cond.1.6_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.4"))
cond.1.6_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.5"))
cond.1.6_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("1.5", "2.6"))

cond.2.1_2.2 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.1", "2.2"))
cond.2.1_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.1", "2.3"))
cond.2.1_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.1", "2.4"))
cond.2.1_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.1", "2.5"))
cond.2.1_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.1", "2.6"))

cond.2.2_2.3 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.2", "2.3"))
cond.2.2_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.2", "2.4"))
cond.2.2_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.2", "2.5"))
cond.2.2_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.2", "2.6"))

cond.2.3_2.4 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.3", "2.4"))
cond.2.3_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.3", "2.5"))
cond.2.3_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.3", "2.6"))
```

```
cond.2.4_2.5 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.4", "2.5"))
cond.2.4_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.4", "2.6"))

cond.2.5_2.6 <- subset_samples(phy.16S.rare_sansT0, condition.no. %in% c("2.5", "2.6"))


#empty vector to store the results of the pair-wise permanova tests
cond_list <- c(cond.1.1_1.2, cond.1.1_1.3, cond.1.1_1.4, cond.1.1_1.5, cond.1.1_1.6, cond.1
.1_2.1, cond.1.1_2.2, cond.1.1_2.3, cond.1.1_2.4, cond.1.1_2.5, cond.1.1_2.6,
               cond.1.2_1.3, cond.1.2_1.4, cond.1.2_1.5, cond.1.2_1.6, cond.1.2_2.1, cond.1
.2_2.2, cond.1.2_2.3, cond.1.2_2.4, cond.1.2_2.5, cond.1.2_2.6,
               cond.1.3_1.4, cond.1.3_1.5, cond.1.3_1.6, cond.1.3_2.1, cond.1.3_2.2, cond.1
.3_2.3, cond.1.3_2.4, cond.1.3_2.5, cond.1.3_2.6,
               cond.1.4_1.5, cond.1.4_1.6, cond.1.4_2.1, cond.1.4_2.2, cond.1.4_2.3, cond.1
.4_2.4, cond.1.4_2.5, cond.1.4_2.6,
               cond.1.5_1.6, cond.1.5_2.1, cond.1.5_2.2, cond.1.5_2.3, cond.1.5_2.4, cond.1
.5_2.5, cond.1.5_2.6,
               cond.1.6_2.1, cond.1.6_2.2, cond.1.6_2.3, cond.1.6_2.4, cond.1.6_2.5, cond.1
.6_2.6,
               cond.2.1_2.2, cond.2.1_2.3, cond.2.1_2.4, cond.2.1_2.5, cond.2.1_2.6,
               cond.2.2_2.3, cond.2.2_2.4, cond.2.2_2.5, cond.2.2_2.6,
               cond.2.3_2.4, cond.2.3_2.5, cond.2.3_2.6,
               cond.2.4_2.5, cond.2.4_2.6,
               cond.2.5_2.6
               )
#Function to run the permanova
run_permanova <- function(X){
  bray <- vegdist(data.frame(otu_table(X)), method = "bray")
metadata <- as(sample_data(X), "data.frame")
metadata$condition.no. <- as.factor(metadata$condition.no.)
head(metadata)
adonis(bray ~ condition.no., data = metadata, permutations = 999)
}

lapply(cond_list, run_permanova)
```

Bar chart

```
rare.split.16_sansT0 <- split_proteo_silva(phy.16S.rare_sansT0) #this breaks out proteobact
eria to the class level (by replacing the phylum column for proteobacteria with the class c
olumn)

phy.glom.16_sansT0 <- tax_glom(rare.split.16_sansT0, taxrank = "Phylum")

colSums(otu_table(phy.glom.16_sansT0))
tax_table(phy.glom.16_sansT0)
## Group together 'rare' phyla, so that my barplot only has ~10-12 colours. I also group al
l the proteobacteria together by prefacing them with 'aa' this just helps ggplot put them t
ogether (likewise with archaea being z)
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Epsilonbacteraeota"] <- "zz
Other"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Euryarchaeota"] <- "zzOther
"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Firmicutes"] <- "zzOther"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Acidobacteria"] <- "zzOther
"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Hydrogenedentes"] <- "zzOth
er"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Chlamydiae"] <- "zzOther"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Fibrobacteres"] <- "zzOther
"

tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Alphaproteobacteria"] <- "a
```

```
aAlphaproteobacteria"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Deltaproteobacteria"] <- "a
aDeltaproteobacteria"
tax_table(phy.glom.16_sansT0)[tax_table(phy.glom.16_sansT0) == "Gammaproteobacteria"] <- "a
aGammaproteobacteria"

phy.glom.16_sansT0 <- tax_glom(phy.glom.16_sansT0, taxrank="Phylum") # glom again with new
phylum names
tax_table(phy.glom.16_sansT0)

phylum.palette <- c("#C7EAFF", "#368DFF", "#00268F", "#FFE0E0", "#FF8585", "#318C48", "#9EF
AAA", "#FFADE1", "#FF459C", "#FAF49B", "#FFC936", "grey80")

### scale_y_continuous numbers are the number of read rarfied to

phy.bar.16_sansT0 <- plot_bar(phy.glom.16_sansT0, fill = "Phylum") + facet_grid(rows = vars
(condition), cols = vars(type), scales = "free", space = "free") +
  scale_y_continuous(expand = c(0,0),limits = c(0,13574), breaks = c(0, 3389, 6787, 10185,
13574), labels = c("0%","25%", "50%", "75%", "100%")) +
  theme(legend.text=element_text(size=10), legend.title= element_blank(), legend.background
= element_rect(color="grey50", size=.3, linetype=1), strip.background = element_blank()) +
  scale_fill_manual(values = phylum.palette, c("Alphaproteobacteria", "Deltaproteobacteria"
, "Gammaproteobacteria", "Acidobacteria","Actinobacteria", "Bacteroidetes","Chloroflexi", "
Cyanobacteria", "Lentisphaerae", "Planctomycetes", "Verrucomicrobia", "Other")) +
  scale_x_discrete(expand = c(0,0)) +
  ylab("") +
  xlab("") +
  geom_bar(stat = "identity", width = 1, size = 0.2, colour = "gray20") +
  theme_ro_legend() +
  theme(axis.text.x = element_text(size = 8, hjust = 0.5, vjust = 0.5)) +
  theme(axis.text.y = element_text(size = 8, hjust = 0), axis.title.y = element_text(size =
10)) +
  theme(legend.text = element_text(size=8), legend.title = element_blank()) +
  theme(strip.text.y = element_text(size = 10, angle = 0), strip.text.x = element_text(size
= 10)) +
  theme(panel.spacing.x = unit(1.75, "lines")) +
  theme(legend.position =  "bottom", legend.spacing.x = unit(0.5, 'lines')) +
  coord_flip()

save_plot("D:/Amplicon seq/Absolon16S/Absolon16S/20101008_16S_bar_sansT0.png", phy.bar.16_s
ansT0, base_width = 13, base_height = 9)

svglite("D:/Amplicon seq/Absolon16S/Absolon16S/20211213_16S_phy_bar_sansT0.svg", width = 7,
height = 7)
phy.bar.16_sansT0
dev.off()
```

Ratio work - glom at Family level - experimental

```
# Phyloseq subdivision ----------------------------------------------------
# Spliting the phyloseq object into cDNA and DNA
# First removing samples with no matching pairs

phy.16S.ratio <- subset_samples(phy.16S.rare_sansT0, bottle_number !="1")
phy.16S.ratio <- subset_samples(phy.16S.ratio, bottle_number !="14")

phy.fam.16_sansT0.ratio <- tax_glom(phy.16S.ratio, taxrank = "Family")

dna.16.fam2 <- subset_samples(phy.fam.16_sansT0.ratio, type == "DNA")
rna.16.fam2 <- subset_samples(phy.fam.16_sansT0.ratio, type == "cDNA")

sample_names(dna.16.fam2) <- sample_data(dna.16.fam2)$bottle_number
sample_names(rna.16.fam2) <- sample_data(rna.16.fam2)$bottle_number
```

```
# Pull out OTU tables as data frames

df.dna.fam2 <- data.frame(otu_table(dna.16.fam2))
df.rna.fam2 <- data.frame(otu_table(rna.16.fam2))

# Manually check the rows match

df.dna.fam2[1:10, 1:10]
df.rna.fam2[1:10, 1:10]

#### Extract asv names

fam_names <- phy.fam.16_sansT0.ratio@tax_table

## calculate a dataframe containing the sum of dna and rna reads
df.sum.fam2 <- df.dna.fam2+df.rna.fam2

## calculate a dataframe which is the ratio between dna and rna reads (note I'm assuming co
rresponding rows match here)
df.dna2.fam2 <- df.dna.fam2 # make a new df.dna object as we are going to manipulate this

df.dna2.fam2[df.dna2.fam2 == 0 & df.rna.fam2>0] <- 1 # reassigned dna 0 values as 1 IF rna
value is >0 (i.e. if RNA is present we assume DNA should also be present)

df.ratio.fam2 <- data.frame(df.rna.fam2/df.dna2.fam2) # calculate ratio

df.ratio.fam2[df.ratio.fam2 == "NaN"] <- 0 # some values will be NaN (dividing when rna = 0
) - these should be replaced with 0.

## Now you have a 4 key dataframes (dna, rna, sum of dna and rna reads, ratio of dna rna re
ads). When you are working with your data, all your rows should match - that is to say, row
1 in each data.frame will correspond bottle_1 (from which both dna and rna are extracted).
Consequently, you can give the rows of your respective dataframes the same name... (this wi
ll be important for plotting).


#sample_codes <- vector()
#for (i in 1:23){
#  sample_codes[i] <- paste0("bottle",i)
#}
#sample_codes

sample_codes2 <- c("T24_Control_Bottle_02",  "T24_Control_Bottle_03",  "T24_Control_Bottle_
04",  "B12_06",  "B12_07",  "B12_08",  "B12_09", "B12_Nut_11", "B12_Nut_12", "B12_Nut_13",
"BVit_mix_16", "BVit_mix_17", "BVit_mix_18", "BVit_mix_19", "BVit_Nut_21", "BVit_Nut_22", "
BVit_Nut_23", "BVit_Nut_24", "Nut_26", "Nut_27", "Nut_28", "Nut_29", "Nut_30")

rownames(df.dna.fam2) <- sample_codes2
rownames(df.rna.fam2) <- sample_codes2
rownames(df.sum.fam2) <- sample_codes2
rownames(df.ratio.fam2) <- sample_codes2

## now we have an identifier for each bottle, we can manipulate these dataframes into a sin
gle object for plotting. This is where the data-wrangling gets a little hairy. We will use
a package called reshape2 and a function called 'melt' to do this. If you're interested, yo
u could rerun this code with a subset (e.g. [1:5,1:5] of each dataframe) to see the nuts an
d bolts.

df.dna.fam2$bottle <- rownames(df.dna.fam2) # this adds a column to the dataframe containin
g the bottle identifier (which was previously assigned to row names)
df.dna.melt.fam2 <- melt(df.dna.fam2, id = "bottle") # this decomposes the dataframe to mak
e asv name a column,  read count a column, and bottle a column
colnames(df.dna.melt.fam2) <- c("bottle", "asv", "dna") # renaming the columns to reflect t
heir contents
```

```
df.rna.fam2$bottle <- rownames(df.rna.fam2)
df.rna.melt.fam2 <- melt(df.rna.fam2, id = "bottle")
colnames(df.rna.melt.fam2) <- c("bottle", "asv", "rna")

df.sum.fam2$bottle <- rownames(df.sum.fam2)
df.sum.melt.fam2 <- melt(df.sum.fam2, id = "bottle")
colnames(df.sum.melt.fam2) <- c("bottle", "asv", "sum")

df.ratio.fam2$bottle <- rownames(df.ratio.fam2)
df.ratio.melt.fam2 <- melt(df.ratio.fam2, id = "bottle")
colnames(df.ratio.melt.fam2) <- c("bottle", "asv", "ratio")

df.complete.fam2 <- data.frame(cbind(df.dna.melt.fam2$bottle, as.character(df.dna.melt.fam2
$asv), df.dna.melt.fam2$dna, df.rna.melt.fam2$rna, df.sum.melt.fam2$sum, df.ratio.melt.fam2
$ratio))
colnames(df.complete.fam2) <- c("bottle", "asv", "dna", "rna","sum", "ratio") # tidy up col
umn names

df.complete.fam2[1:10,1:6] # have a look at the dataframe

df.complete.fam2 <- as.data.frame(df.complete.fam2)
df.slice.fam2 <- subset(df.complete.fam2, asv %in% c("asv_1", "asv_2", "asv_3", "asv_4", "a
sv_8", "asv_11", "asv_13", "asv_16", "asv_18", "asv_22"))

df.slice.fam2$sum <- as.numeric(as.character(df.slice.fam2$sum))
#df.slice$asv <- as.numeric(as.character(df.slice$asv))
df.slice.fam2$ratio <- as.numeric(as.character(df.slice.fam2$ratio))

df.slice.fam2$asv <- factor(df.slice.fam2$asv, levels = rev(c("asv_1", "asv_2", "asv_3", "a
sv_4", "asv_8", "asv_11", "asv_13", "asv_16", "asv_18", "asv_22")))

library(colorspace)
library(pals)

t4 <- colorRampPalette(c("khaki1", "deepskyblue1", "royalblue4"))
t5 <- colorRampPalette(c("khaki1", "firebrick1", "violetred4"), bias = 3)
t6 <- cbind(rev(t4(300)),t5(300))
t7 <- colorRampPalette(t6)

balloon_fam2_asv_ordered_top10 <- ggplot(df.slice.fam2, aes(x = bottle, y = asv, size = sqr
t(sum), fill = ratio)) +
  geom_point(shape = 21, colour = "black", stroke = 0.1) +
  scale_size_continuous(breaks = c(0, sqrt(0.01*(rowSums(otu_table(phy.16S.ratio)*2))[1]),
sqrt(0.05*(rowSums(otu_table(phy.16S.ratio)*2))[1]), sqrt(0.1*(rowSums(otu_table(phy.16S.ra
tio)*2))[1])), labels = c("0%", "1%", "5%", "10%")) +
  xlab("") +
  ylab("") +
  scale_fill_gradientn(colours = t7(35),  values = rescale(c(0,1,10)), na.value = "violetre
d4", limits = c(0,10), breaks = c(0,1,5,10,15,20), guide=guide_colorbar(title="16S Ratio",
title.position = "top", frame.colour = "black", ticks.colour = "black", frame.linewidth = 0
.5)) + # here I'm defining the palette and setting an NA value of violetred4 - this will be
the colour of any points with a ratio >10
  theme_ro_legend() +
  theme(panel.spacing.y = unit(1, "lines"), panel.spacing.x = unit(0.5, "lines")) +
  guides(size = guide_legend(title = "Rel. Abundance", title.position = "top")) +
  theme(axis.text.y = element_text(size = 6.5)) +
  theme(axis.text.x = element_text(angle = 90, size = 6.5, hjust = 1, vjust = 0.5)) +
  theme(legend.position = "none", legend.title = element_text(size = 7.5), legend.text = el
ement_text(size = 7.5),panel.spacing=unit(0,"lines"), panel.grid.major = element_line(colou
r = "grey90", size = 0.1, linetype = 1))


tax_table(dna.16.fam2)[1:10,5] # this is a list of family names (note asv_9 is NA)

balloon.manual <- ggplot(df.slice.fam2, aes(x = bottle, y = asv, size = sqrt(sum), fill = r
atio)) +
```

```
  geom_point(shape = 21, colour = "black", stroke = 0.1) +
  scale_size_continuous(breaks = c(0, sqrt(0.01*(rowSums(otu_table(phy.16S.ratio)*2))[1]), sqrt(0.05*(rowSums(otu_table(phy.16S.ratio)*2))[1]), sqrt(0.1*(rowSums(otu_table(phy.16S.ratio)*2))[1])), labels = c("0%", "1%", "5%", "10%")) +
  xlab("") +
  ylab("") +
  scale_fill_gradientn(colours = t7(35),  values = rescale(c(0,1,10)), na.value = "violetred4", limits = c(0,10), breaks = c(0,1,5,10,15,20), guide=guide_colorbar(title="16S Ratio", title.position = "top", frame.colour = "black", ticks.colour = "black", frame.linewidth = 0.5)) + # here I'm defining the palette and setting an NA value of violetred4 - this will be the colour of any points with a ratio >10
  scale_y_discrete(label = rev(tax_table(dna.16.fam2)[1:10,5])) +
  theme_ro_legend() +
  theme(panel.spacing.y = unit(1, "lines"), panel.spacing.x = unit(0.5, "lines")) +
  guides(size = guide_legend(title = "Rel. Abundance", title.position = "top")) +
  theme(axis.text.y = element_text(size = 6.5)) +
  theme(axis.text.x = element_text(angle = 90, size = 6.5, hjust = 1, vjust = 0.5)) +
  theme(legend.position = "none", legend.title = element_text(size = 7.5), legend.text = element_text(size = 7.5),panel.spacing=unit(0,"lines"), panel.grid.major = element_line(colour = "grey90", size = 0.1, linetype = 1))

save_plot("D:/Amplicon seq/Absolon16S/Absolon16S/20211008_balloon_fam2_top10_named.png", balloon.manual + theme(legend.position = "bottom"), base_height = 5, base_width = 6)

grouped_df.slice.fam <- df.slice.fam2 %>% group_by(cond)

ratio_box <- ggplot(grouped_sf.slice.gen, aes(x = cond, y = ratio, fill = asv))+
  geom_boxplot()+
  theme_ro()+
  facet_wrap(grouped_sf.slice.gen$asv, scales = "free")+
  scale_y_continuous(limits = c(0, 50))+
  theme(axis.text.y = element_text(size = 10))

svglite("D:/Amplicon seq/Absolon16S/Absolon16S/20211213_ratio_box_plot_16S.svg", width = 11, height = 7)
ratio_box
dev.off()
```

Ratio plots for genus level and ASV level as above but 'glom' at appropriate taxonomic rank

```
phy.16S.rare_sansT0.top100 <- prune_taxa(names(sort(taxa_sums(phy.16S.rare_sansT0),TRUE)[1:100]), phy.16S.rare_sansT0)

plot_heatmap(phy.16S.rare_sansT0.top100, low="#000033", high="#CCFF66")+facet_grid(, cols = vars(type))

phy.16S.relativeabun  = transform_sample_counts(phy.16S.rare_sansT0.top100, function(x) x / sum(x))

phy.16S.pcntabun  = transform_sample_counts(phy.glom.16_sansT0, function(x) x / sum(x)*100)

prcntabun <- data.frame(otu_table(phy.16S.pcntabun))

colnames(prcntabun) <- tax_table(phy.16S.pcntabun)[,2:2]

sampledf <- data.frame(sample_data(phy.16S.pcntabun)$condition.no.)
sampledf2 <- data.frame(sample_data(phy.16S.pcntabun)$type)

prcntabun <- cbind(prcntabun, sampledf, sampledf2)

as.factor(prcntabun$sample_data.phy.16S.pcntabun..condition.no.)
```

```
dna.prcntabun <- subset(prcntabun, sample_data.phy.16S.pcntabun..type == "DNA")
rna.prcntabun <- subset(prcntabun, sample_data.phy.16S.pcntabun..type == "cDNA")

dna.prcntabun <- dna.prcntabun[,1:13]
rna.prcntabun <- rna.prcntabun[,1:13]

write.csv(dna.prcntabun,"D:/Amplicon seq/Absolon16S/Absolon16S/20211027_DNA_percent_abundan
ces.csv")

write.csv(rna.prcntabun,"D:/Amplicon seq/Absolon16S/Absolon16S/20211027_cDNA_percent_abunda
nces.csv")


boxplot(dna.prcntabun)


tax_table(phy.16S.pcntabun)[,1:2]
otu_table(phy.16S.pcntabun)[1:20]
```

1.3 - Full pairwise PERMANOVA p-value results for every condition-type combination of 16S samples

Full pairwise PERMANOVA p-value results for every condition-type combination of 16S samples. This test allows the identification of condition-types that are significantly different in beta-diversity compared to all other condition-types. Condition-type refers to the combination of the type of sample (rRNA or rDNA) and the amendment conditions. For full experimental set up see Chapter 2 Section 2.1.2 and this data is referred to in Section 2.2.1.2. Those p-values in bold are below 0.05 significance threshold.

Condition-type 1	Condition-type 2	p-value
rRNA Control	rRNA B_{12}	0.859
rRNA Control	rRNA B_{12} + Nutrients	0.539
rRNA Control	rRNA B vitamin mix	**0.03**
rRNA Control	rRNA B vitamin mix + Nutrients	0.086
rRNA Control	rRNA Nutrients	0.568
rRNA Control	rDNA Control	**0.027**
rRNA Control	rDNA B_{12}	**0.029**
rRNA Control	rDNA B_{12} + Nutrients	**0.039**
rRNA Control	rDNA B vitamin mix	**0.034**
rRNA Control	rDNA B vitamin mix + Nutrients	**0.029**
rRNA Control	rDNA Nutrients	**0.008**
rRNA B_{12}	rRNA B_{12} + Nutrients	0.081
rRNA B_{12}	rRNA B vitamin mix	**0.023**
rRNA B_{12}	rRNA B vitamin mix + Nutrients	**0.038**
rRNA B_{12}	rRNA Nutrients	0.119
rRNA B_{12}	rDNA Control	**0.03**
rRNA B_{12}	rDNA B_{12}	**0.034**
rRNA B_{12}	rDNA B_{12} + Nutrients	**0.032**
rRNA B_{12}	rDNA B vitamin mix	**0.028**
rRNA B_{12}	rDNA B vitamin mix + Nutrients	**0.03**
rRNA B_{12}	rDNA Nutrients	**0.009**
rRNA B_{12} + Nutrients	rRNA B vitamin mix	**0.033**
rRNA B_{12} + Nutrients	rRNA B vitamin mix + Nutrients	**0.021**
rRNA B_{12} + Nutrients	rRNA Nutrients	0.625
rRNA B_{12} + Nutrients	rDNA Control	0.1
rRNA B_{12} + Nutrients	rDNA B_{12}	**0.033**
rRNA B_{12} + Nutrients	rDNA B_{12} + Nutrients	**0.027**
rRNA B_{12} + Nutrients	rDNA B vitamin mix	**0.025**
rRNA B_{12} + Nutrients	rDNA B vitamin mix + Nutrients	**0.037**
rRNA B_{12} + Nutrients	rDNA Nutrients	**0.015**
rRNA B vitamin mix	rRNA +B vitamin mix + Nutrients	0.071
rRNA B vitamin mix	rRNA Nutrients	**0.006**
rRNA B vitamin mix	rDNA Control	**0.04**
rRNA B vitamin mix	rDNA B_{12}	**0.037**
rRNA B vitamin mix	rDNA B_{12} + Nutrients	**0.025**
rRNA B vitamin mix	rDNA B vitamin mix	**0.027**
rRNA B vitamin mix	rDNA B vitamin mix + Nutrients	**0.028**
rRNA B vitamin mix	rDNA Nutrients	**0.009**
rRNA B vitamin mix + Nutrients	rRNA Nutrients	**0.006**
rRNA B vitamin mix + Nutrients	rDNA Control	**0.035**

rRNA B vitamin mix + Nutrients	rDNA B_{12}	**0.028**
rRNA B vitamin mix + Nutrients	rDNA B_{12} + Nutrients	**0.034**
rRNA B vitamin mix + Nutrients	rDNA B vitamin mix	**0.028**
rRNA B vitamin mix + Nutrients	rDNA B vitamin mix + Nutrients	**0.03**
rRNA B vitamin mix + Nutrients	rDNA Nutrients	**0.005**
rRNA Nutrients	rDNA Control	**0.03**
rRNA Nutrients	rDNA B_{12}	**0.034**
rRNA Nutrients	rDNA B_{12} + Nutrients	**0.031**
rRNA Nutrients	rDNA B vitamin mix	**0.026**
rRNA Nutrients	rDNA B vitamin mix + Nutrients	**0.027**
rRNA Nutrients	rDNA Nutrients	**0.009**
rDNA Control	rDNA B_{12}	**0.028**
rDNA Control	rDNA B_{12} + Nutrients	0.466
rDNA Control	rDNA B vitamin mix	**0.034**
rDNA Control	rDNA B vitamin mix + Nutrients	**0.025**
rDNA Control	rDNA Nutrients	**0.023**
rDNA B_{12}	rDNA B_{12} + Nutrients	0.273
rDNA B_{12}	rDNA B vitamin mix	**0.02**
rDNA B_{12}	rDNA B vitamin mix + Nutrients	**0.035**
rDNA B_{12}	rDNA Nutrients	0.571
rDNA B_{12} + Nutrients	rDNA B vitamin mix	0.177
rDNA B_{12} + Nutrients	rDNA B vitamin mix + Nutrients	**0.045**
rDNA B_{12} + Nutrients	rDNA Nutrients	0.126
rDNA B vitamin mix	rDNA B vitamin mix + Nutrients	0.057
rDNA B vitamin mix	rDNA Nutrients	**0.019**
rDNA B vitamin mix + Nutrients	rDNA Nutrients	**0.023**

1.4 Full list of prokaryotic genera predicted from all 16S samples

Full list of prokaryotic genera predicted from all 16S samples. Note that ASV number refers to the ASV number of the first species detected in that genera. All others are collapsed into that ASV number. For full description of taxonomic predictions see Chapter 2 Section 2.2.1. This data is referred to in Chapter 2 Section 2.3.2.2.

	Phylum	Class	Order	Family	Genus
ASV 1	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aurantivirga
ASV 2	Proteobacteria	Alphaproteobacteria	SAR11_clade	Clade_I	Clade_Ia
ASV 3	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae	OM60(NOR5)_clade
ASV 4	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Ascidiaceihabitans
ASV 5	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Formosa
ASV 6	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Amylibacter
ASV 11	Cyanobacteria	Oxyphotobacteria	Synechococcales	Cyanobiaceae	Synechococcus_CC9902
ASV 12	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Planktomarina
ASV 13	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Alteromonas
ASV 15	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	HIMB11
ASV 22	Proteobacteria	Alphaproteobacteria	Caulobacterales	Hyphomonadaceae	Hellea
ASV 24	Verrucomicrobia	Verrucomicrobiae	Opitutales	Puniceicoccaceae	Coraliomargarita
ASV 25	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	NS4_marine_group
ASV 30	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Spongiibacteraceae	Litorivivens
ASV 32	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Porticoccaceae	SAR92_clade
ASV 34	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae	Luminiphilus
ASV 38	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter_4
ASV 43	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter
ASV 44	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Rubritaleaceae	Roseibacillus
ASV 46	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter_2
ASV 48	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Ulvibacter
ASV 50	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Aliiglaciecola
ASV 59	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Vibrio
ASV 63	Bacteroidetes	Bacteroidia	Cytophagales	Cyclobacteriaceae	Marinoscillum
ASV 64	Proteobacteria	Gammaproteobacteria	Alteromonadales	Pseudoalteromonadaceae	Pseudoalteromonas
ASV 65	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Methylophilaceae	OM43_clade
ASV 68	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	NS5_marine_group

ASV 72	Proteobacteria	Gammaproteobacteria	Thiomicrospirales	Thioglobaceae	SUP05_cluster
ASV 81	Proteobacteria	Alphaproteobacteria	Puniceispirillales	SAR116_clade	Candidatus_Puniceispirillum
ASV 88	Verrucomicrobia	Verrucomicrobiae	Opitutales	Puniceicoccaceae	MB11C04_marine_group
ASV 91	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aureicoccus
ASV 97	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flaviramulus
ASV 101	Verrucomicrobia	Verrucomicrobiae	Pedosphaerales	Pedosphaeraceae	SCGC_AAA164-E04
ASV 104	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Nereida
ASV 110	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Tateyamaria
ASV 112	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Glaciecola
ASV 113	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	NS2b_marine_group
ASV 130	Proteobacteria	Alphaproteobacteria	Thalassobaculales	Nisaeaceae	OM75_clade
ASV 131	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Roseobacter_clade_NAC11-7_lineage
ASV 142	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Litoricolaceae	Litoricola
ASV 143	Lentisphaerae	Lentisphaeria	Lentisphaerales	Lentisphaeraceae	Lentisphaera
ASV 148	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Nitrosomonadaceae	IS-44
ASV 154	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Spongiibacteraceae	BD1-7_clade
ASV 155	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Photobacterium
ASV 159	Cyanobacteria	Oxyphotobacteria	Nostocales	Microcystaceae	Atelocyanobacterium_(UCYN-A)
ASV 161	Verrucomicrobia	Verrucomicrobiae	Opitutales	Puniceicoccaceae	Lentimonas
ASV 166	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flavicella
ASV 168	Bacteroidetes	Bacteroidia	Flavobacteriales	Crocinitomicaceae	Fluviicola
ASV 204	Planctomycetes	Planctomycetacia	Pirellulales	Pirellulaceae	Blastopirellula
ASV 219	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Loktanella
ASV 232	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Woeseiaceae	Woeseia
ASV 245	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Pseudohongiellaceae	Pseudohongiella
ASV 250	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aquibacter
ASV 263	Proteobacteria	Gammaproteobacteria	Alteromonadales	Marinobacteraceae	Marinobacter
ASV 269	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Aliiroseovarius
ASV 304	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	Halomonas
ASV 324	Planctomycetes	Phycisphaerae	Phycisphaerales	Phycisphaeraceae	CL500-3

ASV 334	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae	Thalassotalea
ASV 346	Verrucomicrobia	Verrucomicrobiae	Opitutales	Puniceicoccaceae	Pelagicoccus
ASV 349	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Thalassobius
ASV 358	Proteobacteria	Gammaproteobacteria	Coxiellales	Coxiellaceae	Coxiella
ASV 378	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Octadecabacter
ASV 444	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bdellovibrionaceae	OM27_clade
ASV 454	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Pseudooceanicola
ASV 460	Actinobacteria	Acidimicrobiia	Actinomarinales	Actinomarinaceae	Candidatus_Actinomarina
ASV 467	Acidobacteria	Thermoanaerobaculia	Thermoanaerobaculales	Thermoanaerobaculaceae	Subgroup_10
ASV 486	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Porticoccaceae	Porticoccus
ASV 495	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Pseudobowmanella
ASV 506	Bacteroidetes	Bacteroidia	Chitinophagales	Saprospiraceae	Aureispira
ASV 513	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Roseovarius
ASV 515	Planctomycetes	Planctomycetacia	Pirellulales	Pirellulaceae	Rubripirellula
ASV 527	Proteobacteria	Alphaproteobacteria	Rickettsiales	Midichloriaceae	MD3-55
ASV 533	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Jannaschia
ASV 535	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Sulfitobacter
ASV 536	Planctomycetes	Phycisphaerae	Phycisphaerales	Phycisphaeraceae	FS140-16B-02_marine_group
ASV 546	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Roseobacter
ASV 552	Proteobacteria	Alphaproteobacteria	Caulobacterales	Hyphomonadaceae	Maricaulis
ASV 561	Actinobacteria	Acidimicrobiia	Microtrichales	Ilumatobacteraceae	Ilumatobacter
ASV 563	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter_3
ASV 569	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Aestuariibacter
ASV 571	Verrucomicrobia	Verrucomicrobiae	Opitutales	Puniceicoccaceae	Cerasicoccus
ASV 585	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	NS3a_marine_group
ASV 592	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Albimonas
ASV 596	Epsilonbacteraeota	Campylobacteria	Campylobacterales	Arcobacteraceae	Arcobacter
ASV 599	Bacteroidetes	Bacteroidia	Flavobacteriales	Cryomorphaceae	NS10_marine_group
ASV 610	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Tenacibaculum
ASV 613	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Erythrobacter

ASV 615	Proteobacteria	Alphaproteobacteria	Caulobacterales	Hyphomonadaceae	Henriciella
ASV 628	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Alcanivoracaceae	Alcanivorax
ASV 630	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae	Halioglobus
ASV 639	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Saccharospirillaceae	Oleispira
ASV 645	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Maribius
ASV 650	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Planktotalea
ASV 653	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Salinimonas
ASV 656	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Rubritaleaceae	Rubritalea
ASV 660	Proteobacteria	Alphaproteobacteria	Caulobacterales	Hyphomonadaceae	Hyphomonas
ASV 662	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Psychroserpens
ASV 680	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Erwinia
ASV 699	Proteobacteria	Gammaproteobacteria	Salinisphaerales	Salinisphaeraceae	Salinisphaera
ASV 709	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanothermobacteraceae	Methanothermobacter
ASV 737	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oleiphilaceae	Oleiphilus
ASV 751	Bacteroidetes	Bacteroidia	Cytophagales	Cyclobacteriaceae	Fabibacter
ASV 759	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Winogradskyella
ASV 770	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Paraglaciecola
ASV 782	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	Oceanospirillum
ASV 784	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Jindonia
ASV 802	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae	Candidatus_Megaira
ASV 806	Actinobacteria	Acidimicrobiia	Microtrichales	Microtrichaceae	Sva0996_marine_group
ASV 815	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Spongiibacteraceae	Spongiibacter
ASV 820	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Pseudofulvibacter
ASV 850	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Nonlabens
ASV 861	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Aliivibrio
ASV 887	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Snuella
ASV 902	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae	Colwellia
ASV 904	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Silicimonas
ASV 947	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Halodurantibacterium
ASV 966	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Ruegeria
ASV 967	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Pseudahrensia

ASV 969	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Marinomonadaceae	Marinomonas
ASV 983	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae	Shewanella
ASV 1003	Bacteroidetes	Rhodothermia	Balneolales	Balneolaceae	Balneola
ASV 1016	Planctomycetes	Phycisphaerae	Phycisphaerales	Phycisphaeraceae	Urania-1B-19_marine_sediment_group
ASV 1044	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Muricauda
ASV 1059	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Bizionia
ASV 1085	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	Peredibacter
ASV 1123	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Sedimentitalea
ASV 1124	Verrucomicrobia	Verrucomicrobiae	Opitutales	Opitutaceae	Cephaloticoccus
ASV 1139	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
ASV 1140	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Pelagicola
ASV 1159	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
ASV 1161	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnoclostridium
ASV 1174	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
ASV 1196	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Enterovibrio
ASV 1197	Acidobacteria	Holophagae	Acanthopleuribacterales	Acanthopleuribacteraceae	Acanthopleuribacter
ASV 1208	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aureitalea
ASV 1249	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Maribacter
ASV 1262	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Lentibacter
ASV 1264	Proteobacteria	Gammaproteobacteria	Francisellales	Francisellaceae	Fangia
ASV 1320	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Leuconostoc
ASV 1351	Bacteroidetes	Bacteroidia	Chitinophagales	Saprospiraceae	Saprospira
ASV 1400	Actinobacteria	Actinobacteria	Corynebacteriales	Nocardiaceae	Williamsia
ASV 1441	Proteobacteria	Alphaproteobacteria	Acetobacterales	Acetobacteraceae	Acetobacter
ASV 1445	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Nitrincolaceae	Marinobacterium
ASV 1462	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	1174-901-12
ASV 1474	Proteobacteria	Alphaproteobacteria	Caulobacterales	Hyphomonadaceae	Algimonas
ASV 1483	Bacteroidetes	Bacteroidia	Chitinophagales	Saprospiraceae	Lewinella
ASV 1487	Bacteroidetes	Bacteroidia	Chitinophagales	Saprospiraceae	Portibacter
ASV 1522	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Magnetospiraceae	Magnetospira

ASV 1535	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio
ASV 1537	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Hafnia-Obesumbacterium
ASV 1572	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	Weissella
ASV 1582	Planctomycetes	Phycisphaerae	Phycisphaerales	Phycisphaeraceae	SM1A02
ASV 1588	Proteobacteria	Gammaproteobacteria	Francisellales	Francisellaceae	Francisella
ASV 1643	Bacteroidetes	Bacteroidia	Cytophagales	Cyclobacteriaceae	Marivirga
ASV 1659	Proteobacteria	Alphaproteobacteria	Caulobacterales	Hyphomonadaceae	Ponticaulis
ASV 1664	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Oceanibulbus
ASV 1695	Bacteroidetes	Bacteroidia	Cytophagales	Amoebophilaceae	Candidatus_Amoebophilus
ASV 1715	Bacteroidetes	Bacteroidia	Chitinophagales	Saprospiraceae	Phaeodactylibacter
ASV 1749	Proteobacteria	Gammaproteobacteria	Arenicellales	Arenicellaceae	Arenicella
ASV 1773	Proteobacteria	Alphaproteobacteria	SAR11_clade	Clade_I	Clade_Ib
ASV 1779	Actinobacteria	Actinobacteria	Corynebacteriales	Mycobacteriaceae	Mycobacterium
ASV 1815	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Croceibacter
ASV 1817	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospiraceae_UCG-006
ASV 1837	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Paracoccus
ASV 1844	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Halocynthiibacter
ASV 1862	Chlamydiae	Chlamydiae	Chlamydiales	Simkaniaceae	Simkania
ASV 1875	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aureisphaera
ASV 1880	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Altererythrobacter
ASV 1887	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnospiraceae_NK4A136_group
ASV 1894	Proteobacteria	Alphaproteobacteria	Alphaproteobacteria_Incertae_Sedis	Unknown_Family	Acuticoccus
ASV 1902	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	Neiella
ASV 1916	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Catenococcus
ASV 1928	Cyanobacteria	Oxyphotobacteria	Synechococcales	Cyanobiaceae	Synechococcus_MBIC10613
ASV 1948	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Pseudophaeobacter
ASV 1953	Tenericutes	Mollicutes	Entomoplasmatales	Entomoplasmatales_Incertae_Sedis	Candidatus_Hepatoplasma
ASV 1998	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Actibacterium
ASV 2088	Actinobacteria	Rubrobacteria	Rubrobacterales	Rubrobacteriaceae	Rubrobacter

ASV 2114	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Maritimibacter
ASV 2201	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Spongiivirga
ASV 2216	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Citreitalea
ASV 2235	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Aureivirga
ASV 2445	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Roseicyclus
ASV 2677	Proteobacteria	Deltaproteobacteria	Myxococcales	Sandaracinaceae	Sandaracinus
ASV 2920	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Pseudohalocynthiibacter

1.5 Full list of prokaryotic families predicted from all 16S samples

Full list of prokaryotic families predicted from all 16S samples. Note that ASV number refers to the ASV number of the first species detected in that family. All others are collapsed into that ASV number. For full description of taxonomic predictions see Chapter 2 Section 2.2.1. This data is referred to in Chapter 3 Section 2.3.2.2.

	Phylum	Class	Order	Family
ASV 1	Bacteroidetes	Bacteroidia	Flavobacteriales	Flavobacteriaceae
ASV 2	Proteobacteria	Alphaproteobacteria	SAR11_clade	Clade_I
ASV 3	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Halieaceae
ASV 4	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae
ASV 8	Proteobacteria	Alphaproteobacteria	Parvibaculales	PS1_clade
ASV 11	Cyanobacteria	Oxyphotobacteria	Synechococcales	Cyanobiaceae
ASV 13	Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae
ASV 16	Proteobacteria	Alphaproteobacteria	Puniceispirillales	SAR116_clade
ASV 18	Proteobacteria	Alphaproteobacteria	SAR11_clade	Clade_II
ASV 22	Proteobacteria	Alphaproteobacteria	Caulobacterales	Hyphomonadaceae
ASV 24	Verrucomicrobia	Verrucomicrobiae	Opitutales	Puniceicoccaceae
ASV 30	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Spongiibacteraceae
ASV 32	Proteobacteria	Gammaproteobacteria	Cellvibrionales	Porticoccaceae
ASV 44	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Rubritaleaceae
ASV 59	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae
ASV 60	Bacteroidetes	Bacteroidia	Flavobacteriales	NS9_marine_group
ASV 63	Bacteroidetes	Bacteroidia	Cytophagales	Cyclobacteriaceae
ASV 64	Proteobacteria	Gammaproteobacteria	Alteromonadales	Pseudoalteromonadaceae
ASV 65	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Methylophilaceae
ASV 72	Proteobacteria	Gammaproteobacteria	Thiomicrospirales	Thioglobaceae
ASV 73	Bacteroidetes	Bacteroidia	Flavobacteriales	NS7_marine_group
ASV 92	Proteobacteria	Alphaproteobacteria	Rhodospirillales	AEGEAN-169_marine_group
ASV 101	Verrucomicrobia	Verrucomicrobiae	Pedosphaerales	Pedosphaeraceae
ASV 117	Proteobacteria	Alphaproteobacteria	Parvibaculales	OCS116_clade
ASV 129	Bacteroidetes	Bacteroidia	Flavobacteriales	Cryomorphaceae
ASV 130	Proteobacteria	Alphaproteobacteria	Thalassobaculales	Nisaeaceae

ASV 142	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Litoricolaceae
ASV 143	Lentisphaerae	Lentisphaeria	Lentisphaerales	Lentisphaeraceae
ASV 148	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Nitrosomonadaceae
ASV 159	Cyanobacteria	Oxyphotobacteria	Nostocales	Microcystaceae
ASV 168	Bacteroidetes	Bacteroidia	Flavobacteriales	Crocinitomicaceae
ASV 189	Proteobacteria	Gammaproteobacteria	Ectothiorhodospirales	Ectothiorhodospiraceae
ASV 192	Proteobacteria	Alphaproteobacteria	SAR11_clade	Clade_III
ASV 204	Planctomycetes	Planctomycetacia	Pirellulales	Pirellulaceae
ASV 232	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Woeseiaceae
ASV 245	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Pseudohongiellaceae
ASV 263	Proteobacteria	Gammaproteobacteria	Alteromonadales	Marinobacteraceae
ASV 271	Proteobacteria	Alphaproteobacteria	Rickettsiales	S25-593
ASV 292	Bacteroidetes	Bacteroidia	Sphingobacteriales	NS11-12_marine_group
ASV 304	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae
ASV 324	Planctomycetes	Phycisphaerae	Phycisphaerales	Phycisphaeraceae
ASV 332	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae
ASV 334	Proteobacteria	Gammaproteobacteria	Alteromonadales	Colwelliaceae
ASV 347	Proteobacteria	Gammaproteobacteria	Thiotrichales	Thiotrichaceae
ASV 358	Proteobacteria	Gammaproteobacteria	Coxiellales	Coxiellaceae
ASV 379	Bacteroidetes	Bacteroidia	Chitinophagales	Saprospiraceae
ASV 388	Proteobacteria	Alphaproteobacteria	Rickettsiales	Mitochondria
ASV 412	Proteobacteria	Deltaproteobacteria	Desulfarculales	Desulfarculaceae
ASV 444	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bdellovibrionaceae
ASV 448	Bacteroidetes	Bacteroidia	Cytophagales	Amoebophilaceae
ASV 460	Actinobacteria	Acidimicrobiia	Actinomarinales	Actinomarinaceae
ASV 467	Acidobacteria	Thermoanaerobaculia	Thermoanaerobaculales	Thermoanaerobaculaceae
ASV 470	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	DEV007
ASV 527	Proteobacteria	Alphaproteobacteria	Rickettsiales	Midichloriaceae
ASV 529	Chlamydiae	Chlamydiae	Chlamydiales	Simkaniaceae
ASV 548	Proteobacteria	Alphaproteobacteria	SAR11_clade	Clade_IV
ASV 561	Actinobacteria	Acidimicrobiia	Microtrichales	Ilumatobacteraceae
ASV 573	Proteobacteria	Deltaproteobacteria	Myxococcales	BIrii41

ASV 596	Epsilonbacteraeota	Campylobacteria	Campylobacterales	Arcobacteraceae
ASV 613	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae
ASV 617	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Magnetospiraceae
ASV 628	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Alcanivoracaceae
ASV 639	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Saccharospirillaceae
ASV 690	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae
ASV 699	Proteobacteria	Gammaproteobacteria	Salinisphaerales	Salinisphaeraceae
ASV 709	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanothermobacteraceae
ASV 737	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Oleiphilaceae
ASV 802	Proteobacteria	Alphaproteobacteria	Rickettsiales	Rickettsiaceae
ASV 806	Actinobacteria	Acidimicrobiia	Microtrichales	Microtrichaceae
ASV 890	Proteobacteria	Deltaproteobacteria	Oligoflexales	Oligoflexaceae
ASV 909	Proteobacteria	Gammaproteobacteria	Betaproteobacteriales	Burkholderiaceae
ASV 967	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae
ASV 969	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Marinomonadaceae
ASV 983	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae
ASV 1003	Bacteroidetes	Rhodothermia	Balneolales	Balneolaceae
ASV 1017	Proteobacteria	Alphaproteobacteria	Holosporales	Holosporaceae
ASV 1084	Hydrogenedentes	Hydrogenedentia	Hydrogenedentiales	Hydrogenedensaceae
ASV 1107	Proteobacteria	Alphaproteobacteria	Parvibaculales	Parvibaculaceae
ASV 1108	Proteobacteria	Deltaproteobacteria	Myxococcales	P3OB-42
ASV 1124	Verrucomicrobia	Verrucomicrobiae	Opitutales	Opitutaceae
ASV 1139	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae
ASV 1159	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae
ASV 1161	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
ASV 1174	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae
ASV 1191	Chloroflexi	Anaerolineae	SBR1031	A4b
ASV 1197	Acidobacteria	Holophagae	Acanthopleuribacterales	Acanthopleuribacteraceae
ASV 1264	Proteobacteria	Gammaproteobacteria	Francisellales	Francisellaceae
ASV 1320	Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae
ASV 1361	Actinobacteria	Nitriliruptoria	Nitriliruptorales	Nitriliruptoraceae
ASV 1381	Proteobacteria	Deltaproteobacteria	Myxococcales	Sandaracinaceae

ASV 1400	Actinobacteria	Actinobacteria	Corynebacteriales	Nocardiaceae
ASV 1441	Proteobacteria	Alphaproteobacteria	Acetobacterales	Acetobacteraceae
ASV 1445	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Nitrincolaceae
ASV 1462	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae
ASV 1535	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae
ASV 1749	Proteobacteria	Gammaproteobacteria	Arenicellales	Arenicellaceae
ASV 1779	Actinobacteria	Actinobacteria	Corynebacteriales	Mycobacteriaceae
ASV 1803	Bacteroidetes	Rhodothermia	Rhodothermales	Rhodothermaceae
ASV 1860	Proteobacteria	Alphaproteobacteria	Rickettsiales	SM2D12
ASV 1894	Proteobacteria	Alphaproteobacteria	Alphaproteobacteria_Incertae_Sedis	Unknown_Family
ASV 1925	Proteobacteria	Gammaproteobacteria	Oceanospirillales	P13-46
ASV 1953	Tenericutes	Mollicutes	Entomoplasmatales	Entomoplasmatales_Incertae_Sedis
ASV 1969	Proteobacteria	Alphaproteobacteria	Rickettsiales	AB1
ASV 2088	Actinobacteria	Rubrobacteria	Rubrobacterales	Rubrobacteriaceae
ASV 2089	Chloroflexi	Anaerolineae	Caldilineales	Caldilineaceae

1.6 Full pairwise PERMANOVA p-value results for every condition-type combination of 18S samples

Full pairwise PERMANOVA p-value results for every condition-type combination of 18S samples. This test allows the identification of condition-types that are significantly different in beta-diversity compared to all other condition-types. Condition-type refers to the combination of the type of sample (rRNA or rDNA) and the amendment conditions. For full experimental set up see Chapter 2 Section 2.1.2 and this data is referred to in Section 2.3.3.2. Those p-values in bold are below 0.05 significance threshold.

Condition 1	Condition 2	p-value
rRNA Control	RDNA Control	0.067
rRNA Control	rDNA B_{12}	**0.029**
rRNA Control	rDNA B vitamin mix + Nutrients	0.067
rRNA Control	rDNA Nutrients	**0.008**
rRNA B_{12}	RDNA Control	0.067
rRNA B_{12}	rDNA B_{12}	**0.025**
rRNA B_{12}	rDNA B vitamin mix + Nutrients	0.067
rRNA B_{12}	rDNA Nutrients	**0.009**
rRNA B_{12} + Nutrients	RDNA Control	0.1
rRNA B_{12} + Nutrients	rDNA B_{12}	**0.023**
rRNA B_{12} + Nutrients	rDNA B vitamin mix + Nutrients	0.1
rRNA B_{12} + Nutrients	rDNA Nutrients	**0.025**
rRNA B vitamin mix	RDNA Control	0.067
rRNA B vitamin mix	rDNA B_{12}	**0.026**
rRNA B vitamin mix	rDNA B vitamin mix + Nutrients	0.067
rRNA B vitamin mix	rDNA Nutrients	**0.008**
rRNA B vitamin mix + Nutrients	RDNA Control	0.067
rRNA B vitamin mix + Nutrients	rDNA B_{12}	**0.03**
rRNA B vitamin mix + Nutrients	rDNA B vitamin mix + Nutrients	0.067
rRNA B vitamin mix + Nutrients	rDNA Nutrients	**0.004**
rRNA Nutrients	RDNA Control	0.067
rRNA Nutrients	rDNA B_{12}	**0.024**
rRNA Nutrients	rDNA B vitamin mix + Nutrients	0.067
rRNA Nutrients	rDNA Nutrients	**0.004**
rRNA Control	rRNA B_{12}	0.403
rRNA Control	rRNA B_{12} + Nutrients	0.333
rRNA Control	rRNA B vitamin mix	0.099
rRNA Control	rRNA B vitamin mix + Nutrients	**0.027**
rRNA Control	rRNA Nutrients	0.509
rRNA B_{12}	rRNA B_{12} + Nutrients	0.445
rRNA B_{12}	rRNA B vitamin mix	**0.022**
rRNA B_{12}	rRNA B vitamin mix + Nutrients	**0.039**
rRNA B_{12}	rRNA Nutrients	0.756
rRNA B_{12} + Nutrients	rRNA B vitamin mix	0.199
rRNA B_{12} + Nutrients	rRNA B vitamin mix + Nutrients	0.141
rRNA B_{12} + Nutrients	rRNA Nutrients	0.981
rRNA B vitamin mix	rRNA B vitamin mix + Nutrients	0.138
rRNA B vitamin mix	rRNA Nutrients	0.101
rRNA B vitamin mix + Nutrients	rRNA Nutrients	0.448
rDNA Control	rDNA B_{12}	0.2

rDNA Control	rDNA B vitamin mix + Nutrients	0.334
rDNA Control	rDNA Nutrients	0.053
rDNA B_{12}	rDNA B vitamin mix + Nutrients	0.067
rDNA B_{12}	rDNA Nutrients	**0.009**
rDNA B vitamin mix + Nutrients	rDNA Nutrients	0.629

1.7 Full list of eukaryotic genera predicted from all 18S samples

Full list of eukaryotic genera predicted from all 18S samples. Note that ASV number refers to the ASV number of the first species detected in that genera. All others are collapsed into that ASV number. For full description of taxonomic predictions see Chapter 2 Section 2.1.2. This data is referred to in Chapter 2 Section 2.3.4.2.

	Supergroup	Division	Class	Order	Family	Genus
ASV 2	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Gymnodiniaceae*	*Gyrodinium*
ASV 3	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Cryomonadida*	*Cryothecomonas-lineage*	*Cryothecomonas*
ASV 6	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Leptocylindrus*
ASV 7	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Torodiniales*	*Torodiniaceae*	*Torodinium*
ASV 8	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Thalassiosira*
ASV 13	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Dinophyceae_X*	*Dinophyceae_XX*	*Dinophyceae_XXX*
ASV 14	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Chaetoceros*
ASV 15	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Guinardia*
ASV 16	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Bathycoccaceae*	*Ostreococcus*
ASV 19	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Strobilidiidae_I*	*Strobilidiidae_I_X*
ASV 20	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Lynnellidae*	*Lynnella*
ASV 21	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida_D*	*Strombidiida_D_X*	*Strombidiida_D_X X*
ASV 23	*Alveolata*	*Dinoflagellata*	*Noctilucophyceae*	*Noctilucales*	*Noctilucaceae*	*Noctiluca*
ASV 24	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Skeletonema*
ASV 25	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Mamiellaceae*	*Micromonas*
ASV 26	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiales_X*	*Chrysocampanula*
ASV 27	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Kareniaceae*	*Karenia*
ASV 29	*Archaeplastida*	*Chlorophyta*	*Trebouxiophyceae*	*Chlorellales*	*Chlorellales_X*	*Picochlorum*
ASV 30	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Tintinnida*	*Eutintinnidae*	*Eutintinnus*
ASV 32	*Hacrobia*	*Telonemia*	*Telonemia_X*	*Telonemia_XX*	*Telonemia-Group-2*	*Telonemia-Group-2_X*

ASV 33	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Arcocellulus*
ASV 36	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Gymnodiniaceae*	*Lepidodinium*
ASV 38	*Stramenopiles*	*Stramenopiles_X*	*Stramenopiles_X-Group-7*	*Stramenopiles_X-Group-7_X*	*Stramenopiles_X-Group-7_XX*	*Stramenopiles_X-Group-7_XXX*
ASV 39	*Stramenopiles*	*Ochrophyta*	*Chrysophyceae*	*Chrysophyceae_X*	*Chrysophyceae_Clade-C*	*Chrysophyceae_Clade-C_X*
ASV 41	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiaceae*	*Prymnesiaceae_X*
ASV 42	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiophyceae_X*	*Braarudosphaeraceae*	*Braarudosphaera*
ASV 44	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gonyaulacales*	*Ceratiaceae*	*Tripos*
ASV 48	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiaceae*	*Prymnesium*
ASV 49	*Rhizaria*	*Cercozoa*	*Chlorarachniophyceae*	*Chlorarachnida*	*Minorisa-lineage*	*Minorisa*
ASV 50	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Chrysochromulinaceae*	*Chrysochromulina*
ASV 52	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiaceae*	*Haptolina*
ASV 56	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Tintinnida*	*Tintinnidae*	*Tintinnidae_X*
ASV 58	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Tontoniidae_B*	*Tontoniidae_B_X*
ASV 60	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Ventricleftida*	*Ventrifissuridae*	*Ventrifissuridae_X*
ASV 69	*Stramenopiles*	*Pseudofungi*	*MAST-1*	*MAST-1C*	*MAST-1C_X*	*MAST-1C_XX*
ASV 74	*Archaeplastida*	*Chlorophyta*	*Prasino-Clade-V*	*Pseudoscourfieldiales*	*Pycnococcaceae*	*Pycnococcaceae_X*
ASV 83	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Raphid-pennate*	*Pseudo-nitzschia*
ASV 85	*Archaeplastida*	*Chlorophyta*	*Pyramimonadophyceae*	*Pyramimonadales*	*Pterospermaceae*	*Pterosperma*
ASV 91	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Cryomonadida*	*Protaspa-lineage*	*Protaspa-lineage_X*
ASV 93	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Tontoniidae_A*	*Spirotontonia*
ASV 94	*Hacrobia*	*Cryptophyta*	*Cryptophyceae*	*Cryptomonadales*	*Cryptomonadales_X*	*Plagioselmis*
ASV 95	*Hacrobia*	*Picozoa*	*Picozoa_X*	*Picozoa_XX*	*Picozoa_XXX*	*Picozoa_XXXX*
ASV 96	*Alveolata*	*Ciliophora*	*Nassophorea*	*Nassophorea_X*	*Discotrichidae*	*NASSO_1*
ASV 100	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-IV*	*Dino-Group-IV-Syndinium-Group*	*Syndinium*

ASV 102	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Gymnodiniaceae*	*Balechina*
ASV 103	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Rhizosolenia*
ASV 108	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Raphid-pennate*	*Meuniera*
ASV 109	*Alveolata*	*Ciliophora*	*Heterotrichea*	*Heterotrichea_X*	*Heterotrichida*	*Heterotrichida_X*
ASV 113	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Araphid-pennate*	*Delphineis*
ASV 115	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Isochrysidales*	*Noelaerhabdaceae*	*Gephyrocapsa*
ASV 119	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Tontoniidae_B*	*Pseudotontonia*
ASV 126	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae_Q*	*Strombidiidae_Q_X*
ASV 130	*Rhizaria*	*Cercozoa*	*Filosa-Imbricatea*	*Filosa-Imbricatea_X*	*Novel-clade-2*	*Novel-clade-2_X*
ASV 131	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Prorocentrales*	*Prorocentraceae*	*Prorocentrum*
ASV 137	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Strobilidiidae_I*	*Pelagostrobilidium*
ASV 139	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae_M*	*Strombidium_M*
ASV 140	*Stramenopiles*	*Ochrophyta*	*Chrysophyceae*	*Chrysophyceae_X*	*Chrysophyceae_Clade-G*	*Chrysophyceae_Clade-G_X*
ASV 147	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Cerataulina*
ASV 148	*Alveolata*	*Ciliophora*	*Ciliophora_X*	*Ciliophora_XX*	*Ciliophora_XXX*	*Mesodinium*
ASV 149	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae_L*	*Strombidiidae_L_X*
ASV 155	*Alveolata*	*Ciliophora*	*CONThreeP*	*CONThreeP_X*	*CONThreeP_XX*	*Askenasia*
ASV 156	*Stramenopiles*	*Sagenista*	*MAST-4*	*MAST-4A*	*MAST-4A_X*	*MAST-4A_XX*
ASV 159	*Hacrobia*	*Centroheliozoa*	*Centroheliozoa_X*	*Pterocystida*	*Pterocystida_X*	*Pterocystida_XX*
ASV 160	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiophyceae_X*	*Prymnesiophyceae_XX*	*Prymnesiophyceae_XXX*
ASV 161	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiophyceae_Clade_D*	*Prymnesiophyceae_Clade_D_X*	*Prymnesiophyceae_Clade_D_XX*
ASV 164	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-I*	*Dino-Group-I-Clade-1*	*Dino-Group-I-Clade-1_X*
ASV 165	*Rhizaria*	*Cercozoa*	*Filosa-Imbricatea*	*Marimonadida*	*Marimonadida_X*	*Marimonadida_XX*
ASV 168	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Ventricleftida*	*CCW10-lineage*	*CCW10-lineage_X*
ASV 170	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Chytriodiniaceae*	*Dissodinium*

ASV 173	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Ebriida*	*Botuliformidae*	*Botuliformidae_X*
ASV 179	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Gymnodiniaceae*	*Gymnodinium*
ASV 184	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Minutocellus*
ASV 185	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-I*	*Dino-Group-I-Clade-4*	*Dino-Group-I-Clade-4_X*
ASV 188	*Rhizaria*	*Cercozoa*	*Cercozoa_X*	*Cercozoa_XX*	*Cercozoa_XXX*	*Cercozoa_XXXX*
ASV 189	*Archaeplastida*	*Prasinodermophyta*	*Prasinodermophyceae*	*Prasinococcales*	*Prasinococcales-Clade-B*	*Prasinoderma*
ASV 190	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Minidiscus*
ASV 191	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Heterocapsaceae*	*Heterocapsa*
ASV 192	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Protoperidiniaceae*	*Protoperidinium*
ASV 199	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-6*	*Dino-Group-II-Clade-6_X*
ASV 204	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Thoracosphaeraceae*	*Scrippsiella*
ASV 207	*Hacrobia*	*Katablepharidophyta*	*Katablepharidaceae*	*Katablepharidales*	*Katablepharidales_X*	*Katablepharidales_XX*
ASV 208	*Stramenopiles*	*Ochrophyta*	*MOCH-3*	*MOCH-3_X*	*MOCH-3_XX*	*MOCH-3_XXX*
ASV 209	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-3*	*Dino-Group-II-Clade-3_X*
ASV 210	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-4*	*Dino-Group-II-Clade-4_X*
ASV 213	*Hacrobia*	*Cryptophyta*	*Cryptophyceae*	*Cryptomonadales*	*Cryptomonadales_X*	*Cryptomonadales_XX*
ASV 216	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Syracosphaerales*	*Syracosphaerales_X*	*Syracosphaerales_XX*
ASV 217	*Archaeplastida*	*Chlorophyta*	*Pyramimonadophyceae*	*Pyramimonadales*	*Pyramimonadaceae*	*Pyramimonas*
ASV 218	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-I*	*Dino-Group-I-Clade-5*	*Dino-Group-I-Clade-5_X*
ASV 221	*Stramenopiles*	*Ochrophyta*	*Chrysophyceae*	*Chrysophyceae_X*	*Chrysophyceae_Clade-F*	*Paraphysomonas*

ASV 222	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Thoracosphaeraceae*	*Pentapharsodinium*
ASV 226	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Raphid-pennate*	*Cylindrotheca*
ASV 228	*Alveolata*	*Ciliophora*	*Oligohymenophorea*	*Scuticociliatia_1*	*Philasterida*	*Porpostoma*
ASV 236	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-1*	*Dino-Group-II-Clade-1_X*
ASV 240	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Lauderia*
ASV 245	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Phaeocystales*	*Phaeocystaceae*	*Phaeocystis*
ASV 249	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Suessiales*	*Suessiaceae*	*Protodinium*
ASV 252	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-I*	*Dino-Group-I-Clade-4*	*Euduboscquella*
ASV 254	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-23*	*Dino-Group-II-Clade-23_X*
ASV 255	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Suessiales*	*Suessiaceae*	*Pelagodinium*
ASV 257	*Hacrobia*	*Cryptophyta*	*Cryptophyceae*	*Cryptomonadales*	*Cryptomonadales_X*	*Teleaulax*
ASV 258	*Archaeplastida*	*Chlorophyta*	*Pyramimonadophyceae*	*Pyramimonadales*	*Pyramimonadales_X*	*Pyramimonadales_XX*
ASV 260	*Stramenopiles*	*Sagenista*	*Labyrinthulomycetes*	*Thraustochytriales*	*Thraustochytriaceae*	*Thraustochytriaceae_X*
ASV 265	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-III*	*Dino-Group-III_X*	*Dino-Group-III_XX*
ASV 271	*Archaeplastida*	*Chlorophyta*	*Chlorodendrophyceae*	*Chlorodendrales*	*Chlorodendraceae*	*Chlorodendrales_XX*
ASV 274	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae*	*Parallelostrombidium*
ASV 276	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Dolichomastigales*	*Crustomastigaceae*	*Crustomastigaceae-AB*
ASV 279	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Thoracosphaeraceae*	*Luciella*
ASV 280	*Stramenopiles*	*Sagenista*	*MAST-7*	*MAST-7A*	*MAST-7A_X*	*MAST-7A_XX*
ASV 284	*Stramenopiles*	*Ochrophyta*	*Chrysophyceae*	*Chrysophyceae_X*	*Chrysophyceae_Clade-I*	*Chrysophyceae_Clade-I_X*
ASV 291	*Stramenopiles*	*Ochrophyta*	*Pelagophyceae*	*Pelagomonadales*	*Pelagomonadaceae*	*Pelagomonadaceae_clade_C*

ASV 293	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Ebriida*	*TAGIRI1-lineage*	*TAGIRI1-lineage_X*
ASV 296	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Dactyliosolen*
ASV 300	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiophyceae_Clade_E*	*Prymnesiophyceae_Clade_E_X*	*Prymnesiophyceae_Clade_E_XX*
ASV 312	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-32*	*Dino-Group-II-Clade-32_X*
ASV 313	*Rhizaria*	*Cercozoa*	*Chlorarachniophyceae*	*Chlorarachniophyceae_X*	*NPK2-lineage*	*NPK2-lineage_X*
ASV 314	*Amoebozoa*	*Conosa*	*Variosea*	*Variosea_X*	*Mariager-Fjord-lineage*	*Mariager-Fjord-lineage_X*
ASV 318	*Stramenopiles*	*Ochrophyta*	*MOCH-1*	*MOCH-1_X*	*MOCH-1_XX*	*MOCH-1_XXX*
ASV 319	*Stramenopiles*	*Sagenista*	*MAST-4*	*MAST-4B*	*MAST-4B_X*	*MAST-4B_XX*
ASV 323	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Proboscia*
ASV 326	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Cyclotella*
ASV 327	*Stramenopiles*	*Sagenista*	*Labyrinthulomycetes*	*Thraustochytriales*	*Thraustochytriaceae*	*Labyrinthuloides*
ASV 336	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Suessiales*	*Suessiaceae*	*Biecheleria*
ASV 337	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiaceae*	*Dicrateria*
ASV 339	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Mamiellaceae*	*Mamiella*
ASV 341	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiophyceae_X*	*Braarudosphaeraceae*	*Braarudosphaeraceae_X*
ASV 342	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Diplopsalidaceae*	*Diplopsalis*
ASV 350	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiophyceae_Clade_B4*	*Prymnesiophyceae_Clade_B4_X*
ASV 353	*Stramenopiles*	*Sagenista*	*Labyrinthulomycetes*	*Labyrinthulales*	*Labyrinthulaceae*	*Aplanochytrium*
ASV 356	*Rhizaria*	*Cercozoa*	*Filosa-Imbricatea*	*Thaumatomonadida*	*Thaumatomonadidae*	*Thaumatomastix*
ASV 357	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Kareniaceae*	*Karlodinium*
ASV 364	*Archaeplastida*	*Chlorophyta*	*Pyramimonadophyceae*	*Pyramimonadales*	*Halosphaeraceae*	*Halosphaera*
ASV 372	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Mamiellaceae*	*RCC391*

ASV 377	*Rhizaria*	*Cercozoa*	*Filosa-Imbricatea*	*Euglyphida*	*Paulinellidae*	*Paulinella*
ASV 382	*Stramenopiles*	*Sagenista*	*MAST-7*	*MAST-7B*	*MAST-7B_X*	*MAST-7B_XX*
ASV 391	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-8*	*Dino-Group-II-Clade-8_X*
ASV 400	*Stramenopiles*	*Ochrophyta*	*Bolidophyceae*	*Parmales*	*Triparmaceae*	*Triparma*
ASV 404	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Choreotrichida_X*	*Choreotrichida_XX*
ASV 416	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Ventricleftida*	*Ventricleftida_X*	*Ventricleftida_XX*
ASV 418	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-7*	*Dino-Group-II-Clade-7_X*
ASV 419	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida_G*	*Strombidiida_G_X*	*Strombidiida_G_XX*
ASV 426	*Alveolata*	*Ciliophora*	*Litostomatea*	*Litostomatea_X*	*Litostomatea_XX*	*Litostomatea_XXX*
ASV 428	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Dolichomastigales*	*Crustomastigaceae*	*Crustomastigaceae_X*
ASV 432	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-12*	*Dino-Group-II-Clade-12_X*
ASV 437	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Filosa-Thecofilosea_X*	*Filosa-Thecofilosea_XX*	*Filosa-Thecofilosea_XXX*
ASV 448	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Dolichomastigales*	*Crustomastigaceae*	*Crustomastix*
ASV 451	*Rhizaria*	*Cercozoa*	*Filosa-Imbricatea*	*Thaumatomonadida*	*Thaumatomonadidae*	*Reckertia*
ASV 459	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-14*	*Dino-Group-II-Clade-14_X*
ASV 464	*Archaeplastida*	*Chlorophyta*	*Nephroselmidophyceae*	*Nephroselmidales*	*Nephroselmidales_X*	*Nephroselmis*
ASV 466	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Amphidomataceae*	*Azadinium*
ASV 467	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Filosa-Thecofilosea_X*	*Mataza-lineage*	*Mataza-lineage_X*
ASV 468	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II_X*	*Dino-Group-II_XX*
ASV 479	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Thoracosphaeraceae*	*Thoracosphaeraceae_X*

ASV 490	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-10-and-11*	*Dino-Group-II-Clade-10-and-11_X*
ASV 496	*Stramenopiles*	*Ochrophyta*	*Pelagophyceae*	*Pelagomonadales*	*Pelagomonadaceae*	*Aureococcus*
ASV 497	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Raphid-pennate*	*Nitzschia*
ASV 499	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-I*	*Dino-Group-I-Clade-3*	*Dino-Group-I-Clade-3_X*
ASV 504	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Thoracosphaeraceae*	*Apocalathium*
ASV 507	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Cryomonadida*	*Protaspa-lineage*	*Protaspa*
ASV 509	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-5*	*Dino-Group-II-Clade-5_X*
ASV 520	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Calcihaptophycidae*	*Watznaueriaceae*	*Tergestiella*
ASV 534	*Stramenopiles*	*Sagenista*	*MAST-7*	*MAST-7C*	*MAST-7C_X*	*MAST-7C_XX*
ASV 545	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gonyaulacales*	*Pyrophacaceae*	*Fragilidium*
ASV 547	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiophyceae_Clade_B5*	*Prymnesiophyceae_Clade_B5_X*
ASV 549	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Strombidinopsidae*	*Strombidinopsis*
ASV 550	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-44*	*Dino-Group-II-Clade-44_X*
ASV 556	*Rhizaria*	*Cercozoa*	*Filosa*	*Filosa_X*	*Filosa_XX*	*Discomonas*
ASV 566	*Rhizaria*	*Cercozoa*	*Filosa-Sarcomonadea*	*Cercomonadida*	*Cercomonadidae*	*Cercomonas*
ASV 567	*Stramenopiles*	*Ochrophyta*	*MOCH-2*	*MOCH-2_X*	*MOCH-2_XX*	*MOCH-2_XXX*
ASV 574	*Rhizaria*	*Radiolaria*	*Acantharea*	*Acantharea_X*	*Acantharea_XX*	*Trizona*
ASV 577	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae_B*	*Strombidiidae_B_X*
ASV 581	*Stramenopiles*	*Stramenopiles_X*	*Stramenopiles_XX*	*Stramenopiles_XXX*	*Stramenopiles_XXXX*	*Stramenopiles_XXXXX*
ASV 587	*Rhizaria*	*Cercozoa*	*Chlorarachniophyceae*	*Chlorarachnida*	*Chlorarachnida_X*	*Chlorarachnida_XX*
ASV 588	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Araphid-pennate*	*Thalassionema*
ASV 591	*Stramenopiles*	*Ochrophyta*	*Dictyochophyceae*	*Dictyochophyceae_X*	*Dictyochales*	*Dictyochales_X*

ASV 604	*Stramenopiles*	*Ochrophyta*	*Chrysophyceae*	*Chrysophyceae_X*	*Chrysophyceae_Clade-H*	*Chrysophyceae_Clade-H_X*
ASV 611	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Calcihaptophycidae*	*Calcihaptophycidae_X*	*Calcihaptophycidae_XX*
ASV 614	*Rhizaria*	*Cercozoa*	*Chlorarachniophyceae*	*Chlorarachnida*	*Minorisa-lineage*	*Minorisa-lineage_X*
ASV 617	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Tintinnida*	*Eutintinnidae*	*Eutintinnidae_X*
ASV 619	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Warnowiaceae*	*Warnowia*
ASV 629	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiales*	*Prymnesiophyceae_Clade_B3*	*Prymnesiophyceae_Clade_B3_X*
ASV 630	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Brockmanniella*
ASV 648	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-55*	*Dino-Group-II-Clade-55_X*
ASV 656	*Stramenopiles*	*Opalozoa*	*MAST-3*	*MAST-3I*	*MAST-3I_X*	*MAST-3I_XX*
ASV 657	*Stramenopiles*	*Pseudofungi*	*MAST-1*	*MAST-1B*	*MAST-1B_X*	*MAST-1B_XX*
ASV 659	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Mamiellaceae*	*Mantoniella*
ASV 671	*Hacrobia*	*Centroheliozoa*	*Centroheliozoa_X*	*Centroheliozoa_XX*	*Centroheliozoa_XXX*	*Centroheliozoa_XXXX*
ASV 678	*Stramenopiles*	*Opalozoa*	*Bicoecea*	*Bicoecales*	*Bicoecaceae*	*Bicoecaceae_X*
ASV 686	*Stramenopiles*	*Pseudofungi*	*MAST-2*	*MAST-2B*	*MAST-2B_X*	*MAST-2B_XX*
ASV 687	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-30*	*Dino-Group-II-Clade-30_X*
ASV 696	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II_X*	*Dino-Group-II-Clade-52_X*
ASV 712	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Leegaardiellidae_B*	*Leegaardiella*
ASV 713	*Archaeplastida*	*Chlorophyta*	*Ulvophyceae*	*Ulvales-relatives*	*Ulvales-relatives_X*	*Ulvales-relatives_XX*
ASV 715	*Archaeplastida*	*Streptophyta*	*Embryophyceae*	*Embryophyceae_X*	*Embryophyceae_XX*	*Silene*
ASV 727	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida_F*	*Strombidiida_F_X*	*Strombidiida_F_XX*
ASV 735	*Alveolata*	*Ciliophora*	*CONTH_8*	*CONTH_8_X*	*CONTH_8_XX*	*CONTH_8_XXX*

ASV 739	*Hacrobia*	*Haptophyta*	*Haptophyta_Clade_HAP3*	*Haptophyta_Clade_HAP3_X*	*Haptophyta_Clade_HAP3_XX*	*Haptophyta_Clade_HAP3_XXX*
ASV 743	*Alveolata*	*Ciliophora*	*Phyllopharyngea*	*Cyrtophoria_1*	*PHYLL_4*	*PHYLL_4_X*
ASV 756	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Filosa-Thecofilosea_X*	*Novel-Clade-4*	*Novel-Clade-4_X*
ASV 764	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae_R*	*Strombidium_R*
ASV 774	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Raphid-pennate*	*Pleurosigma*
ASV 782	*Stramenopiles*	*Sagenista*	*Labyrinthulomycetes*	*Labyrinthulales*	*Labyrinthulaceae*	*Labyrinthulaceae_X*
ASV 785	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Coccolithales*	*Coccolithaceae*	*Coccolithus*
ASV 786	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Actinocyclus*
ASV 788	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-I*	*Dino-Group-I-Clade-8*	*Dino-Group-I-Clade-8_X*
ASV 810	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Leegaardiellidae_A*	*Leegaardiellidae_A_X*
ASV 822	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Dinophysiales*	*Dinophysiaceae*	*Dinophysis*
ASV 837	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Podolampadaceae*	*Lessardia*
ASV 838	*Stramenopiles*	*Sagenista*	*MAST-8*	*MAST-8B*	*MAST-8B_X*	*MAST-8B_XX*
ASV 839	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Corethron*
ASV 843	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Dinophysiales*	*Amphisoleniaceae*	*Amphisoleniaceae_X*
ASV 848	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Calcihaptophycidae*	*Rhabdosphaeraceae*	*Algirosphaera*
ASV 849	*Archaeplastida*	*Prasinodermophyta*	*Prasinodermophyceae*	*Prasinococcales*	*Prasinococcales-Clade-B*	*Prasinococcales-Clade-B_X*
ASV 877	*Hacrobia*	*Haptophyta*	*Haptophyta_X*	*Haptophyta_XX*	*Haptophyta_XXX*	*Haptophyta_XXXX*
ASV 884	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-22*	*Dino-Group-II-Clade-22_X*
ASV 892	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-39*	*Dino-Group-II-Clade-39_X*
ASV 895	*Rhizaria*	*Radiolaria*	*Acantharea*	*Chaunacanthida*	*Chaunacanthida_X*	*Acanthometron*

ASV 910	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gymnodiniales*	*Gymnodiniaceae*	*Paragymnodinium*
ASV 924	*Alveolata*	*Dinoflagellata*	*Noctilucophyceae*	*Noctilucales*	*Noctilucaceae*	*Spatulodinium*
ASV 934	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-47*	*Dino-Group-II-Clade-47_X*
ASV 947	*Stramenopiles*	*Sagenista*	*MAST-4*	*MAST-4E*	*MAST-4E_X*	*MAST-4E_XX*
ASV 952	*Stramenopiles*	*Ochrophyta*	*Bolidophyceae*	*Parmales*	*Parmales_env_3*	*Parmales_env_3_X*
ASV 954	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Diplopsalidaceae*	*Gotoius*
ASV 971	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-57*	*Dino-Group-II-Clade-57_X*
ASV 984	*Stramenopiles*	*Opalozoa*	*MAST-3*	*MAST-3D*	*MAST-3D_X*	*MAST-3D_XX*
ASV 992	*Alveolata*	*Ciliophora*	*Colpodea*	*Colpodea_X*	*Cyrtolophosidida*	*Aristerostoma*
ASV 1007	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-42*	*Dino-Group-II-Clade-42_X*
ASV 1044	*Hacrobia*	*Telonemia*	*Telonemia_X*	*Telonemia_XX*	*Telonemia-Group-1*	*Telonemia-Group-1_X*
ASV 1050	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae_I*	*Strombidiidae_I_X*
ASV 1067	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Gonyaulacales*	*Goniodomataceae*	*Alexandrium*
ASV 1107	*Alveolata*	*Apicomplexa*	*Gregarinomorphea*	*Gregarines_GRE2*	*Gregarines_GRE2_X*	*Gregarines_GRE2_XX*
ASV 1113	*Stramenopiles*	*Sagenista*	*MAST-7*	*MAST-7_X*	*MAST-7_XX*	*MAST-7_XXX*
ASV 1118	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Strombidiida*	*Strombidiidae_K*	*Strombidium_K*
ASV 1123	*Alveolata*	*Dinoflagellata*	*Dinophyceae*	*Peridiniales*	*Amphidiniopsidaceae*	*Islandinium*
ASV 1138	*Hacrobia*	*Haptophyta*	*Prymnesiophyceae*	*Prymnesiophyceae_Clade_F*	*Prymnesiophyceae_Clade_F_X*	*Prymnesiophyceae_Clade_F_XX*
ASV 1139	*Stramenopiles*	*Ochrophyta*	*Dictyochophyceae*	*Dictyochophyceae_X*	*Dictyochales*	*Dictyocha*
ASV 1210	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Raphid-pennate*	*Navicula*
ASV 1228	*Stramenopiles*	*Ochrophyta*	*Pelagophyceae*	*Pelagomonadales*	*Pelagomonadaceae*	*Pelagomonas*
ASV 1237	*Archaeplastida*	*Chlorophyta*	*Trebouxiophyceae*	*Watanabea-Clade*	*Watanabea-Clade_X*	*Diplosphaera*
ASV 1246	*Alveolata*	*Ciliophora*	*Oligohymenophorea*	*Apostomatia*	*Foettingeriidae*	*Synophrya*
ASV 1274	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Polar-centric-Mediophyceae*	*Eucampia*

ASV 1283	*Stramenopiles*	*Ochrophyta*	*Bacillariophyta*	*Bacillariophyta_X*	*Radial-centric-basal-Coscinodiscophyceae*	*Coscinodiscus*
ASV 1350	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Strobilidiidae_G*	*Strobilidiidae_G_X*
ASV 1365	*Alveolata*	*Dinoflagellata*	*Syndiniales*	*Dino-Group-II*	*Dino-Group-II-Clade-50*	*Dino-Group-II-Clade-50_X*
ASV 1390	*Archaeplastida*	*Chlorophyta*	*Mamiellophyceae*	*Mamiellales*	*Bathycoccaceae*	*Bathycoccus*
ASV 1500	*Stramenopiles*	*Opalozoa*	*MAST-3*	*MAST-3F*	*MAST-3F_X*	*MAST-3F_XX*
ASV 1659	*Alveolata*	*Dinoflagellata*	*Noctilucophyceae*	*Noctilucales*	*Noctilucaceae*	*Noctilucales_X*
ASV 1694	*Archaeplastida*	*Chlorophyta*	*Trebouxiophyceae*	*Chlorellales*	*Chlorellales_X*	*Nannochloris*
ASV 1704	*Hacrobia*	*Centroheliozoa*	*Centroheliozoa_X*	*Pterocystida*	*Pterocystidae*	*Raineriophrys*
ASV 1746	*Stramenopiles*	*Opalozoa*	*MAST-3*	*MAST-3J*	*MAST-3J_X*	*MAST-3J_XX*
ASV 1823	*Rhizaria*	*Cercozoa*	*Filosa-Thecofilosea*	*Ventricleftida*	*Ventrifissuridae*	*Ventrifissura*
ASV 1930	*Apusozoa*	*Apusomonadidae*	*Apusomonadidae_Group-1*	*Apusomonadidae_Group-1_X*	*Apusomonadidae_Group-1_XX*	*Amastigomonas*
ASV 1982	*Stramenopiles*	*Pseudofungi*	*Oomycota*	*Oomycota_X*	*Oomycota_XX*	*Oomycota_XXX*
ASV 2006	*Alveolata*	*Ciliophora*	*Spirotrichea*	*Choreotrichida*	*Strombidinopsidae*	*Parastrombidinopsis*

1.8 Full list of eukaryotic species predicted from all 18S samples

Full list of eukaryotic species predicted from all 18S samples. Note that ASV number refers to the ASV number of the first ASV detected for that species. All others are collapsed into that ASV number. For full description of taxonomic predictions see Chapter 2 Section 2.1.2. This data is referred to in Chapter 2 Section 2.3.4.2.

	Supergroup	Division	Class	Order	Family	Genus	Species
ASV 2	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_fusiforme
ASV 3	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	Cryothecomonas_aestivalis
ASV 6	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Leptocylindrus	Leptocylindrus_sp.
ASV 7	Alveolata	Dinoflagellata	Dinophyceae	Torodiniales	Torodiniaceae	Torodinium	Torodinium_robustum

ASV 8	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	NA
ASV 11	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_curvisetus_2b
ASV 13	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 14	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_tenuissimus
ASV 15	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Guinardia	Guinardia_delicatula
ASV 16	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Ostreococcus	Ostreococcus_lucimarinus
ASV 19	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_I	Strobilidiidae_I_X	Strobilidiidae_I_X_sp.
ASV 20	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Lynnellidae	Lynnella	Lynnella_semiglobulosa
ASV 21	Alveolata	Ciliophora	Spirotrichea	Strombidiida_D	Strombidiida_D_X	Strombidiida_D_XX	Strombidiida_D_XX_sp.
ASV 23	Alveolata	Dinoflagellata	Noctilucophyceae	Noctilucales	Noctilucaceae	Noctiluca	Noctiluca_scintillans
ASV 24	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Skeletonema	NA
ASV 25	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas	Micromonas_commoda_A2
ASV 26	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiales_X	Chrysocampanula	Chrysocampanula_spinifera
ASV 27	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karenia	Karenia_brevis
ASV 29	Archaeplastida	Chlorophyta	Trebouxiophyceae	Chlorellales	Chlorellales_X	Picochlorum	Picochlorum_sp.
ASV 30	Alveolata	Ciliophora	Spirotrichea	Tintinnida	Eutintinnidae	Eutintinnus	Eutintinnus_sp.

ASV 32	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 33	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Arcocellulus	Arcocellulus_cornucervis
ASV 34	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_sp.
ASV 36	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Lepidodinium	Lepidodinium_chlorophorum
ASV 37	Alveolata	Dinoflagellata	Noctilucophyceae	Noctilucales	Noctilucaceae	Noctiluca	Noctiluca_scintillans
ASV 38	Stramenopiles	Stramenopiles_X	Stramenopiles_X-Group-7	Stramenopiles_X-Group-7_X	Stramenopiles_X-Group-7_XX	Stramenopiles_X-Group-7_XXX	Stramenopiles_X-Group-7_XXX_sp.
ASV 39	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-C	Chrysophyceae_Clade-C_X	Chrysophyceae_Clade-C_X_sp.
ASV 41	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesiaceae_X	Prymnesiaceae_X_sp.
ASV 42	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	Braarudosphaera	Braarudosphaera_bigelowii
ASV 44	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	NA
ASV 47	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	NA	NA	NA
ASV 48	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 49	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachnida	Minorisa-lineage	Minorisa	Minorisa_minuta
ASV 50	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 52	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Haptolina	Haptolina_sp.
ASV 56	Alveolata	Ciliophora	Spirotrichea	Tintinnida	Tintinnidae	Tintinnidae_X	Tintinnidae_X_sp.

ASV 57	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	Cryothecomonas_aestivalis
ASV 58	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_B	Tontoniidae_B_X	Tontoniidae_B_X_sp.
ASV 59	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 60	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 63	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-C	Chrysophyceae_Clade-C_X	Chrysophyceae_Clade-C_X_sp.
ASV 65	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	NA	NA	NA
ASV 66	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	Peronosporales	NA	NA
ASV 67	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Haptolina	Haptolina_sp.
ASV 68	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 69	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
ASV 71	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas	Micromonas_bravo_B1
ASV 72	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas	Micromonas_pusilla
ASV 74	Archaeplastida	Chlorophyta	Prasino-Clade-V	Pseudoscourfieldiales	Pycnococcaceae	Pycnococcaceae_X	Pycnococcaceae_X_sp.
ASV 77	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_delicatula
ASV 78	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Leptocylindrus	Leptocylindrus_minimus
ASV 80	Alveolata	Ciliophora	Spirotrichea	Tintinnida	Eutintinnidae	NA	NA

ASV 82	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas	Micromonas_bravo_B2
ASV 83	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	NA
ASV 85	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_sp.
ASV 86	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Heterocapsaceae	Heterocapsa	Heterocapsa_pygmaea
ASV 87	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Heterocapsaceae	Heterocapsa	Heterocapsa_nei/rotundata
ASV 89	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_I	Strobilidiidae_I_X	Strobilidiidae_I_X_sp.
ASV 91	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 92	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-F	NA	NA
ASV 93	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_A	Spirotontonia	Spirotontonia_sp.
ASV 94	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Plagioselmis	Plagioselmis_prolonga
ASV 95	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 96	Alveolata	Ciliophora	Nassophorea	Nassophorea_X	Discotrichidae	NASSO_1	NASSO_1_sp.
ASV 100	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-IV	Dino-Group-IV-Syndinium-Group	Syndinium	Syndinium_sp.
ASV 101	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
ASV 102	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Balechina	Balechina_pachyderma ta
ASV 103	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Rhizosolenia	Rhizosolenia_similoides

ASV 105	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	NA	NA
ASV 106	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Rhizosolenia	Rhizosolenia_delicatula
ASV 108	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Meuniera	Meuniera_membranacea
ASV 109	Alveolata	Ciliophora	Heterotrichea	Heterotrichea_X	Heterotrichida	Heterotrichida_X	Heterotrichida_X_sp.
ASV 110	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 111	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	Biecheleria_sp.
ASV 113	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Araphid-pennate	Delphineis	Delphineis_sp.
ASV 115	Hacrobia	Haptophyta	Prymnesiophyceae	Isochrysidales	Noelaerhabdaceae	Gephyrocapsa	Gephyrocapsa_oceanica
ASV 117	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_cordatum
ASV 118	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_dominans
ASV 119	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_B	Pseudotontonia	Pseudotontonia_simplicidens
ASV 122	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Protodinium	Protodinium_simplex
ASV 125	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Leptocylindrus	Leptocylindrus_convexus
ASV 126	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_Q	Strombidiidae_Q_X	Strombidiidae_Q_X_sp.
ASV 127	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
ASV 128	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	Biecheleria_cincta

ASV 129	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_socialis
ASV 130	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 131	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_cordatum
ASV 132	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 133	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_sp.
ASV 135	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 136	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Leptocylindrus	Leptocylindrus_minimus
ASV 137	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_I	Pelagostrobilidium	NA
ASV 138	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Balechina	Balechina_pachyderma
ASV 139	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_M	Strombidium_M	NA
ASV 140	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-G	Chrysophyceae_Clade-G_X	Chrysophyceae_Clade-G_X_sp.
ASV 143	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_B	Pseudotontonia	Pseudotontonia_simplicidens
ASV 144	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	Braarudosphaera	Braarudosphaera_bigelowii
ASV 145	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_spirale
ASV 146	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_sp.
ASV 147	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Cerataulina	Cerataulina_pelagica

ASV 148	Alveolata	Ciliophora	Ciliophora_X	Ciliophora_XX	Ciliophora_XXX	Mesodinium	Mesodinium_sp.
ASV 149	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_L	Strombidiidae_L_X	Strombidiidae_L_X_sp.
ASV 150	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-C	Chrysophyceae_Clade-C_X	Chrysophyceae_Clade-C_X_sp.
ASV 152	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 153	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	NA
ASV 154	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesiaceae_X	Prymnesiaceae_X_sp.
ASV 155	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 156	Stramenopiles	Sagenista	MAST-4	MAST-4A	MAST-4A_X	MAST-4A_XX	MAST-4A_XX_sp.
ASV 157	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_donghaiense
ASV 159	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystida_X	Pterocystida_XX	Pterocystida_XX_sp.
ASV 160	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Prymnesiophyceae_XX	Prymnesiophyceae_XXX	Prymnesiophyceae_XXX_sp.
ASV 161	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_Clade_D	Prymnesiophyceae_Clade_D_X	Prymnesiophyceae_Clade_D_XX	Prymnesiophyceae_Clade_D_XX_sp.
ASV 162	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_cordatum
ASV 163	Archaeplastida	Chlorophyta	Trebouxiophyceae	Chlorellales	Chlorellales_X	Picochlorum	Picochlorum_sp.
ASV 164	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
ASV 165	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.

ASV 168	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	CCW10-lineage	CCW10-lineage_X	CCW10-lineage_X_sp.
ASV 169	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesiaceae_X	Prymnesiaceae_X_sp.
ASV 170	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Chytriodiniaceae	Dissodinium	Dissodinium_pseudolunula
ASV 171	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 173	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	Botuliformidae	Botuliformidae_X	Botuliformidae_X_sp.
ASV 174	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	NA	NA
ASV 177	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 179	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gymnodinium	Gymnodinium_sp.
ASV 180	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 182	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_cristatum
ASV 183	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_sp.
ASV 184	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Minutocellus	Minutocellus_polymorphus
ASV 185	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Dino-Group-I-Clade-4_X	Dino-Group-I-Clade-4_X_sp.
ASV 186	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 187	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	Tripos_fusus
ASV 188	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.

ASV 189	Archaeplastida	Prasinodermophyta	Prasinodermophyceae	Prasinococcales	Prasinococcales-Clade-B	Prasinoderma	Prasinoderma_coloniale
ASV 190	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Minidiscus	Minidiscus_variabilis
ASV 191	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Heterocapsaceae	Heterocapsa	NA
ASV 192	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_bipes
ASV 193	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_sp_Clade_Na11C3
ASV 196	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_costatus
ASV 197	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	Pseudo-nitzschia_pungens
ASV 198	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.
ASV 199	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-6	Dino-Group-II-Clade-6_X	Dino-Group-II-Clade-6_X_sp.
ASV 201	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 202	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
ASV 203	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_danicus
ASV 204	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Scrippsiella	NA
ASV 205	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Prymnesiophyceae_XX	Prymnesiophyceae_XXX	Prymnesiophyceae_XXX_sp.
ASV 207	Hacrobia	Katablepharidophyta	Katablepharidaceae	Katablepharidales	Katablepharidales_X	Katablepharidales_XX	Katablepharidales_XX_sp.
ASV 208	Stramenopiles	Ochrophyta	MOCH-3	MOCH-3_X	MOCH-3_XX	MOCH-3_XXX	MOCH-3_XXX_sp.

ASV 209	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 210	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 211	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 212	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiales_X	Chrysocampanula	Chrysocampanula_spinifera
ASV 213	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Cryptomonadales_XX	Cryptomonadales_XX_sp.
ASV 214	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachnida	Chlorarachnida_X	NA	NA
ASV 216	Hacrobia	Haptophyta	Prymnesiophyceae	Syracosphaerales	Syracosphaerales_X	Syracosphaerales_XX	Syracosphaerales_XX_sp.
ASV 217	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pyramimonadaceae	Pyramimonas	Pyramimonas_sp.
ASV 218	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-5	Dino-Group-I-Clade-5_X	Dino-Group-I-Clade-5_X_sp.
ASV 219	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_I	Strobilidiidae_I_X	Strobilidiidae_I_X_sp.
ASV 220	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_dominans
ASV 221	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-F	Paraphysomonas	Paraphysomonas_sp.
ASV 222	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Pentapharsodinium	NA
ASV 223	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	NA
ASV 224	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_L	Strombidiidae_L_X	Strombidiidae_L_X_sp.
ASV 225	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.

ASV 226	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Cylindrotheca	Cylindrotheca_sp.
ASV 227	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Cerataulina	Cerataulina_pelagica
ASV 228	Alveolata	Ciliophora	Oligohymenophorea	Scuticociliatia_1	Philasterida	Porpostoma	Porpostoma_notata
ASV 229	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	Botuliformidae	Botuliformidae_X	Botuliformidae_X_sp.
ASV 230	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-III	Dino-Group-III_X	Dino-Group-III_XX	Dino-Group-III_XX_sp.
ASV 231	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gymnodinium	Gymnodinium_sp.
ASV 233	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 236	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 238	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	NA	NA	NA
ASV 240	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Lauderia	Lauderia_annulata
ASV 242	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 243	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-III	Dino-Group-III_X	Dino-Group-III_XX	Dino-Group-III_XX_sp.
ASV 244	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_B	Pseudotontonia	Pseudotontonia_sp.
ASV 245	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_sp.
ASV 249	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Protodinium	Protodinium_simplex
ASV 251	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	Pseudo-nitzschia_delicatissima

ASV 252	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Eduboscquella	Eduboscquella_crenulata
ASV 254	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-23	Dino-Group-II-Clade-23_X	Dino-Group-II-Clade-23_X_sp.
ASV 255	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Pelagodinium	Pelagodinium_beii
ASV 256	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas	Micromonas_pusilla
ASV 257	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Teleaulax	Teleaulax_gracilis
ASV 258	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pyramimonadales_X	Pyramimonadales_XX	Pyramimonadales_XX_sp.
ASV 259	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Guinardia	Guinardia_flaccida
ASV 260	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Thraustochytriaceae_X	Thraustochytriaceae_X_sp.
ASV 261	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 262	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gymnodinium	Gymnodinium_sp.
ASV 264	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-F	NA	NA
ASV 265	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-III	Dino-Group-III_X	Dino-Group-III_XX	Dino-Group-III_XX_sp.
ASV 266	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 267	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_protuberans
ASV 268	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 271	Archaeplastida	Chlorophyta	Chlorodendrophyceae	Chlorodendrales	Chlorodendraceae	Chlorodendrales_XX	Chlorodendrales_XX_sp.

ASV 273	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_divergens
ASV 274	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae	Parallelostrombidium	Parallelostrombidium_conicum
ASV 276	Archaeplastida	Chlorophyta	Mamiellophyceae	Dolichomastigales	Crustomastigaceae	Crustomastigaceae-AB	Crustomastigaceae-AB_sp.
ASV 277	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_B	Pseudotontonia	Pseudotontonia_sp.
ASV 278	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-F	Paraphysomonas	Paraphysomonas_imperforata
ASV 279	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Luciella	Luciella_sp.
ASV 280	Stramenopiles	Sagenista	MAST-7	MAST-7A	MAST-7A_X	MAST-7A_XX	MAST-7A_XX_sp.
ASV 281	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 282	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_decipiens
ASV 283	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 284	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-I	Chrysophyceae_Clade-I_X	Chrysophyceae_Clade-I_X_sp.
ASV 288	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 291	Stramenopiles	Ochrophyta	Pelagophyceae	Pelagomonadales	Pelagomonadaceae	Pelagomonadaceae_clade_C	Pelagomonadaceae_clade_C_sp.
ASV 292	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	NA
ASV 293	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	TAGIRI1-lineage	TAGIRI1-lineage_X	TAGIRI1-lineage_X_sp.
ASV 296	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Dactyliosolen	Dactyliosolen_blavyanus

ASV 297	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 298	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	NA	NA	NA
ASV 300	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_Clade_E	Prymnesiophyceae_Clade_E_X	Prymnesiophyceae_Clade_E_XX	Prymnesiophyceae_Clade_E_XX_sp.
ASV 301	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_I	Pelagostrobilidium	Pelagostrobilidium_neptuni
ASV 302	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_lorenzianus_2
ASV 304	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
ASV 305	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-IV	Dino-Group-IV-Syndinium-Group	Syndinium	Syndinium_sp.
ASV 308	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystidae	NA	NA
ASV 309	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	Oomycota_XX	NA	NA
ASV 310	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	NA	NA	NA
ASV 311	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 312	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-32	Dino-Group-II-Clade-32_X	Dino-Group-II-Clade-32_X_sp.
ASV 313	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachniophyceae_X	NPK2-lineage	NPK2-lineage_X	NPK2-lineage_X_sp.
ASV 314	Amoebozoa	Conosa	Variosea	Variosea_X	Mariager-Fjord-lineage	Mariager-Fjord-lineage_X	Mariager-Fjord-lineage_X_sp.
ASV 315	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_sp.
ASV 318	Stramenopiles	Ochrophyta	MOCH-1	MOCH-1_X	MOCH-1_XX	MOCH-1_XXX	MOCH-1_XXX_sp.

ASV 319	Stramenopiles	Sagenista	MAST-4	MAST-4B	MAST-4B_X	MAST-4B_XX	MAST-4B_XX_sp.
ASV 320	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 321	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Cylindrotheca	Cylindrotheca_sp.
ASV 322	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 323	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Proboscia	Proboscia_indica
ASV 325	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystida_X	Pterocystida_XX	Pterocystida_XX_sp.
ASV 326	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Cyclotella	NA
ASV 327	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Labyrinthuloides	Labyrinthuloides_minuta
ASV 328	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 329	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 330	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_brevis_3
ASV 331	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	NA	NA	NA
ASV 332	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gymnodinium	Gymnodinium_dorsalisulcum
ASV 333	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 334	Alveolata	Ciliophora	Spirotrichea	Strombidiida_G	Strombidiida_G_X	Strombidiida_G_XX	Strombidiida_G_XX_sp.
ASV 335	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Thraustochytriaceae_X	Thraustochytriaceae_X_sp.

ASV 336	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	Biecheleria_sp.
ASV 337	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Dicrateria	Dicrateria_rotunda
ASV 338	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	Biecheleria_sp.
ASV 339	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Mamiella	Mamiella_gilva
ASV 340	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 341	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	Braarudosphaeraceae_X	Braarudosphaeraceae_X_sp.
ASV 342	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Diplopsalidaceae	Diplopsalis	Diplopsalis_caspica
ASV 343	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 344	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 345	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_monovelum
ASV 346	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 347	Alveolata	Ciliophora	Spirotrichea	Tintinnida	Eutintinnidae	Eutintinnus	Eutintinnus_sp.
ASV 349	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Thraustochytriaceae_X	Thraustochytriaceae_X_sp.
ASV 350	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiophyceae_Clade_B4	Prymnesiophyceae_Clade_B4_X	Prymnesiophyceae_Clade_B4_X_sp.
ASV 351	Rhizaria	Cercozoa	Filosa-Imbricatea	Thaumatomonadida	NA	NA	NA
ASV 353	Stramenopiles	Sagenista	Labyrinthulomycetes	Labyrinthulales	Labyrinthulaceae	Aplanochytrium	Aplanochytrium_sp.

ASV 354	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachniophyceae_X	NPK2-lineage	NPK2-lineage_X	NPK2-lineage_X_sp.
ASV 356	Rhizaria	Cercozoa	Filosa-Imbricatea	Thaumatomonadida	Thaumatomonadidae	Thaumatomastix	Thaumatomastix_sp.
ASV 357	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karlodinium	NA
ASV 358	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	NA
ASV 361	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	RCC391	RCC391_sp.
ASV 362	Hacrobia	Katablepharidophyta	Katablepharidaceae	Katablepharidales	Katablepharidales_X	Katablepharidales_XX	Katablepharidales_XX_sp.
ASV 363	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	Biecheleria_sp.
ASV 364	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Halosphaeraceae	Halosphaera	Halosphaera_sp.
ASV 366	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_globosa
ASV 368	Archaeplastida	Chlorophyta	Trebouxiophyceae	Chlorellales	Chlorellales_X	Picochlorum	Picochlorum_sp.
ASV 369	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_didymus_2
ASV 372	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	RCC391	RCC391_sp.
ASV 374	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_sp.
ASV 376	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 377	Rhizaria	Cercozoa	Filosa-Imbricatea	Euglyphida	Paulinellidae	Paulinella	Paulinella_sp.
ASV 378	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	NA	NA	NA

ASV 379	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 380	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 381	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_lorenzianus_2
ASV 382	Stramenopiles	Sagenista	MAST-7	MAST-7B	MAST-7B_X	MAST-7B_XX	MAST-7B_XX_sp.
ASV 383	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 384	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Minidiscus	Minidiscus_comicus
ASV 385	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_L	Strombidiidae_L_X	Strombidiidae_L_X_sp.
ASV 386	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	Pseudo-nitzschia_sp.
ASV 388	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 390	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 391	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-8	Dino-Group-II-Clade-8_X	Dino-Group-II-Clade-8_X_sp.
ASV 394	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 397	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 398	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Guinardia	Guinardia_delicatula
ASV 399	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 400	Stramenopiles	Ochrophyta	Bolidophyceae	Parmales	Triparmaceae	Triparma	Triparma_laevis_clade

ASV 401	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Teleaulax	Teleaulax_acuta
ASV 402	Alveolata	Ciliophora	Heterotrichea	Heterotrichea_X	NA	NA	NA
ASV 403	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Guinardia	Guinardia_flaccida
ASV 404	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Choreotrichida_X	Choreotrichida_XX	Choreotrichida_XX_sp.
ASV 405	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	NA	NA	NA
ASV 408	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Choreotrichida_X	Choreotrichida_XX	Choreotrichida_XX_sp.
ASV 409	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_tenuissimus
ASV 411	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Haptolina	Haptolina_sp.
ASV 412	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 413	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
ASV 414	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 415	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_sp.
ASV 416	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventricleftida_X	Ventricleftida_XX	Ventricleftida_XX_sp.
ASV 417	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 418	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-7	Dino-Group-II-Clade-7_X	Dino-Group-II-Clade-7_X_sp.
ASV 419	Alveolata	Ciliophora	Spirotrichea	Strombidiida_G	Strombidiida_G_X	Strombidiida_G_XX	Strombidiida_G_XX_sp.

ASV 420	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 421	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	Cryothecomonas_aestivalis
ASV 422	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 423	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_rotula
ASV 424	Alveolata	Ciliophora	Spirotrichea	Strombidiida	NA	NA	NA
ASV 425	Hacrobia	Haptophyta	Prymnesiophyceae	Calcihaptophycidae	Rhabdosphaeraceae	Algirosphaera	Algirosphaera_robusta
ASV 426	Alveolata	Ciliophora	Litostomatea	Litostomatea_X	Litostomatea_XX	Litostomatea_XXX	Litostomatea_XXX_sp.
ASV 427	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	NA	NA	NA
ASV 428	Archaeplastida	Chlorophyta	Mamiellophyceae	Dolichomastigales	Crustomastigaceae	Crustomastigaceae_X	Crustomastigaceae_X_sp.
ASV 429	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Eduboscquella	Eduboscquella_crenulata
ASV 430	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 432	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 433	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Teleaulax	Teleaulax_sp.
ASV 435	Alveolata	Dinoflagellata	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum	Prorocentrum_sp.
ASV 437	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Filosa-Thecofilosea_XX	Filosa-Thecofilosea_XXX	Filosa-Thecofilosea_XXX_sp.
ASV 438	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.

ASV 439	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karlodinium	Karlodinium_sp.
ASV 441	Stramenopiles	Sagenista	MAST-7	MAST-7B	MAST-7B_X	MAST-7B_XX	MAST-7B_XX_sp.
ASV 442	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 443	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	NA	NA	NA
ASV 444	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 445	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	Braarudosphaera	Braarudosphaera_bigelowii
ASV 447	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_B	Pseudotontonia	Pseudotontonia_sp.
ASV 448	Archaeplastida	Chlorophyta	Mamiellophyceae	Dolichomastigales	Crustomastigaceae	Crustomastix	Crustomastix_didyma
ASV 449	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Pelagodinium	Pelagodinium_beii
ASV 450	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Protodinium	Protodinium_simplex
ASV 451	Rhizaria	Cercozoa	Filosa-Imbricatea	Thaumatomonadida	Thaumatomonadidae	Reckertia	Reckertia_filosa
ASV 452	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-7	Dino-Group-II-Clade-7_X	Dino-Group-II-Clade-7_X_sp.
ASV 453	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 456	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 457	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_sp.

ASV 458	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Filosa-Thecofilosea_XX	Filosa-Thecofilosea_XXX	Filosa-Thecofilosea_XXX_sp.
ASV 459	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 460	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	Botuliformidae	Botuliformidae_X	Botuliformidae_X_sp.
ASV 462	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	NA	NA	NA
ASV 463	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Thraustochytriaceae_X	Thraustochytriaceae_X_sp.
ASV 464	Archaeplastida	Chlorophyta	Nephroselmidophyceae	Nephroselmidales	Nephroselmidales_X	Nephroselmis	Nephroselmis_pyriformis
ASV 465	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	Bathycoccus_prasinos
ASV 466	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Amphidomataceae	Azadinium	NA
ASV 467	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Mataza-lineage	Mataza-lineage_X	Mataza-lineage_X_sp.
ASV 468	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	Dino-Group-II_XX	Dino-Group-II_XX_sp.
ASV 469	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-C	Chrysophyceae_Clade-C_X	Chrysophyceae_Clade-C_X_sp.
ASV 470	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	Tripos_furca
ASV 472	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventricleftida_X	Ventricleftida_XX	Ventricleftida_XX_sp.
ASV 473	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-III	Dino-Group-III_X	Dino-Group-III_XX	Dino-Group-III_XX_sp.
ASV 474	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	Dino-Group-II_XX	Dino-Group-II_XX_sp.
ASV 475	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.

ASV 476	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 477	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 478	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_danicus
ASV 479	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Thoracosphaeraceae_X	Thoracosphaeraceae_X_sp.
ASV 481	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 482	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-F	Paraphysomonas	NA
ASV 483	Hacrobia	Katablepharidophyta	Katablepharidaceae	Katablepharidales	Katablepharidales_X	Katablepharidales_XX	Katablepharidales_XX_sp.
ASV 484	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 485	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	NA	NA
ASV 486	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karlodinium	Karlodinium_veneficum
ASV 487	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 489	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.
ASV 490	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 491	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karlodinium	Karlodinium_veneficum
ASV 492	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karlodinium	NA

ASV 493	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 494	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Dino-Group-I-Clade-4_X	Dino-Group-I-Clade-4_X_sp.
ASV 495	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 496	Stramenopiles	Ochrophyta	Pelagophyceae	Pelagomonadales	Pelagomonadaceae	Aureococcus	Aureococcus_anophagefferens
ASV 497	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Nitzschia	NA
ASV 499	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-3	Dino-Group-I-Clade-3_X	Dino-Group-I-Clade-3_X_sp.
ASV 500	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 501	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_Q	Strombidiidae_Q_X	Strombidiidae_Q_X_sp.
ASV 502	Alveolata	Ciliophora	Nassophorea	Nassophorea_X	Discotrichidae	NASSO_1	NASSO_1_sp.
ASV 504	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Apocalathium	Apocalathium_aciculiferum
ASV 505	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 506	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	NA
ASV 507	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa	NA
ASV 508	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 509	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 510	Stramenopiles	Ochrophyta	MOCH-1	MOCH-1_X	MOCH-1_XX	MOCH-1_XXX	MOCH-1_XXX_sp.

ASV 512	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
ASV 514	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	NA	NA
ASV 515	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 516	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_sp.
ASV 520	Hacrobia	Haptophyta	Prymnesiophyceae	Calcihaptophycidae	Watznaueriaceae	Tergestiella	Tergestiellaadriatica
ASV 521	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_sp.
ASV 523	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 525	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gymnodinium	NA
ASV 527	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-III	Dino-Group-III_X	Dino-Group-III_XX	Dino-Group-III_XX_sp.
ASV 528	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 529	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 532	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	NA
ASV 533	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 534	Stramenopiles	Sagenista	MAST-7	MAST-7C	MAST-7C_X	MAST-7C_XX	MAST-7C_XX_sp.
ASV 536	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 537	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karlodinium	Karlodinium_sp.

ASV 538	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	NA
ASV 539	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 541	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae	NA	NA
ASV 542	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 543	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_Clade_D	Prymnesiophyceae_Clade_D_X	Prymnesiophyceae_Clade_D_XX	Prymnesiophyceae_Clade_D_XX_sp.
ASV 544	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 545	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Pyrophacaceae	Fragilidium	NA
ASV 546	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 547	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiophyceae_Clade_B5	Prymnesiophyceae_Clade_B5_X	Prymnesiophyceae_Clade_B5_X_sp.
ASV 548	Hacrobia	Haptophyta	Prymnesiophyceae	Calcihaptophycidae	NA	NA	NA
ASV 549	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strombidinopsidae	Strombidinopsis	Strombidinopsis_sp.
ASV 550	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-44	Dino-Group-II-Clade-44_X	Dino-Group-II-Clade-44_X_sp.
ASV 551	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_minuscula
ASV 554	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	NA	NA	NA
ASV 555	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 556	Rhizaria	Cercozoa	Filosa	Filosa_X	Filosa_XX	Discomonas	Discomonas_retusa

ASV 557	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_sp.
ASV 558	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 559	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	NA	NA
ASV 561	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	NA
ASV 563	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 564	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_globosa
ASV 565	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiophyceae_Clade_B4	Prymnesiophyceae_Clade_B4_X	Prymnesiophyceae_Clade_B4_X_sp.
ASV 566	Rhizaria	Cercozoa	Filosa-Sarcomonadea	Cercomonadida	Cercomonadidae	Cercomonas	NA
ASV 567	Stramenopiles	Ochrophyta	MOCH-2	MOCH-2_X	MOCH-2_XX	MOCH-2_XXX	MOCH-2_XXX_sp.
ASV 568	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 571	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Heterocapsaceae	Heterocapsa	NA
ASV 572	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 573	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	NA	NA
ASV 574	Rhizaria	Radiolaria	Acantharea	Acantharea_X	Acantharea_XX	Trizona	Trizona_brandti
ASV 575	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 576	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.

ASV 577	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_B	Strombidiidae_B_X	Strombidiidae_B_X_sp.
ASV 578	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 579	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 580	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	Biecheleria_sp.
ASV 581	Stramenopiles	Stramenopiles_X	Stramenopiles_XX	Stramenopiles_XXX	Stramenopiles_XXXX	Stramenopiles_XXXXX	Stramenopiles_XXXXX_sp.
ASV 583	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	NA	NA	NA
ASV 584	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachniophyceae_X	NPK2-lineage	NPK2-lineage_X	NPK2-lineage_X_sp.
ASV 585	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Plagioselmis	Plagioselmis_prolonga
ASV 587	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachnida	Chlorarachnida_X	Chlorarachnida_XX	Chlorarachnida_XX_sp.
ASV 588	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Araphid-pennate	Thalassionema	Thalassionema_sp.
ASV 589	Stramenopiles	Ochrophyta	Bolidophyceae	Parmales	Triparmaceae	Triparma	Triparma_pacifica
ASV 591	Stramenopiles	Ochrophyta	Dictyochophyceae	Dictyochophyceae_X	Dictyochales	Dictyochales_X	Dictyochales_X_sp.
ASV 592	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 594	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 595	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	TAGIRI1-lineage	TAGIRI1-lineage_X	TAGIRI1-lineage_X_sp.
ASV 597	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pseudo-nitzschia	Pseudo-nitzschia_galaxiae

ASV 599	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 600	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Minidiscus	Minidiscus_comicus
ASV 601	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 602	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 604	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-H	Chrysophyceae_Clade-H_X	Chrysophyceae_Clade-H_X_sp.
ASV 605	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_sp.
ASV 606	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.
ASV 607	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 608	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_tenuissimus
ASV 610	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 611	Hacrobia	Haptophyta	Prymnesiophyceae	Calcihaptophycidae	Calcihaptophycidae_X	Calcihaptophycidae_XX	Calcihaptophycidae_XX_sp.
ASV 613	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	NA	NA
ASV 614	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachnida	Minorisa-lineage	Minorisa-lineage_X	Minorisa-lineage_X_sp.
ASV 616	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_sp_Clade_Na13C1
ASV 617	Alveolata	Ciliophora	Spirotrichea	Tintinnida	Eutintinnidae	Eutintinnidae_X	Eutintinnidae_X_sp.
ASV 619	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Warnowiaceae	Warnowia	Warnowia_sp.

ASV 620	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 622	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 623	Alveolata	Ciliophora	Spirotrichea	Strombidiida	NA	NA	NA
ASV 624	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	Braarudosphaera	Braarudosphaera_bigelowii
ASV 628	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	NA
ASV 629	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiophyceae_Clade_B3	Prymnesiophyceae_Clade_B3_X	Prymnesiophyceae_Clade_B3_X_sp.
ASV 630	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Brockmanniella	Brockmanniella brockmannii
ASV 631	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 632	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
ASV 633	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 634	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystida_X	Pterocystida_XX	Pterocystida_XX_sp.
ASV 635	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_americanum
ASV 637	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 639	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Rhizosolenia	Rhizosolenia_robusta
ASV 640	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Mataza-lineage	Mataza-lineage_X	Mataza-lineage_X_sp.
ASV 642	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.

ASV 643	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 645	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 646	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 647	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 648	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-55	Dino-Group-II-Clade-55_X	Dino-Group-II-Clade-55_X_sp.
ASV 650	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Thoracosphaeraceae_X	Thoracosphaeraceae_X_sp.
ASV 651	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Cylindrotheca	Cylindrotheca_closterium
ASV 652	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 653	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_cristatum
ASV 655	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Pelagodinium	Pelagodinium_beii
ASV 656	Stramenopiles	Opalozoa	MAST-3	MAST-3I	MAST-3I_X	MAST-3I_XX	MAST-3I_XX_sp.
ASV 657	Stramenopiles	Pseudofungi	MAST-1	MAST-1B	MAST-1B_X	MAST-1B_XX	MAST-1B_XX_sp.
ASV 659	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Mantoniella	NA
ASV 660	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.
ASV 661	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	NA	NA	NA
ASV 662	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.

ASV 663	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 664	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 665	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Proboscia	Proboscia_alata
ASV 667	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	Cryothecomonas_sp.
ASV 669	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_decipiens
ASV 670	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-5	Dino-Group-I-Clade-5_X	Dino-Group-I-Clade-5_X_sp.
ASV 671	Hacrobia	Centroheliozoa	Centroheliozoa_X	Centroheliozoa_XX	Centroheliozoa_XXX	Centroheliozoa_XXXX	Centroheliozoa_XXXX_sp.
ASV 672	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 673	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-H	Chrysophyceae_Clade-H_X	Chrysophyceae_Clade-H_X_sp.
ASV 675	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 676	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 678	Stramenopiles	Opalozoa	Bicoecea	Bicoecales	Bicoecaceae	Bicoecaceae_X	Bicoecaceae_X_sp.
ASV 682	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	NA	NA
ASV 683	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 684	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 685	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.

ASV 686	Stramenopiles	Pseudofungi	MAST-2	MAST-2B	MAST-2B_X	MAST-2B_XX	MAST-2B_XX_sp.
ASV 687	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-30	Dino-Group-II-Clade-30_X	Dino-Group-II-Clade-30_X_sp.
ASV 689	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	NA
ASV 691	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 694	Alveolata	Ciliophora	Phyllopharyngea	Cyrtophoria_1	PHYLL_4	PHYLL_4_X	PHYLL_4_X_sp.
ASV 696	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	Dino-Group-II-Clade-52_X	Dino-Group-II-Clade-52_X_sp.
ASV 697	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	NA	NA
ASV 698	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Guinardia	Guinardia_striata
ASV 699	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 700	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 702	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-6	Dino-Group-II-Clade-6_X	Dino-Group-II-Clade-6_X_sp.
ASV 704	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Haptolina	Haptolina_sp.
ASV 706	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Mataza-lineage	Mataza-lineage_X	Mataza-lineage_X_sp.
ASV 707	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 708	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 709	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.

ASV 710	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_decipiens
ASV 711	Stramenopiles	Opalozoa	MAST-3	MAST-3I	MAST-3I_X	MAST-3I_XX	MAST-3I_XX_sp.
ASV 712	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Leegaardiellidae_B	Leegaardiella	Leegaardiella_sp.
ASV 713	Archaeplastida	Chlorophyta	Ulvophyceae	Ulvales-relatives	Ulvales-relatives_X	Ulvales-relatives_XX	Ulvales-relatives_XX_sp.
ASV 714	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae	NA	NA
ASV 715	Archaeplastida	Streptophyta	Embryophyceae	Embryophyceae_X	Embryophyceae_XX	Silene	Silene_latifolia
ASV 716	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachniophyceae_X	NPK2-lineage	NPK2-lineage_X	NPK2-lineage_X_sp.
ASV 717	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	NA
ASV 718	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 719	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Eduboscquella	Eduboscquella_crenulata
ASV 721	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	Tripos_fusus
ASV 722	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_sp.
ASV 723	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Skeletonema	Skeletonema_pseudocostatum
ASV 724	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 725	Alveolata	Ciliophora	Spirotrichea	Strombidiida	NA	NA	NA
ASV 727	Alveolata	Ciliophora	Spirotrichea	Strombidiida_F	Strombidiida_F_X	Strombidiida_F_XX	Strombidiida_F_XX_sp.

ASV 729	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 730	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 731	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Leptocylindrus	Leptocylindrus_minimus
ASV 732	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 733	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 734	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Tontoniidae_B	Pseudotontonia	Pseudotontonia_simplicidens
ASV 735	Alveolata	Ciliophora	CONTH_8	CONTH_8_X	CONTH_8_XX	CONTH_8_XXX	CONTH_8_XXX_sp.
ASV 736	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strombidinopsidae	Strombidinopsis	Strombidinopsis_acuminata
ASV 739	Hacrobia	Haptophyta	Haptophyta_Clade_HAP3	Haptophyta_Clade_HAP3_X	Haptophyta_Clade_HAP3_XX	Haptophyta_Clade_HAP3_XXX	Haptophyta_Clade_HAP3_XXX_sp.
ASV 740	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	Cryothecomonas_aestivalis
ASV 741	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	NA	NA	NA
ASV 743	Alveolata	Ciliophora	Phyllopharyngea	Cyrtophoria_1	PHYLL_4	PHYLL_4_X	PHYLL_4_X_sp.
ASV 745	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 746	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	Tripos_petersenii
ASV 747	Alveolata	Ciliophora	Spirotrichea	Strombidiida	NA	NA	NA
ASV 749	Alveolata	Ciliophora	Phyllopharyngea	Cyrtophoria_1	NA	NA	NA

ASV 750	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 751	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 752	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 753	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.
ASV 754	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_sp.
ASV 755	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Lepidodinium	Lepidodinium_chlorophorum
ASV 756	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Novel-Clade-4	Novel-Clade-4_X	Novel-Clade-4_X_sp.
ASV 758	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Filosa-Thecofilosea_XX	Filosa-Thecofilosea_XXX	Filosa-Thecofilosea_XXX_sp.
ASV 759	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 760	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_L	Strombidiidae_L_X	Strombidiidae_L_X_sp.
ASV 761	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae	NA	NA
ASV 763	Alveolata	Ciliophora	Oligohymenophorea	Oligohymenophorea_X	Oligohymenophorea_XX	NA	NA
ASV 764	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_R	Strombidium_R	Strombidium_R_sp.
ASV 765	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
ASV 766	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	Dino-Group-II_XX	Dino-Group-II_XX_sp.

ASV 767	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 768	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Filosa-Thecofilosea_XX	Filosa-Thecofilosea_XXX	Filosa-Thecofilosea_XXX_sp.
ASV 770	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_affinis
ASV 772	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 774	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pleurosigma	Pleurosigma_sp.
ASV 776	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 777	Hacrobia	Katablepharidophyta	Katablepharidaceae	Katablepharidales	Katablepharidales_X	Katablepharidales_XX	Katablepharidales_XX_sp.
ASV 780	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	Pterosperma_sp.
ASV 781	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_sp.
ASV 782	Stramenopiles	Sagenista	Labyrinthulomycetes	Labyrinthulales	Labyrinthulaceae	Labyrinthulaceae_X	Labyrinthulaceae_X_sp.
ASV 783	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_tenera
ASV 784	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 785	Hacrobia	Haptophyta	Prymnesiophyceae	Coccolithales	Coccolithaceae	Coccolithus	NA
ASV 786	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Actinocyclus	Actinocyclus_sp.
ASV 787	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	NA	NA
ASV 788	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-8	Dino-Group-I-Clade-8_X	Dino-Group-I-Clade-8_X_sp.

ASV 789	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 790	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
ASV 792	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 793	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 794	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 795	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.
ASV 796	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-H	Chrysophyceae_Clade-H_X	Chrysophyceae_Clade-H_X_sp.
ASV 797	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strombidinopsidae	Strombidinopsis	Strombidinopsis_sp.
ASV 799	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 800	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 803	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 804	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 806	Alveolata	Ciliophora	Nassophorea	Nassophorea_X	Discotrichidae	NA	NA
ASV 807	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 808	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	NA	NA	NA

ASV 809	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Thoracosphaeraceae_X	Thoracosphaeraceae_X_sp.
ASV 810	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Leegaardiellidae_A	Leegaardiellidae_A_X	Leegaardiellidae_A_X_sp.
ASV 812	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 813	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiophyceae_Clade_B4	Prymnesiophyceae_Clade_B4_X	Prymnesiophyceae_Clade_B4_X_sp.
ASV 814	Stramenopiles	Ochrophyta	MOCH-2	MOCH-2_X	MOCH-2_XX	MOCH-2_XXX	MOCH-2_XXX_sp.
ASV 815	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	Tripos_concilians
ASV 816	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 817	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_sp.
ASV 818	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Haptolina	Haptolina_sp.
ASV 819	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Dino-Group-I-Clade-4_X	Dino-Group-I-Clade-4_X_sp.
ASV 820	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 822	Alveolata	Dinoflagellata	Dinophyceae	Dinophysiales	Dinophysiaceae	Dinophysis	Dinophysis_acuminata
ASV 823	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 825	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_cordata
ASV 826	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 828	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.

ASV 830	Archaeplastida	Chlorophyta	Prasino-Clade-V	Pseudoscourfieldiales	Pycnococcaceae	NA	NA
ASV 831	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_tenera
ASV 833	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Dino-Group-I-Clade-4_X	Dino-Group-I-Clade-4_X_sp.
ASV 834	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	NA
ASV 835	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 836	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Balechina	Balechina_pachyderma ta
ASV 837	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Podolampadaceae	Lessardia	Lessardia_elongata
ASV 838	Stramenopiles	Sagenista	MAST-8	MAST-8B	MAST-8B_X	MAST-8B_XX	MAST-8B_XX_sp.
ASV 839	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Corethron	Corethron_hystrix
ASV 840	Hacrobia	Haptophyta	Prymnesiophyceae	Syracosphaerales	Syracosphaerales_X	Syracosphaerales_XX	Syracosphaerales_XX_sp.
ASV 841	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 843	Alveolata	Dinoflagellata	Dinophyceae	Dinophysiales	Amphisoleniaceae	Amphisoleniaceae_X	Amphisoleniaceae_X_sp.
ASV 844	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 847	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Filosa-Thecofilosea_XX	Filosa-Thecofilosea_XXX	Filosa-Thecofilosea_XXX_sp.
ASV 848	Hacrobia	Haptophyta	Prymnesiophyceae	Calcihaptophycidae	Rhabdosphaeraceae	Algirosphaera	Algirosphaera_robusta
ASV 849	Archaeplastida	Prasinodermophyta	Prasinodermophyceae	Prasinococcales	Prasinococcales-Clade-B	Prasinococcales-Clade-B_X	Prasinococcales-Clade-B1_X_sp.

ASV 851	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_decipiens
ASV 852	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Mataza-lineage	Mataza-lineage_X	Mataza-lineage_X_sp.
ASV 853	Hacrobia	Haptophyta	Haptophyta_Clade_HAP3	Haptophyta_Clade_HAP3_X	Haptophyta_Clade_HAP3_XX	Haptophyta_Clade_HAP3_XXX	Haptophyta_Clade_HAP3_XXX_sp.
ASV 854	Stramenopiles	Stramenopiles_X	Stramenopiles_XX	Stramenopiles_XXX	Stramenopiles_XXXX	Stramenopiles_XXXXX	Stramenopiles_XXXXX_sp.
ASV 855	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 856	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 857	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 858	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_L	Strombidiidae_L_X	Strombidiidae_L_X_sp.
ASV 859	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
ASV 861	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 862	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	Dino-Group-II_XX	Dino-Group-II_XX_sp.
ASV 863	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-30	Dino-Group-II-Clade-30_X	Dino-Group-II-Clade-30_X_sp.
ASV 866	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 867	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae	NA	NA
ASV 868	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 869	Alveolata	Ciliophora	Spirotrichea	Strombidiida	NA	NA	NA

ASV 870	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Heterocapsaceae	Heterocapsa	NA
ASV 871	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Thraustochytriaceae_X	Thraustochytriaceae_X_sp.
ASV 872	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 875	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	TAGIRI1-lineage	TAGIRI1-lineage_X	TAGIRI1-lineage_X_sp.
ASV 876	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 877	Hacrobia	Haptophyta	Haptophyta_X	Haptophyta_XX	Haptophyta_XXX	Haptophyta_XXXX	Haptophyta_XXXX_sp.
ASV 881	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 882	Alveolata	Ciliophora	CONTH_8	CONTH_8_X	CONTH_8_XX	CONTH_8_XXX	CONTH_8_XXX_sp.
ASV 883	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventricleftida_X	Ventricleftida_XX	Ventricleftida_XX_sp.
ASV 884	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-22	Dino-Group-II-Clade-22_X	Dino-Group-II-Clade-22_X_sp.
ASV 886	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 887	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 889	Archaeplastida	Chlorophyta	Mamiellophyceae	Dolichomastigales	Crustomastigaceae	Crustomastigaceae-AB	Crustomastigaceae-AB_sp.
ASV 890	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-5	Dino-Group-I-Clade-5_X	Dino-Group-I-Clade-5_X_sp.
ASV 892	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-39	Dino-Group-II-Clade-39_X	Dino-Group-II-Clade-39_X_sp.
ASV 895	Rhizaria	Radiolaria	Acantharea	Chaunacanthida	Chaunacanthida_X	Acanthometron	Acanthometron_sp.

ASV 896	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Mataza-lineage	Mataza-lineage_X	Mataza-lineage_X_sp.
ASV 897	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 898	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_elegans
ASV 899	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_leadbeateri
ASV 902	Alveolata	Ciliophora	Spirotrichea	Strombidiida	NA	NA	NA
ASV 903	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 905	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Guinardia	Guinardia_striata
ASV 906	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystida_X	Pterocystida_XX	Pterocystida_XX_sp.
ASV 908	Hacrobia	Haptophyta	Haptophyta_Clade_HAP3	Haptophyta_Clade_HAP3_X	Haptophyta_Clade_HAP3_XX	Haptophyta_Clade_HAP3_XXX	Haptophyta_Clade_HAP3_XXX_sp.
ASV 909	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 910	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Paragymnodinium	Paragymnodinium_shiwhaense
ASV 912	Stramenopiles	Ochrophyta	Pelagophyceae	Pelagomonadales	Pelagomonadales_clade_A	NA	NA
ASV 913	Stramenopiles	Opalozoa	Bicoecea	Bicoecales	Bicoecaceae	Bicoecaceae_X	Bicoecaceae_X_sp.
ASV 916	Hacrobia	Katablepharidophyta	Katablepharidaceae	Katablepharidales	Katablepharidales_X	Katablepharidales_XX	Katablepharidales_XX_sp.
ASV 920	Hacrobia	Haptophyta	Haptophyta_X	Haptophyta_XX	Haptophyta_XXX	Haptophyta_XXXX	Haptophyta_XXXX_sp.
ASV 921	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiophyceae_Clade_B4	Prymnesiophyceae_Clade_B4_X	Prymnesiophyceae_Clade_B4_X_sp.

ASV 922	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachniophyceae_X	NPK2-lineage	NPK2-lineage_X	NPK2-lineage_X_sp.
ASV 923	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 924	Alveolata	Dinoflagellata	Noctilucophyceae	Noctilucales	Noctilucaceae	Spatulodinium	Spatulodinium_pseudonoctiluca
ASV 925	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 926	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Proboscia	Proboscia_alata
ASV 927	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	NA
ASV 928	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 930	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 932	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Proboscia	NA
ASV 934	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-47	Dino-Group-II-Clade-47_X	Dino-Group-II-Clade-47_X_sp.
ASV 935	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_concaviuscula
ASV 936	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 939	Stramenopiles	Stramenopiles_X	Stramenopiles_XX	Stramenopiles_XXX	Stramenopiles_XXXX	Stramenopiles_XXXXX	Stramenopiles_XXXXX_sp.
ASV 940	Alveolata	Ciliophora	Spirotrichea	Strombidiida_D	Strombidiida_D_X	Strombidiida_D_XX	Strombidiida_D_XX_sp.
ASV 943	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	NA
ASV 944	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	NA

ASV 947	Stramenopiles	Sagenista	MAST-4	MAST-4E	MAST-4E_X	MAST-4E_XX	MAST-4E_XX_sp.
ASV 948	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gymnodinium	Gymnodinium_dorsalisulcum
ASV 949	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	Biecheleria_sp.
ASV 952	Stramenopiles	Ochrophyta	Bolidophyceae	Parmales	Parmales_env_3	Parmales_env_3_X	Parmales_env_3_X_sp.
ASV 953	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 954	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Diplopsalidaceae	Gotoius	Gotoius_excentricus
ASV 955	Stramenopiles	Opalozoa	MAST-3	MAST-3I	MAST-3I_X	MAST-3I_XX	MAST-3I_XX_sp.
ASV 956	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 957	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	Peronosporales	NA	NA
ASV 958	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_elegans
ASV 959	Stramenopiles	Ochrophyta	MOCH-3	MOCH-3_X	MOCH-3_XX	MOCH-3_XXX	MOCH-3_XXX_sp.
ASV 960	Stramenopiles	Sagenista	Labyrinthulomycetes	Labyrinthulales	Labyrinthulaceae	NA	NA
ASV 964	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_lauderi
ASV 968	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_sp.
ASV 969	Hacrobia	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis_sp.
ASV 971	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-57	Dino-Group-II-Clade-57_X	Dino-Group-II-Clade-57_X_sp.

ASV 973	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 975	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 979	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	Prymnesium_sp.
ASV 980	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 982	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	NA	NA
ASV 984	Stramenopiles	Opalozoa	MAST-3	MAST-3D	MAST-3D_X	MAST-3D_XX	MAST-3D_XX_sp.
ASV 985	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 986	Alveolata	Ciliophora	Ciliophora_X	Ciliophora_XX	Ciliophora_XXX	Mesodinium	Mesodinium_sp.
ASV 987	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	Peronosporales	NA	NA
ASV 989	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	NA	NA
ASV 991	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 992	Alveolata	Ciliophora	Colpodea	Colpodea_X	Cyrtolophosidida	Aristerostoma	Aristerostoma_sp.
ASV 994	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-H	Chrysophyceae_Clade-H_X	Chrysophyceae_Clade-H_X_sp.
ASV 1000	Stramenopiles	Stramenopiles_X	Stramenopiles_XX	Stramenopiles_XXX	Stramenopiles_XXXX	Stramenopiles_XXXXX	Stramenopiles_XXXXX_sp.
ASV 1001	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 1006	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.

ASV 1007	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-42	Dino-Group-II-Clade-42_X	Dino-Group-II-Clade-42_X_sp.
ASV 1008	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.
ASV 1010	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 1011	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Ceratiaceae	Tripos	Tripos_furca
ASV 1012	Stramenopiles	Opalozoa	MAST-3	MAST-3I	MAST-3I_X	MAST-3I_XX	MAST-3I_XX_sp.
ASV 1014	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_cf_tortissimus
ASV 1015	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 1016	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 1019	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Mataza-lineage	Mataza-lineage_X	Mataza-lineage_X_sp.
ASV 1024	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 1026	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 1027	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 1031	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 1032	Rhizaria	Cercozoa	Filosa-Imbricatea	Thaumatomonadida	NA	NA	NA
ASV 1035	Archaeplastida	Prasinodermophyta	Prasinodermophyceae	Prasinococcales	Prasinococcales-Clade-B	Prasinoderma	Prasinoderma_sp.
ASV 1038	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Pleurosigma	Pleurosigma_sp.

ASV 1042	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_pseudocurvisetus
ASV 1044	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-1	Telonemia-Group-1_X	Telonemia-Group-1_X_sp.
ASV 1046	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 1047	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 1048	Hacrobia	Haptophyta	Haptophyta_Clade_HAP3	Haptophyta_Clade_HAP3_X	Haptophyta_Clade_HAP3_XX	Haptophyta_Clade_HAP3_XXX	Haptophyta_Clade_HAP3_XXX_sp.
ASV 1050	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_I	Strombidiidae_I_X	Strombidiidae_I_X_sp.
ASV 1054	Stramenopiles	Sagenista	Labyrinthulomycetes	Labyrinthulales	Labyrinthulaceae	Labyrinthulaceae_X	Labyrinthulaceae_X_sp.
ASV 1055	Alveolata	Ciliophora	Spirotrichea	Tintinnida	NA	NA	NA
ASV 1057	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 1067	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulacales	Goniodomataceae	Alexandrium	Alexandrium_hiranoi
ASV 1068	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 1072	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 1075	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-32	Dino-Group-II-Clade-32_X	Dino-Group-II-Clade-32_X_sp.
ASV 1076	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	NA	NA	NA
ASV 1083	Stramenopiles	Ochrophyta	MOCH-3	MOCH-3_X	MOCH-3_XX	MOCH-3_XXX	MOCH-3_XXX_sp.
ASV 1085	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-2	Telonemia-Group-2_X	Telonemia-Group-2_X_sp.

ASV 1086	Rhizaria	Cercozoa	Filosa-Imbricatea	Filosa-Imbricatea_X	Novel-clade-2	Novel-clade-2_X	Novel-clade-2_X_sp.
ASV 1089	Stramenopiles	Opalozoa	MAST-3	MAST-3I	MAST-3I_X	MAST-3I_XX	MAST-3I_XX_sp.
ASV 1090	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_I	Strombidiidae_I_X	Strombidiidae_I_X_sp.
ASV 1092	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Pelagodinium	Pelagodinium_beii
ASV 1093	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_brevis_3
ASV 1095	Alveolata	Ciliophora	CONThreeP	CONThreeP_X	CONThreeP_XX	Askenasia	Askenasia_sp.
ASV 1096	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 1100	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 1101	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	NA	NA
ASV 1102	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Thoracosphaeraceae	Pentapharsodinium	Pentapharsodinium_tyrrhenicum
ASV 1105	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	TAGIRI1-lineage	TAGIRI1-lineage_X	TAGIRI1-lineage_X_sp.
ASV 1107	Alveolata	Apicomplexa	Gregarinomorphea	Gregarines_GRE2	Gregarines_GRE2_X	Gregarines_GRE2_XX	Gregarines_GRE2_XX_sp.
ASV 1110	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-10-and-11	Dino-Group-II-Clade-10-and-11_X	Dino-Group-II-Clade-10-and-11_X_sp.
ASV 1111	Archaeplastida	Chlorophyta	Mamiellophyceae	Dolichomastigales	Crustomastigaceae	Crustomastigaceae-AB	Crustomastigaceae-AB_sp.
ASV 1112	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karlodinium	NA
ASV 1113	Stramenopiles	Sagenista	MAST-7	MAST-7_X	MAST-7_XX	MAST-7_XXX	MAST-7_XXX_sp.

ASV 1114	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 1115	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ebriida	TAGIRI1-lineage	TAGIRI1-lineage_X	TAGIRI1-lineage_X_sp.
ASV 1118	Alveolata	Ciliophora	Spirotrichea	Strombidiida	Strombidiidae_K	Strombidium_K	Strombidium_capitatum
ASV 1119	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 1120	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pterospermaceae	Pterosperma	NA
ASV 1123	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Amphidiniopsidaceae	Islandinium	NA
ASV 1127	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Haptolina	NA
ASV 1128	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	CCW10-lineage	CCW10-lineage_X	CCW10-lineage_X_sp.
ASV 1130	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Labyrinthuloides	Labyrinthuloides_haliotidis
ASV 1132	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Guinardia	Guinardia_striata
ASV 1134	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiophyceae_Clade_B5	Prymnesiophyceae_Clade_B5_X	Prymnesiophyceae_Clade_B5_X_sp.
ASV 1138	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_Clade_F	Prymnesiophyceae_Clade_F_X	Prymnesiophyceae_Clade_F_XX	Prymnesiophyceae_Clade_F_XX_sp.
ASV 1139	Stramenopiles	Ochrophyta	Dictyochophyceae	Dictyochophyceae_X	Dictyochales	Dictyocha	Dictyocha_speculum
ASV 1140	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 1142	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_spirale
ASV 1143	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.

ASV 1151	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_decipiens
ASV 1153	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 1154	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa-lineage_X	Protaspa-lineage_X_sp.
ASV 1157	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X	Dino-Group-I-Clade-1_X_sp.
ASV 1160	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 1166	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 1169	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 1170	Archaeplastida	Streptophyta	Embryophyceae	Embryophyceae_X	Embryophyceae_XX	NA	NA
ASV 1172	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 1175	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	Oomycota_XX	Oomycota_XXX	Oomycota_XXX_sp.
ASV 1177	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Amphidomataceae	Azadinium	NA
ASV 1179	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 1180	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 1184	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 1191	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_anguste-lineata
ASV 1192	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.

ASV 1193	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 1201	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 1203	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-30	Dino-Group-II-Clade-30_X	Dino-Group-II-Clade-30_X_sp.
ASV 1206	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 1207	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-7	Dino-Group-II-Clade-7_X	Dino-Group-II-Clade-7_X_sp.
ASV 1208	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-I	Dino-Group-I-Clade-4	Eududoscquella	Eududoscquella_crenulata
ASV 1209	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 1210	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Raphid-pennate	Navicula	Navicula_sp.
ASV 1211	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 1216	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 1218	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Leptocylindrus	Leptocylindrus_minimus
ASV 1219	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 1228	Stramenopiles	Ochrophyta	Pelagophyceae	Pelagomonadales	Pelagomonadaceae	Pelagomonas	Pelagomonas_calceolata
ASV 1232	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 1234	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	Pyramimonadales_X	Pyramimonadales_XX	Pyramimonadales_XX_sp.
ASV 1237	Archaeplastida	Chlorophyta	Trebouxiophyceae	Watanabea-Clade	Watanabea-Clade_X	Diplosphaera	Diplosphaera_sp.

ASV 1240	Stramenopiles	Sagenista	Labyrinthulomycetes	Labyrinthulales	Labyrinthulaceae	Labyrinthulaceae_X	Labyrinthulaceae_X_sp.
ASV 1243	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_tenera
ASV 1246	Alveolata	Ciliophora	Oligohymenophorea	Apostomatia	Foettingeriidae	Synophrya	Synophrya_sp.
ASV 1261	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-6	Dino-Group-II-Clade-6_X	Dino-Group-II-Clade-6_X_sp.
ASV 1262	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-F	Paraphysomonas	Paraphysomonas_foraminifera
ASV 1266	Alveolata	Ciliophora	Ciliophora_X	Ciliophora_XX	Ciliophora_XXX	Mesodinium	Mesodinium_pulex
ASV 1268	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 1271	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade-I	Chrysophyceae_Clade-I_X	Chrysophyceae_Clade-I_X_sp.
ASV 1274	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Eucampia	Eucampia_sp.
ASV 1276	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 1283	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Radial-centric-basal-Coscinodiscophyceae	Coscinodiscus	Coscinodiscus_granii
ASV 1293	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Labyrinthuloides	Labyrinthuloides_haliotidis
ASV 1294	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 1299	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 1305	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_sp_Clade_Na13C1
ASV 1307	Stramenopiles	Sagenista	Labyrinthulomycetes	Labyrinthulales	Labyrinthulaceae	Labyrinthulaceae_X	Labyrinthulaceae_X_sp.

ASV 1323	Stramenopiles	Stramenopiles_X	Stramenopiles_XX	Stramenopiles_XXX	Stramenopiles_XXXX	Stramenopiles_XXXXX	Stramenopiles_XXXXX_sp.
ASV 1326	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 1329	Archaeplastida	Prasinodermophyta	Prasinodermophyceae	Prasinococcales	Prasinococcales-Clade-B	Prasinococcales-Clade-B_X	Prasinococcales-Clade-B1_X_sp.
ASV 1335	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissuridae_X	Ventrifissuridae_X_sp.
ASV 1336	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Lauderia	Lauderia_annulata
ASV 1342	Rhizaria	Cercozoa	Chlorarachniophyceae	Chlorarachniophyceae_X	NPK2-lineage	NPK2-lineage_X	NPK2-lineage_X_sp.
ASV 1350	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strobilidiidae_G	Strobilidiidae_G_X	Strobilidiidae_G_X_sp.
ASV 1353	Stramenopiles	Opalozoa	MAST-3	MAST-3D	MAST-3D_X	MAST-3D_XX	MAST-3D_XX_sp.
ASV 1355	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	NA
ASV 1361	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	NA	NA
ASV 1365	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-50	Dino-Group-II-Clade-50_X	Dino-Group-II-Clade-50_X_sp.
ASV 1382	Rhizaria	Cercozoa	Filosa-Thecofilosea	Filosa-Thecofilosea_X	Mataza-lineage	Mataza-lineage_X	Mataza-lineage_X_sp.
ASV 1388	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_pellucidum
ASV 1390	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	Bathycoccus_prasinos
ASV 1391	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Lauderia	Lauderia_annulata
ASV 1398	Archaeplastida	Chlorophyta	Pyramimonadophyceae	Pyramimonadales	NA	NA	NA

ASV 1401	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Cryothecomonas-lineage	Cryothecomonas	Cryothecomonas_aestivalis
ASV 1403	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa	NA
ASV 1406	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 1420	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	NA
ASV 1424	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_sp.
ASV 1425	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-12	Dino-Group-II-Clade-12_X	Dino-Group-II-Clade-12_X_sp.
ASV 1434	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	NA
ASV 1435	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	Dino-Group-II_XX	Dino-Group-II_XX_sp.
ASV 1436	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-1	Dino-Group-II-Clade-1_X	Dino-Group-II-Clade-1_X_sp.
ASV 1446	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 1453	Hacrobia	Cryptophyta	Cryptophyceae	Cryptomonadales	Cryptomonadales_X	Plagioselmis	Plagioselmis_prolonga
ASV 1454	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II_X	Dino-Group-II_XX	Dino-Group-II_XX_sp.
ASV 1455	Rhizaria	Cercozoa	Cercozoa_X	Cercozoa_XX	Cercozoa_XXX	Cercozoa_XXXX	Cercozoa_XXXX_sp.
ASV 1462	Stramenopiles	Sagenista	Labyrinthulomycetes	Thraustochytriales	Thraustochytriaceae	Thraustochytriaceae_X	Thraustochytriaceae_X_sp.
ASV 1464	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_decipiens
ASV 1475	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_minima

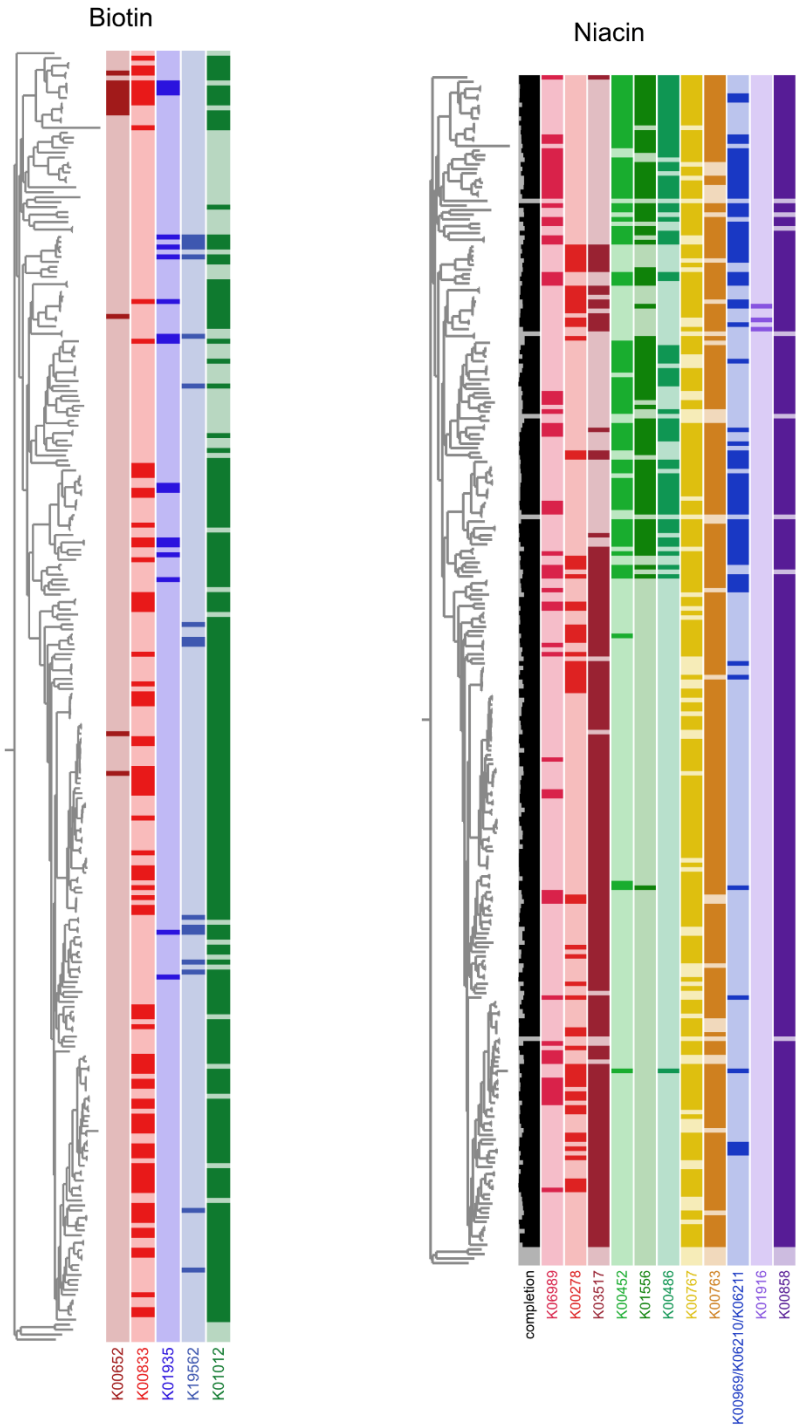
ASV 1490	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	NA	NA	NA
ASV 1496	Alveolata	Apicomplexa	Gregarinomorphea	Gregarines_GRE2	Gregarines_GRE2_X	Gregarines_GRE2_XX	Gregarines_GRE2_XX_sp.
ASV 1498	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Balechina	Balechina_pachyderma ta
ASV 1499	Alveolata	Dinoflagellata	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium_sp.
ASV 1500	Stramenopiles	Opalozoa	MAST-3	MAST-3F	MAST-3F_X	MAST-3F_XX	MAST-3F_XX_sp.
ASV 1505	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_sp_P._quinquecorne_endosymbiont
ASV 1523	Archaeplastida	Prasinodermophyta	Prasinodermophyceae	Prasinococcales	Prasinococcales-Clade-B	Prasinoderma	Prasinoderma_sp.
ASV 1524	Alveolata	Ciliophora	Phyllopharyngea	Cyrtophoria_1	PHYLL_4	PHYLL_4_X	PHYLL_4_X_sp.
ASV 1525	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophyceae_XXX	Dinophyceae_XXX_sp.
ASV 1537	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	NA	NA	NA
ASV 1539	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-4	Dino-Group-II-Clade-4_X	Dino-Group-II-Clade-4_X_sp.
ASV 1580	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	NA	NA	NA
ASV 1636	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-14	Dino-Group-II-Clade-14_X	Dino-Group-II-Clade-14_X_sp.
ASV 1641	Stramenopiles	Pseudofungi	MAST-1	MAST-1C	MAST-1C_X	MAST-1C_XX	MAST-1C_XX_sp.
ASV 1656	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-5	Dino-Group-II-Clade-5_X	Dino-Group-II-Clade-5_X_sp.
ASV 1659	Alveolata	Dinoflagellata	Noctilucophyceae	Noctilucales	Noctilucaceae	Noctilucales_X	Noctilucales_X_sp.

ASV 1674	Hacrobia	Picozoa	Picozoa_X	Picozoa_XX	Picozoa_XXX	Picozoa_XXXX	Picozoa_XXXX_sp.
ASV 1682	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium	Gyrodinium_heterogrammum
ASV 1694	Archaeplastida	Chlorophyta	Trebouxiophyceae	Chlorellales	Chlorellales_X	Nannochloris	Nannochloris_sp.
ASV 1704	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystidae	Raineriophrys	NA
ASV 1727	Archaeplastida	Chlorophyta	Mamiellophyceae	Dolichomastigales	Dolichomastigaceae	NA	NA
ASV 1730	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Thalassiosira	Thalassiosira_delicatula
ASV 1746	Stramenopiles	Opalozoa	MAST-3	MAST-3J	MAST-3J_X	MAST-3J_XX	MAST-3J_XX_sp.
ASV 1762	Alveolata	Dinoflagellata	Dinophyceae	Gymnodiniales	Kareniaceae	Karenia	NA
ASV 1766	Stramenopiles	Stramenopiles_X	Stramenopiles_XX	Stramenopiles_XXX	Stramenopiles_XXXX	Stramenopiles_XXXXX	Stramenopiles_XXXXX_sp.
ASV 1823	Rhizaria	Cercozoa	Filosa-Thecofilosea	Ventricleftida	Ventrifissuridae	Ventrifissura	Ventrifissura_artocarpoidea
ASV 1829	Alveolata	Ciliophora	Spirotrichea	Tintinnida	NA	NA	NA
ASV 1832	Rhizaria	Cercozoa	Filosa-Thecofilosea	Cryomonadida	Protaspa-lineage	Protaspa	NA
ASV 1860	Alveolata	Ciliophora	Ciliophora_X	Ciliophora_XX	Ciliophora_XXX	Mesodinium	Mesodinium_pulex
ASV 1897	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	NA	NA
ASV 1903	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina	Chrysochromulina_sp.
ASV 1921	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	Bathycoccus_oceanicus

ASV 1930	Apusozoa	Apusomonadidae	Apusomonadidae_Group-1	Apusomonadidae_Group-1_X	Apusomonadidae_Group-1_XX	Amastigomonas	Amastigomonas_sp.
ASV 1947	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyta_X	Polar-centric-Mediophyceae	Chaetoceros	Chaetoceros_socialis_debilis
ASV 1965	Hacrobia	Telonemia	Telonemia_X	Telonemia_XX	Telonemia-Group-1	Telonemia-Group-1_X	Telonemia-Group-1_X_sp.
ASV 1982	Stramenopiles	Pseudofungi	Oomycota	Oomycota_X	Oomycota_XX	Oomycota_XXX	Oomycota_XXX_sp.
ASV 1986	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Protodinium	Protodinium_simplex
ASV 1988	Rhizaria	Cercozoa	Filosa-Imbricatea	Marimonadida	Marimonadida_X	Marimonadida_XX	Marimonadida_XX_sp.
ASV 2006	Alveolata	Ciliophora	Spirotrichea	Choreotrichida	Strombidinopsidae	Parastrombidinopsis	NA
ASV 2024	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	NA	NA	NA
ASV 2026	Alveolata	Dinoflagellata	Dinophyceae	Suessiales	Suessiaceae	Biecheleria	NA
ASV 2046	Alveolata	Dinoflagellata	Syndiniales	Dino-Group-II	Dino-Group-II-Clade-3	Dino-Group-II-Clade-3_X	Dino-Group-II-Clade-3_X_sp.
ASV 2078	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Prymnesiaceae	Prymnesium	NA

Appendix 2 – Data associated with Chapter 3 - Comparative genomics of B vitamin biosynthesis pathways in the Stramenopila

2.1 Presence and absence of biosynthesis KOs for Biotin and Niacin with a genome completeness cut off of 70%



Appendix 3 – Data associated with Chapter 4 - Physiology, Imaging and Metagenomics of the MAST3 species *Incisomonas marina* and associated bacterial consortia

3.1 MaSuRCA config file for draft assembly of *I. marina*

```
# DATA is specified as type {PE,JUMP,OTHER,PACBIO} and 5 fields:

# 1)two_letter_prefix 2)mean 3)stdev 4)fastq(.gz)_fwd_reads

# 5)fastq(.gz)_rev_reads. The PE reads are always assumed to be

# innies, i.e. --->.<---, and JUMP are assumed to be outties

# <---.--->. If there are any jump libraries that are innies, such as

# longjump, specify them as JUMP and specify NEGATIVE mean. Reverse reads

# are optional for PE libraries and mandatory for JUMP libraries. Any

# OTHER sequence data (454, Sanger, Ion torrent, etc) must be first

# converted into Celera Assembler compatible .frg files (see

# http://wgs-assembler.sourceforge.com)

DATA

#Illumina paired end reads supplied as <two-character prefix> <fragment mean> <fragment stdev> <forward_reads> <reverse_reads>

#if single-end, do not specify <reverse_reads>

#If mean/stdev are unknown use 500 and 50 -- these are safe values that will work for most runs

#MUST HAVE Illumina paired end reads to use MaSuRCA

PE= pe 250 30  /rds/user/dea33/hpc-work/Inci-Genomics/data/short_reads/SRR2962707.1.fastq

#Illumina mate pair reads supplied as <two-character prefix> <fragment mean> <fragment stdev> <forward_reads> <reverse_reads>

#JUMP= sh 3600 200  /FULL_PATH/short_1.fastq  /FULL_PATH/short_2.fastq

#pacbio OR nanopore reads must be in a single fasta or fastq file with absolute path, can be gzipped

#if you have both types of reads supply them both as NANOPORE type

#PACBIO=/FULL_PATH/pacbio.fa

NANOPORE=/rds/user/dea33/hpc-work/Inci-Genomics/data/long_reads/all_inciONT-clean.fastq

#Legacy reads (Sanger, 454, etc) in one frg file, concatenate your frg files into one if you have many

#OTHER=/FULL_PATH/file.frg
```

#synteny-assisted assembly, concatenate all reference genomes into one reference.fa; works for Illumina-only data

#REFERENCE=/FULL_PATH/nanopore.fa

END

PARAMETERS

#PLEASE READ all comments to essential parameters below, and set the parameters according to your project

#set this to 1 if your Illumina mate pair (jumping) library reads are shorter than 100bp

EXTEND_JUMP_READS=0

#this is k-mer size for deBruijn graph values between 25 and 127 are supported, auto will compute the optimal size based on the read data and GC content

GRAPH_KMER_SIZE = auto

#set this to 1 for all Illumina-only assemblies

#set this to 0 if you have more than 15x coverage by long reads (Pacbio or Nanopore) or any other long reads/mate pairs (Illumina MP, Sanger, 454, etc)

USE_LINKING_MATES = 0

#specifies whether to run the assembly on the grid

USE_GRID=0

#specifies grid engine to use SGE or SLURM

GRID_ENGINE=SLURM

#specifies queue (for SGE) or partition (for SLURM) to use when running on the grid MANDATORY

GRID_QUEUE=all.q

#batch size in the amount of long read sequence for each batch on the grid

GRID_BATCH_SIZE=500000000

#use at most this much coverage by the longest Pacbio or Nanopore reads, discard the rest of the reads

#can increase this to 30 or 35 if your long reads reads have N50<7000bp

LHE_COVERAGE=25

#this parameter is useful if you have too many Illumina jumping library reads. Typically set it to 60 for bacteria and 300 for the other organisms

LIMIT_JUMP_COVERAGE = 300

#these are the additional parameters to Celera Assembler; do not worry about performance, number or processors or batch sizes -- these are computed automatically.

#CABOG ASSEMBLY ONLY: set cgwErrorRate=0.25 for bacteria and 0.1<=cgwErrorRate<=0.15 for other organisms.

CA_PARAMETERS = cgwErrorRate=0.15

#CABOG ASSEMBLY ONLY: whether to attempt to close gaps in scaffolds with Illumina or long read data

CLOSE_GAPS=1

#number of cpus to use, set this to the number of CPUs/threads per node you will be using

NUM_THREADS = 32

#this is mandatory jellyfish hash size -- a safe value is estimated_genome_size*20

JF_SIZE = 1200000000

#ILLUMINA ONLY. Set this to 1 to use SOAPdenovo contigging/scaffolding module.

#Assembly will be worse but will run faster. Useful for very large (>=8Gbp) genomes from Illumina-only data

SOAP_ASSEMBLY=0

#If you are doing Hybrid Illumina paired end + Nanopore/PacBio assembly ONLY (no Illumina mate pairs or OTHER frg files).

#Set this to 1 to use Flye assembler for final assembly of corrected mega-reads.

#A lot faster than CABOG, AND QUALITY IS THE SAME OR BETTER.

#DO NOT use if you have less than 20x coverage by long reads.

FLYE_ASSEMBLY=1

END

3.2 Colony PCR sequences of 16S by Sanger sequencing (Section 4.3.2 and Figure 4.6)

Seq 1 (white) – SILVA prediction: Rhodobacteracea

TAAANATGCAGTCGAGCGCTCTCTTCGGAGGGAGCGGCGGACGGGTTAGTAACGCGTGGGAACGTGCCCTTCTCTAAGGAATAGCCA
CTGGAAACGGTGAGTAATACCTTATACGCCCTTCGGGGGAAAGATTTATCGGAGAAGGATCGGCCCGCGTTAGATTAGATAGTTGGTG
GGGTAACGGCCTACCAAGTCTACGATCTATAGCTGGTTTTAGAGGATGATCAGCAACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATCTTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTCGTAAA
GCTCTTTCGCCTGTGAAGATAATGACTGTAGCAGGTAAAGAAACCCCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGT
TAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCGCGTAGGCGGACTGGAAAGTTGGGGGTGAAATCCCAGGGCTCAACCCTGGAACT
GCCTCCAAAACTATCAGTCTAGAGTTCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACC
AGTGGCGAAGGCGGCTCACTGGCTCGATACTGACGCTGAGGTGCGAAAGTGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCA
CACCGTAAACGATGAATGCCAGTCGTCGGGTAGCATGCTATTCGGTGACACACCTAACGGATTAAGCATTCCGCCTGGGGAGTACGGT
CGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTAC

CAACCCTTGACATCCTAGGACCGCCAGAGAGATTTGGCTTTCACTTCGGTGACCTAGTGACAGGTGCTGCATGGCTGTCGTCAGCTCG
TGTCGTGAGATGTTCGGTTAAGTCCGGCAACGAGCGCAACCCACATCCTTAGTTGCCAGCAGTTCGGCTGGGCACTCTAGGGAAACTG
CCCGTGANAAGCGGAAGGAAGGGTGGGGGATGACGTCAAGTCCTCAGGGCCCTTACGGGGTGGGGCTACNACCGTGCTACAATGGCC
AGGGACAATGGGGTTAATCCCCAAAAAACTGGNNNAAGTTCGGAATTGGGGGCNGGCAACNCGACCCCCAGGAAAGCCGGGAANCC

Seq 2 (white) – SILVA prediction: Alteromonas

TGCAGTCGANGGAACATGTCTAGCTTGCTAGATGATGTCGAGTGGCGGACGGGTGAGTAATGCTTGGGAACTTGCCTTTGCGAGGGG
GATAACAGTTGGAAACGACTGCTAATACCGCATAATGTCTACGGACCAAACAGGGCTTCGGCTCTGGCGCAAAGAGAGGCCCAAGTGA
GATTAGCTAGTTGGTGAGGTAAAGGCTCACCAAGGCAACGATCTCTAGCTGTTCTGAGAGGAAGATCAGCCACACTGGGACTGAGACA
CGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGTGTGAAGAAG
GCCTTCGGGTTGTAAAGCACTTTCAGTTGTGAGGAAAAGTTAGTAGTTAATACCTGCTAGCCGTGACGTTAACAACAGAAGAAGCACCG
GCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTG
TTAAGCTAGATGTGAAAGCCCCGGGCTCAACCTGGGACGGTCATTTAGAACTGGCAGACTAGAGTCTTGGAGAGGGGAGTGGAATTCC
AGGTGTAGCGGTGAAATGCGTAGATATCTGGAGGAACATCAGTGGCGAAGGCGACTCCCTGGCCAAAGACTGACGCTCATGTGCGAAA
GTGTGGGTAGCGAACAGGATTAGATACCCTGGTAGTCCACACCGTAAACGCTGTCTACTAGCTGTTTGTGGCTTTAAGCCGTGAGTAG
CGAAGCTAACGCGATAAGGAAGACCGCCTG

Seq 3 (yellow) SILVA prediction: Winogradskyella

CAGTCGAACGGTAACATTGATGCTTGCATCAGATGACGAGTGGCGCACGGGTGAGTAACGCGTATACAATCTGCCTTTTACAGGGGGA
TAGCCTTTAGAAATGAAGATTAATATCCCATAGTAAGTTGTTATTGCATGATAGTAACTTTAAAGTTTCGGCGGTAAAAGATGAGTATGC
GTTCTATTAGCTAGATGGTGTGGTAACGGCACACCATGGCGACGATAGATAGGGGCCCTGAGAGGGGGATCCCCCACACTGGTACTGA
GACACGGACCAGACTCCTACGGGAGGCAGCAGTGAGGAATATTGGACAATGGGCGAGAGCCTGATCCAGCCATGCCGCGTGCAGGAA
GACTGCCCTATGGGTTGTAAACTGCTTTTATACAGGAAGAAACCTTTTCACGTGTGAAAAGCTGACGGTACTGTAAGAATAAGGATCGG
CTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGATCCAAGCGTTATCCGGAATCATTGGGTTTAAAGGGTCCGTAGGTTGATAATT
AAGTCAGGGGTGAAAGTTTGCAGCTCAACTGTAAAATTGCCTTTGATACTGGTTATCTTGAATTATTGTGAAGTGGTTAGAATATGTAG
TGTAGCGGTGAAATGCATAGATATTACATAGAATACCAATTGCGAAGGCAGATCACTAACAATATATTGACACTGATGGACGAAAGCGT
GGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGCTGTTCGGTTTTCGGACTGAGTGGCTAAGCG
AAAGTGATAAGTATCCCCACCTGGGGGAGTACGTTTCGCAAGAATGAAAACTCAAAGGAAATTGACGGGGGGCCCGCACAAGCGGGGG
GAGCATGGGGGTTTAATTCCGATGGATACGCGGAGGAACCCTTACCNAGGGCTTAAAATGGAAAGTTGAAATAAAGTTAA

Seq 4 (clear) SILVA prediction: Marinobacter

TGCAGTCGAGCGGTANAGGGGTAGCTTGCTACCCGCTGACGAGCGGCGGACGGGTGAGTAATGCTTAGGAATCTGCCCAGTAGTGGG
GGATAGCCCGGGGAAACCCGGATTAATACCGCATACGTCCTACGGGAGAAAGCAGGGGATCTTCGGACCTTGCGCTATTGGATGAGCC
TAAGTCGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACATCGGGAC
TGAGACACGGCCCGAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGTGT
GAAGAAGGCTTTCGGGTTGTAAAGCACTTTCAGTAGGGAGGAAAACCTTATGGCTAATACCCATGAGGCTTGACGTTACCTACAGAAGA
AGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGG
TGGTTTGGTAAGCGAGATGTGAAAGCCCCGGGCTTAACCTGGGAACGGCATTTCGAACTGTCAGACTAGAGTGTGGTAGAGGGTAGTG
GAATTTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGGCTACCTGGACCAACACTGACACTGAGG
NGCGAAAGCGNGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCAACTAGCCGTTGGGACTCTTGAAGTC
TTAGTGGCGCAGCTAACGCACTAAGTTGACCGCCTGGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGGCCCGC
ACAAGCGGNGGGAGCATGG

3.3 – RepeatMasker output

```
sequences:           212
total length:    68108360 bp  (68108160 bp excl N/X-runs)
GC level:          52.49 %
bases masked:    13964093 bp ( 20.50 %)
==================================================
               number of      length   percentage
               elements*    occupied  of sequence
--------------------------------------------------
Retroelements         3110      2698197 bp    3.96 %
   SINEs:                0            0 bp    0.00 %
   Penelope:             0            0 bp    0.00 %
   LINEs:             2190      1836443 bp    2.70 %
    CRE/SLACS         1105      1110978 bp    1.63 %
     L2/CR1/Rex          0            0 bp    0.00 %
     R1/LOA/Jockey       0            0 bp    0.00 %
     R2/R4/NeSL          0            0 bp    0.00 %
     RTE/Bov-B           0            0 bp    0.00 %
     L1/CIN4             0            0 bp    0.00 %
   LTR elements:       920       861754 bp    1.27 %
     BEL/Pao            40         8687 bp    0.01 %
     Ty1/Copia          19         4200 bp    0.01 %
     Gypsy/DIRS1       575       722253 bp    1.06 %
       Retroviral        0            0 bp    0.00 %

DNA transposons       2061       839607 bp    1.23 %
   hobo-Activator      364        88395 bp    0.13 %
   Tc1-IS630-Pogo       33        27821 bp    0.04 %
   En-Spm                0            0 bp    0.00 %
   MULE-MuDR             0            0 bp    0.00 %
   PiggyBac            833       235490 bp    0.35 %
   Tourist/Harbinger     0            0 bp    0.00 %
   Other (Mirage,        0            0 bp    0.00 %
    P-element, Transib)

Rolling-circles          0            0 bp    0.00 %

Unclassified:        12510      3830983 bp    5.62 %

Total interspersed repeats:     7368787 bp   10.82 %


Small RNA:              46       162019 bp    0.24 %

Satellites:           1266       294631 bp    0.43 %
Simple repeats:      65143      5834774 bp    8.57 %
Low complexity:       4582       303882 bp    0.45 %
==================================================

* most repeats fragmented by insertions or deletions
  have been counted as one element


RepeatMasker version 4.1.5 , default mode

run with rmblastn version 2.14.0+
```

3.4 – BLAST results from query of I. marina agains NCBI virus RefSeq database

Hits in the Incisomonas marina genome from NCBI RefSeq Virus sequences

Query - Viral seq NCBI ID	I. marina *contig ID*	% Pairwise Identity	Bit-Score	E Value	Grade	Additional query ID
NC_019410.1	***contig_237***	70.40%	117.459	8.79E-25	64.60%	_cds_YP_006989560.1_180
NC_049451.1	***contig_237***	75.10%	196.865	8.98E-49	59.60%	_cds_YP_009882599.1_42
NC_053512.1	***contig_254***	74.50%	169.165	1.90E-40	58.00%	_cds_YP_010013577.1_87
NC_041868.1	***contig_305***	81.70%	276.271	4.68E-73	79.10%	_cds_YP_009593515.1_38
NC_049342.1	***contig_347***	71.60%	226.412	1.09E-57	80.00%	_cds_YP_009949256.1_57
NC_055026.1	***contig_347***	73.50%	361.217	3.65E-98	78.80%	_cds_YP_010082065.1_26
NC_003324.1	***contig_412***	81.80%	195.019	1.23E-48	69.60%	_cds_YP_010115352.1_35
NC_052968.1	***contig_416***	74.90%	169.165	3.62E-40	47.80%	_cds_YP_009997126.1_44
NC_047790.1	***contig_453***	81.60%	1914.25	0	78.00%	_cds_YP_009787988.1_19
NC_047790.1	***contig_453***	89.10%	1212.52	0	91.90%	_cds_YP_009787993.1_24
NC_047790.1	***contig_453***	89.90%	2285.43	0	94.80%	_cds_YP_009787994.1_25
NC_047790.1	***contig_453***	87.20%	948.452	0	93.60%	_cds_YP_009787995.1_26
NC_047790.1	***contig_453***	81.10%	797.026	0	90.20%	_cds_YP_009787996.1_27
NC_047790.1	***contig_453***	83.40%	1013.08	0	91.40%	_cds_YP_009788002.1_33
NC_047790.1	***contig_453***	78.00%	1092.49	0	89.00%	_cds_YP_009788009.1_40
NC_047790.1	***contig_453***	77.50%	682.534	0	87.50%	_cds_YP_009788011.1_42
NC_047790.1	***contig_453***	86.50%	580.968	1.09E-164	93.20%	_cds_YP_009787998.1_29
NC_047790.1	***contig_453***	85.20%	531.109	1.61E-149	76.60%	_cds_YP_009788015.1_46
NC_047790.1	***contig_453***	85.60%	470.169	2.06E-131	92.70%	_cds_YP_009788003.1_34
NC_047790.1	***contig_453***	79.50%	460.936	1.84E-128	89.70%	_cds_YP_009788001.1_32
NC_047790.1	***contig_453***	79.00%	412.923	4.89E-114	89.50%	_cds_YP_009787979.1_10
NC_047790.1	***contig_453***	81.80%	411.077	1.39E-113	90.90%	_cds_YP_009788014.1_45
NC_047790.1	***contig_453***	80.00%	383.377	3.27E-105	90.00%	_cds_YP_009787997.1_28
NC_047790.1	***contig_453***	90.10%	327.977	7.76E-89	90.70%	_cds_YP_009787989.1_20
NC_047790.1	***contig_453***	88.90%	233.798	1.22E-60	91.50%	_cds_YP_009788004.1_35
NC_047790.1	***contig_453***	81.50%	204.252	1.40E-51	84.70%	_cds_YP_009787987.1_18
NC_047790.1	***contig_453***	86.10%	174.705	9.42E-43	76.00%	_cds_YP_009788007.1_38
NC_048066.1	***contig_453***	72.90%	340.904	6.93E-92	67.10%	_cds_YP_009812525.1_20
NC_055862.1	***contig_453***	73.10%	326.131	2.08E-87	62.80%	_cds_YP_010109753.1_25
NC_028829.1	***contig_463***	73.60%	147.006	2.04E-33	45.70%	_cds_YP_009201475.1_372
NC_049343.1	***contig_482***	72.00%	213.485	1.08E-53	67.70%	_cds_YP_009949297.1_41
NC_047756.1	***contig_506***	79.20%	171.012	2.60E-41	65.00%	_cds_YP_009785463.1_75
NC_049451.1	***contig_537***	79.80%	494.176	5.23E-138	58.60%	_cds_YP_009882596.1_39
NC_049453.1	***contig_537***	79.80%	494.176	5.23E-138	58.60%	_cds_YP_009882647.1_4
NC_000902.1	***contig_596***	84.30%	350.137	2.92E-95	80.80%	_cds_NP_050546.1_50
NC_000924.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_NP_049509.1_49
NC_000924.1	***contig_596***	87.60%	457.243	1.66E-127	86.40%	_cds_NP_049507.1_47
NC_000924.1	***contig_596***	84.90%	185.785	3.17E-46	92.50%	_cds_NP_049508.1_48
NC_001416.1	***contig_596***	100.00%	881.972	0	100.00%	_cds_NP_040645.1_68
NC_001416.1	***contig_596***	100.00%	854.272	0	100.00%	_cds_NP_040646.1_69
NC_001416.1	***contig_596***	100.00%	987.231	0	100.00%	_cds_NP_597781.1_72
NC_001416.1	***contig_596***	99.70%	593.895	8.26E-169	99.80%	_cds_NP_040644.1_66
NC_001416.1	***contig_596***	99.70%	582.815	1.75E-165	99.80%	_cds_YP_001551775.1_67
NC_001416.1	***contig_596***	100.00%	544.035	7.59E-154	100.00%	_cds_NP_597780.1_71
NC_001416.1	***contig_596***	100.00%	383.377	1.18E-105	100.00%	_cds_NP_597782.1_73
NC_001416.1	***contig_596***	100.00%	339.057	2.26E-92	100.00%	_cds_YP_001551744.1_70
NC_002166.1	***contig_596***	96.00%	776.713	0	98.00%	_cds_NP_037696.1_35

NC_002166.1	***contig_596***	98.50%	571.735	3.87E-162	99.20%	_cds_NP_037695.1_34
NC_002167.1	***contig_596***	96.40%	787.793	0	98.20%	_cds_NP_037753.1_57
NC_002167.1	***contig_596***	97.80%	556.962	1.07E-157	98.90%	_cds_NP_037752.1_56
NC_002167.1	***contig_596***	95.70%	189.479	6.27E-47	61.20%	_cds_NP_037754.1_58
NC_002371.2	***contig_596***	90.90%	412.923	2.51E-114	92.40%	_cds_NP_059621.1_64
NC_002371.2	***contig_596***	90.80%	407.383	1.13E-112	93.20%	_cds_YP_063731.1_65
NC_002730.1	***contig_596***	95.40%	760.093	0	97.70%	_cds_NP_112069.1_36
NC_002730.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_NP_112068.1_35
NC_002730.1	***contig_596***	92.20%	348.29	1.05E-94	72.30%	_cds_NP_112070.1_37
NC_003444.1	***contig_596***	83.80%	375.99	4.91E-103	83.30%	_cds_NP_599082.1_50
NC_004813.1	***contig_596***	84.50%	355.677	6.27E-97	80.90%	_cds_YP_001449288.1_51
NC_004813.1	***contig_596***	85.80%	150.699	1.46E-35	73.80%	_cds_YP_001449289.1_52
NC_004913.3	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_NP_859242.1_164
NC_004913.3	***contig_596***	87.60%	457.243	1.66E-127	86.40%	_cds_NP_859238.1_160
NC_004913.3	***contig_596***	89.60%	351.984	4.66E-96	94.80%	_cds_NP_859240.1_162
NC_004913.3	***contig_596***	84.20%	263.345	2.19E-69	92.10%	_cds_NP_859239.1_161
NC_004913.3	***contig_596***	84.90%	185.785	3.17E-46	92.50%	_cds_NP_859241.1_163
NC_004913.3	***contig_596***	95.70%	185.785	4.35E-46	71.40%	_cds_NP_859243.1_165
NC_004914.3	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_NP_859413.1_167
NC_004914.3	***contig_596***	87.60%	457.243	1.66E-127	86.40%	_cds_NP_859409.1_163
NC_004914.3	***contig_596***	89.60%	351.984	4.66E-96	94.80%	_cds_NP_859411.1_165
NC_004914.3	***contig_596***	84.20%	263.345	2.19E-69	92.10%	_cds_NP_859410.1_164
NC_004914.3	***contig_596***	84.90%	185.785	3.17E-46	92.50%	_cds_NP_859412.1_166
NC_004914.3	***contig_596***	95.70%	185.785	4.35E-46	71.40%	_cds_NP_859414.1_168
NC_005344.1	***contig_596***	95.00%	749.013	0	97.50%	_cds_NP_958236.1_60
NC_005344.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_NP_958234.1_58
NC_005344.1	***contig_596***	98.70%	566.195	1.76E-160	99.40%	_cds_NP_958235.1_59
NC_005344.1	***contig_596***	92.30%	167.319	2.94E-40	59.50%	_cds_NP_958237.1_61
NC_005841.1	***contig_596***	85.40%	254.111	2.35E-66	68.90%	_cds_YP_006398.1_42
NC_006949.1	***contig_596***	90.90%	412.923	2.51E-114	92.40%	_cds_YP_224212.1_74
NC_006949.1	***contig_596***	90.80%	407.383	1.13E-112	93.20%	_cds_YP_224213.1_75
NC_006949.1	***contig_596***	82.30%	340.904	1.77E-92	83.60%	_cds_YP_224215.1_77
NC_008464.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_794058.1_11
NC_008464.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_794056.1_9
NC_008464.1	***contig_596***	82.30%	158.086	6.91E-38	91.10%	_cds_YP_794057.1_10
NC_009514.1	***contig_596***	85.10%	433.236	2.82E-120	88.60%	_cds_YP_001272572.1_59
NC_010237.1	***contig_596***	98.60%	610.515	8.75E-174	99.30%	_cds_YP_001648941.1_50
NC_010237.1	***contig_596***	87.40%	451.703	7.72E-126	86.30%	_cds_YP_001648940.1_49
NC_011356.1	***contig_596***	84.00%	344.597	1.36E-93	80.70%	_cds_YP_002274183.1_47
NC_011356.1	***contig_596***	85.80%	150.699	1.46E-35	73.80%	_cds_YP_002274184.1_48
NC_011357.1	***contig_596***	90.60%	156.239	6.36E-37	58.70%	_cds_YP_002274258.1_47
NC_011357.1	***contig_596***	85.10%	143.312	1.61E-33	86.60%	_cds_YP_002274261.1_50
NC_011802.1	***contig_596***	85.40%	254.111	2.35E-66	68.90%	_cds_YP_002455882.1_46
NC_011976.1	***contig_596***	95.00%	749.013	0	97.50%	_cds_YP_002533525.1_66
NC_011976.1	***contig_596***	97.80%	555.115	3.94E-157	98.00%	_cds_YP_002533524.1_65
NC_011976.1	***contig_596***	94.00%	178.399	1.36E-43	60.40%	_cds_YP_002533526.1_67
NC_013059.1	***contig_596***	95.20%	754.553	0	97.60%	_cds_YP_003090277.1_60
NC_013059.1	***contig_596***	97.80%	555.115	3.94E-157	98.00%	_cds_YP_003090276.1_59
NC_013059.1	***contig_596***	94.00%	178.399	1.36E-43	60.40%	_cds_YP_003090278.1_61
NC_016160.1	***contig_596***	96.60%	793.333	0	98.30%	_cds_YP_004934160.1_53
NC_016160.1	***contig_596***	98.50%	571.735	3.87E-162	99.20%	_cds_YP_004934159.1_52
NC_016160.1	***contig_596***	94.90%	183.939	2.92E-45	60.80%	_cds_YP_004934161.1_54
NC_017985.1	***contig_596***	90.20%	320.591	2.29E-86	71.30%	_cds_YP_006383883.1_44

NC_018275.1	***contig_596***	93.30%	156.239	6.36E-37	58.70%	_cds_YP_006560617.1_67
NC_018279.1	***contig_596***	90.20%	320.591	2.29E-86	71.30%	_cds_YP_006560883.1_77
NC_019442.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_007001465.1_45
NC_019442.1	***contig_596***	84.90%	403.69	2.19E-111	85.70%	_cds_YP_007001464.1_44
NC_019501.1	***contig_596***	95.40%	760.093	0	97.70%	_cds_YP_007004318.1_14
NC_019501.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_YP_007004317.1_13
NC_019501.1	***contig_596***	97.40%	200.558	2.90E-50	62.10%	_cds_YP_007004319.1_15
NC_019705.1	***contig_596***	96.90%	798.873	0	98.40%	_cds_YP_007111496.1_60
NC_019705.1	***contig_596***	98.50%	571.735	3.87E-162	99.20%	_cds_YP_007111495.1_59
NC_019705.1	***contig_596***	99.10%	211.638	1.34E-53	62.90%	_cds_YP_007111497.1_61
NC_019706.1	***contig_596***	95.60%	765.633	0	97.80%	_cds_YP_007111566.1_63
NC_019706.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_YP_007111565.1_62
NC_019706.1	***contig_596***	94.70%	176.552	4.88E-43	60.20%	_cds_YP_007111567.1_64
NC_019708.1	***contig_596***	95.80%	771.173	0	97.90%	_cds_YP_007111633.1_57
NC_019708.1	***contig_596***	98.10%	566.195	1.80E-160	99.10%	_cds_YP_007111632.1_56
NC_019708.1	***contig_596***	84.50%	267.038	2.96E-70	71.70%	_cds_YP_007111634.1_58
NC_019709.1	***contig_596***	95.40%	760.093	0	97.70%	_cds_YP_007111696.1_59
NC_019709.1	***contig_596***	97.50%	555.115	3.90E-157	98.80%	_cds_YP_007111695.1_58
NC_019709.1	***contig_596***	96.50%	187.632	2.26E-46	61.10%	_cds_YP_007111697.1_60
NC_019710.1	***contig_596***	95.80%	771.173	0	97.90%	_cds_YP_007111769.1_66
NC_019710.1	***contig_596***	98.50%	571.735	3.87E-162	99.20%	_cds_YP_007111768.1_65
NC_019710.1	***contig_596***	97.40%	200.558	2.90E-50	62.10%	_cds_YP_007111770.1_67
NC_019710.1	***contig_596***	98.90%	171.012	7.28E-42	79.90%	_cds_YP_007111767.1_64
NC_019711.1	***contig_596***	96.00%	776.713	0	98.00%	_cds_YP_007111838.1_64
NC_019711.1	***contig_596***	100.00%	854.272	0	100.00%	_cds_YP_007111839.1_65
NC_019711.1	***contig_596***	100.00%	987.231	0	100.00%	_cds_YP_007111842.1_68
NC_019711.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_YP_007111837.1_63
NC_019711.1	***contig_596***	100.00%	544.035	7.59E-154	100.00%	_cds_YP_007111841.1_67
NC_019711.1	***contig_596***	100.00%	383.377	1.18E-105	100.00%	_cds_YP_007111843.1_69
NC_019711.1	***contig_596***	100.00%	339.057	2.26E-92	100.00%	_cds_YP_007111840.1_66
NC_019714.1	***contig_596***	95.80%	771.173	0	97.90%	_cds_YP_007112009.1_56
NC_019714.1	***contig_596***	98.50%	571.735	3.87E-162	99.20%	_cds_YP_007112008.1_55
NC_019714.1	***contig_596***	97.40%	200.558	2.90E-50	62.10%	_cds_YP_007112010.1_57
NC_019715.1	***contig_596***	93.50%	710.234	0	96.80%	_cds_YP_007112069.1_56
NC_019715.1	***contig_596***	98.10%	566.195	1.80E-160	99.10%	_cds_YP_007112068.1_55
NC_019715.1	***contig_596***	96.50%	187.632	2.26E-46	61.10%	_cds_YP_007112070.1_57
NC_019715.1	***contig_596***	98.90%	171.012	7.28E-42	79.90%	_cds_YP_007112067.1_54
NC_019716.1	***contig_596***	86.10%	263.345	3.91E-69	69.10%	_cds_YP_007112131.1_57
NC_019719.1	***contig_596***	92.00%	636.368	0	93.40%	_cds_YP_007112338.1_62
NC_019719.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_YP_007112337.1_61
NC_019720.1	***contig_596***	95.60%	765.633	0	97.80%	_cds_YP_007112412.1_69
NC_019720.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_YP_007112411.1_68
NC_019720.1	***contig_596***	94.70%	176.552	4.88E-43	60.20%	_cds_YP_007112413.1_70
NC_019723.1	***contig_596***	99.80%	876.432	0	99.90%	_cds_YP_007112602.1_64
NC_019723.1	***contig_596***	100.00%	854.272	0	100.00%	_cds_YP_007112603.1_65
NC_019723.1	***contig_596***	100.00%	987.231	0	100.00%	_cds_YP_007112606.1_68
NC_019723.1	***contig_596***	99.70%	593.895	8.26E-169	99.80%	_cds_YP_007112600.1_62
NC_019723.1	***contig_596***	99.70%	582.815	1.75E-165	99.80%	_cds_YP_007112601.1_63
NC_019723.1	***contig_596***	100.00%	544.035	7.59E-154	100.00%	_cds_YP_007112605.1_67
NC_019723.1	***contig_596***	100.00%	383.377	1.18E-105	100.00%	_cds_YP_007112607.1_69
NC_019723.1	***contig_596***	100.00%	339.057	2.26E-92	100.00%	_cds_YP_007112604.1_66
NC_019767.1	***contig_596***	95.60%	765.633	0	97.80%	_cds_YP_007151667.1_60
NC_019767.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_YP_007151666.1_59

NC_019767.1	***contig_596***	94.00%	178.399	1.36E-43	60.40%	_cds_YP_007151668.1_61
NC_019768.1	***contig_596***	90.00%	586.508	2.08E-166	92.40%	_cds_YP_007151732.1_62
NC_019768.1	***contig_596***	98.10%	566.195	1.80E-160	99.10%	_cds_YP_007151731.1_61
NC_019769.1	***contig_596***	96.20%	782.253	0	98.10%	_cds_YP_007151787.1_52
NC_019769.1	***contig_596***	98.10%	566.195	1.80E-160	99.10%	_cds_YP_007151786.1_51
NC_019769.1	***contig_596***	94.90%	183.939	2.92E-45	60.80%	_cds_YP_007151788.1_53
NC_021857.1	***contig_596***	88.10%	470.169	2.19E-131	85.50%	_cds_YP_008318530.1_52
NC_022749.1	***contig_596***	87.60%	459.089	4.74E-128	85.20%	_cds_YP_008766915.1_51
NC_027339.1	***contig_596***	79.70%	263.345	3.70E-69	82.10%	_cds_YP_009147510.1_63
NC_027398.1	***contig_596***	95.60%	765.633	0	97.80%	_cds_YP_009153133.1_59
NC_027398.1	***contig_596***	98.80%	577.275	8.32E-164	99.40%	_cds_YP_009153131.1_57
NC_027398.1	***contig_596***	98.70%	566.195	1.76E-160	99.40%	_cds_YP_009153132.1_58
NC_027398.1	***contig_596***	91.50%	161.779	1.37E-38	59.10%	_cds_YP_009153134.1_60
NC_028449.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009180940.1_58
NC_028449.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009180939.1_57
NC_028449.1	***contig_596***	93.20%	108.226	7.85E-23	64.60%	_cds_YP_009180941.1_59
NC_028656.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009187370.1_56
NC_028656.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009187369.1_55
NC_028656.1	***contig_596***	95.70%	185.785	4.35E-46	71.40%	_cds_YP_009187371.1_57
NC_028696.2	***contig_596***	89.30%	316.897	3.20E-85	69.80%	_cds_YP_009191506.2_54
NC_029120.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009226876.1_44
NC_029120.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009226875.1_43
NC_029120.1	***contig_596***	95.00%	187.632	1.23E-46	71.40%	_cds_YP_009226877.1_45
NC_030919.1	***contig_596***	90.20%	320.591	2.29E-86	71.30%	_cds_YP_009274683.1_40
NC_031019.1	***contig_596***	90.90%	412.923	2.51E-114	92.40%	_cds_YP_009279794.1_17
NC_041935.1	***contig_596***	98.30%	516.336	1.65E-145	99.10%	_cds_YP_009601938.1_62
NC_041935.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009601937.1_61
NC_042057.1	***contig_596***	99.80%	876.432	0	99.90%	_cds_YP_009617226.1_29
NC_042057.1	***contig_596***	100.00%	854.272	0	100.00%	_cds_YP_009617227.1_30
NC_042057.1	***contig_596***	100.00%	987.231	0	100.00%	_cds_YP_009617229.1_32
NC_042057.1	***contig_596***	99.70%	593.895	8.26E-169	99.80%	_cds_YP_009617225.1_28
NC_042057.1	***contig_596***	99.30%	532.955	1.64E-150	99.70%	_cds_YP_009617228.1_31
NC_042057.1	***contig_596***	100.00%	383.377	1.18E-105	100.00%	_cds_YP_009617230.1_33
NC_049917.1	***contig_596***	84.00%	344.597	1.36E-93	80.70%	_cds_YP_009907801.1_14
NC_049918.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009907919.1_54
NC_049918.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009907918.1_53
NC_049918.1	***contig_596***	95.70%	185.785	4.35E-46	71.40%	_cds_YP_009907920.1_55
NC_049919.1	***contig_596***	99.00%	527.415	7.64E-149	99.50%	_cds_YP_009907999.1_54
NC_049919.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009907997.1_52
NC_049919.1	***contig_596***	81.70%	152.546	3.22E-36	90.90%	_cds_YP_009907998.1_53
NC_049920.1	***contig_596***	99.00%	527.415	7.64E-149	99.50%	_cds_YP_009908083.1_59
NC_049920.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009908081.1_57
NC_049920.1	***contig_596***	95.70%	185.785	4.35E-46	71.40%	_cds_YP_009908084.1_60
NC_049920.1	***contig_596***	81.70%	152.546	3.22E-36	90.90%	_cds_YP_009908082.1_58
NC_049921.1	***contig_596***	99.00%	527.415	7.64E-149	99.50%	_cds_YP_009908158.1_49
NC_049921.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009908156.1_47
NC_049921.1	***contig_596***	95.70%	185.785	4.35E-46	71.40%	_cds_YP_009908159.1_50
NC_049921.1	***contig_596***	81.70%	152.546	3.22E-36	90.90%	_cds_YP_009908157.1_48
NC_049922.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009908238.1_54
NC_049922.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009908237.1_53
NC_049922.1	***contig_596***	93.20%	108.226	7.85E-23	64.60%	_cds_YP_009908239.1_55
NC_049923.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009908308.1_37
NC_049923.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009908306.1_35

NC_049923.1	***contig_596***	95.20%	163.625	1.93E-39	69.80%	_cds_YP_009908309.1_38
NC_049923.1	***contig_596***	82.30%	158.086	6.91E-38	91.10%	_cds_YP_009908307.1_36
NC_049924.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009908401.1_60
NC_049924.1	***contig_596***	87.60%	457.243	1.66E-127	86.40%	_cds_YP_009908399.1_58
NC_049924.1	***contig_596***	84.90%	185.785	3.17E-46	92.50%	_cds_YP_009908400.1_59
NC_049924.1	***contig_596***	95.20%	163.625	1.93E-39	69.80%	_cds_YP_009908402.1_61
NC_049925.1	***contig_596***	98.60%	521.875	3.56E-147	99.30%	_cds_YP_009908484.1_59
NC_049925.1	***contig_596***	84.70%	398.15	1.02E-109	85.60%	_cds_YP_009908483.1_58
NC_049925.1	***contig_596***	93.20%	108.226	7.85E-23	64.60%	_cds_YP_009908485.1_60
NC_049926.1	***contig_596***	96.60%	793.333	0	98.30%	_cds_YP_009908530.1_17
NC_049926.1	***contig_596***	98.50%	571.735	3.87E-162	99.20%	_cds_YP_009908529.1_16
NC_049926.1	***contig_596***	96.50%	187.632	2.26E-46	61.10%	_cds_YP_009908531.1_18
NC_049926.1	***contig_596***	100.00%	176.552	1.57E-43	80.40%	_cds_YP_009908528.1_15
NC_049941.1	***contig_596***	84.30%	350.137	2.92E-95	80.80%	_cds_YP_009909082.1_37
NC_049941.1	***contig_596***	84.30%	171.012	8.88E-42	90.00%	_cds_YP_009909083.1_38
NC_049941.1	***contig_596***	85.90%	150.699	1.46E-35	73.90%	_cds_YP_009909084.1_39
NC_049943.1	***contig_596***	90.60%	156.239	6.36E-37	58.70%	_cds_YP_009909236.1_46
NC_049943.1	***contig_596***	85.80%	150.699	1.46E-35	73.80%	_cds_YP_009909239.1_49
NC_049944.1	***contig_596***	86.50%	156.239	3.14E-37	74.20%	_cds_YP_009909316.1_51
NC_049944.1	***contig_596***	90.60%	156.239	6.36E-37	58.70%	_cds_YP_009909312.1_47
NC_049945.1	***contig_596***	90.60%	156.239	6.36E-37	58.70%	_cds_YP_009909389.1_46
NC_049945.1	***contig_596***	85.80%	150.699	1.46E-35	73.80%	_cds_YP_009909393.1_50
NC_049946.1	***contig_596***	79.70%	263.345	3.70E-69	82.10%	_cds_YP_009909488.1_67
NC_049947.1	***contig_596***	99.10%	832.113	0	99.60%	_cds_YP_009909561.1_70
NC_049947.1	***contig_596***	93.20%	433.236	1.71E-120	96.60%	_cds_YP_009909562.1_71
NC_049947.1	***contig_596***	96.90%	324.284	6.80E-88	97.90%	_cds_YP_009909563.1_72
NC_049947.1	***contig_596***	94.60%	259.651	1.81E-68	92.90%	_cds_YP_009909564.1_73
NC_049948.1	***contig_596***	98.70%	821.033	0	99.40%	_cds_YP_009909632.1_68
NC_049948.1	***contig_596***	93.20%	433.236	1.81E-120	94.20%	_cds_YP_009909633.1_69
NC_049949.1	***contig_596***	99.40%	837.653	0	99.70%	_cds_YP_009909704.1_67
NC_049949.1	***contig_596***	93.50%	438.776	3.89E-122	94.30%	_cds_YP_009909705.1_68
NC_049949.1	***contig_596***	95.20%	265.191	4.11E-70	91.10%	_cds_YP_009909707.1_70
NC_049949.1	***contig_596***	96.30%	222.718	1.67E-57	98.10%	_cds_YP_009909706.1_69
NC_049950.1	***contig_596***	99.00%	350.137	1.12E-95	99.50%	_cds_YP_009909708.1_1
NC_049951.1	***contig_596***	99.80%	876.432	0	99.90%	_cds_YP_009909825.1_60
NC_049951.1	***contig_596***	100.00%	854.272	0	100.00%	_cds_YP_009909826.1_61
NC_049951.1	***contig_596***	100.00%	987.231	0	100.00%	_cds_YP_009909829.1_64
NC_049951.1	***contig_596***	99.70%	593.895	8.26E-169	99.80%	_cds_YP_009909823.1_58
NC_049951.1	***contig_596***	99.70%	582.815	1.75E-165	99.80%	_cds_YP_009909824.1_59
NC_049951.1	***contig_596***	99.30%	532.955	1.64E-150	99.70%	_cds_YP_009909828.1_63
NC_049951.1	***contig_596***	100.00%	383.377	1.18E-105	100.00%	_cds_YP_009909830.1_65
NC_049951.1	***contig_596***	100.00%	361.217	5.18E-99	100.00%	_cds_YP_009909831.1_66
NC_049951.1	***contig_596***	100.00%	339.057	2.26E-92	100.00%	_cds_YP_009909827.1_62
NC_049951.1	***contig_596***	100.00%	211.638	2.93E-54	100.00%	_cds_YP_009909832.1_67
NC_049952.1	***contig_596***	86.60%	278.118	1.51E-73	68.40%	_cds_YP_009909897.1_65
NC_049953.1	***contig_596***	83.70%	383.377	2.51E-105	91.80%	_cds_YP_009909969.1_71
NC_049953.1	***contig_596***	95.10%	257.805	8.60E-68	82.40%	_cds_YP_009909970.1_72
NC_049953.1	***contig_596***	93.90%	250.418	1.11E-65	91.30%	_cds_YP_009909971.1_73
NC_049953.1	***contig_596***	95.70%	189.479	6.27E-47	61.20%	_cds_YP_009909966.1_68
NC_049954.1	***contig_596***	99.40%	837.653	0	99.70%	_cds_YP_009910035.1_64
NC_049954.1	***contig_596***	93.50%	438.776	3.68E-122	96.80%	_cds_YP_009910036.1_65
NC_049954.1	***contig_596***	97.40%	333.517	1.13E-90	98.70%	_cds_YP_009910037.1_66
NC_049955.1	***contig_596***	99.10%	832.113	0	99.60%	_cds_YP_009910102.1_65

NC_049955.1	***contig_596***	93.20%	433.236	1.71E-120	96.60%	_cds_YP_009910103.1_66
NC_049955.1	***contig_596***	96.90%	327.977	5.26E-89	98.50%	_cds_YP_009910105.1_68
NC_049456.2	***contig_686***	75.80%	106.379	1.34E-21	47.20%	_cds_YP_009883012.1_53
NC_019406.1	***contig_689***	76.20%	106.379	9.81E-22	51.60%	_cds_YP_006988375.1_141
NC_048046.1	***contig_689***	84.80%	111.919	2.11E-23	50.90%	_cds_YP_009809318.1_149
NC_047790.1	***contig_88***	87.20%	250.418	1.52E-65	88.10%	_cds_YP_009787991.1_22

3.5 – BLAST results for I. marina vs CroV protein sequences

Best hit BLAST results for CroV protein sequences against the I. marina predicted peptide sequences (e<10-20)

Query Name	Hit Name	% Identical Sites	Bit-Score	E Value
tr|E3T5H4|E3T5H4_CROVB	contig_193	57.20%	681.019	0
tr|E3T4K4|E3T4K4_CROVB	contig_339	39.90%	456.062	3.12E-139
tr|E3T4I7|E3T4I7_CROVB	contig_423	36.40%	379.793	1.04E-114
tr|E3T5M3|E3T5M3_CROVB	contig_339	53.80%	343.969	1.20E-106
tr|E3T4R9|E3T4R9_CROVB	contig_316	39.80%	344.739	1.62E-104
tr|E3T5S6|E3T5S6_CROVB	contig_339	27.60%	343.969	1.36E-98
tr|E3T5E1|E3T5E1_CROVB	contig_131	40.10%	322.398	2.91E-98
tr|E3T4Y8|E3T4Y8_CROVB	contig_339	49.20%	291.197	1.87E-88
tr|E3T5D9|E3T5D9_CROVB	contig_217	27.80%	317.005	1.30E-86
tr|E3T4J5|E3T4J5_CROVB	contig_339	34.60%	286.189	2.45E-83
tr|E3T596|E3T596_CROVB	contig_115	44.20%	288.115	3.62E-82
tr|E3T5M5|E3T5M5_CROVB	contig_172	41.70%	289.656	1.13E-78
tr|E3T4F6|E3T4F6_CROVB	contig_689	32.70%	276.944	2.05E-77
tr|E3T4N5|E3T4N5_CROVB	contig_1	34.70%	255.373	1.04E-73
tr|E3T4Z4|E3T4Z4_CROVB	contig_238	46.10%	252.677	8.06E-67
tr|E3T5P9|E3T5P9_CROVB	contig_237	35.40%	247.669	3.33E-66
tr|E3T507|E3T507_CROVB	contig_630	33.10%	241.121	1.24E-64
tr|E3T4M6|E3T4M6_CROVB	contig_339	43.20%	214.542	1.30E-62
tr|E3T4G3|E3T4G3_CROVB	contig_339	37.90%	226.483	1.44E-62
tr|E3T537|E3T537_CROVB	contig_327	42.60%	220.705	1.01E-61
tr|E3T575|E3T575_CROVB	contig_310	26.30%	202.216	1.27E-54
tr|E3T5Q0|E3T5Q0_CROVB	contig_310	30.90%	194.512	1.53E-53
tr|E3T4W6|E3T4W6_CROVB	contig_201	41.90%	186.808	4.54E-52
tr|E3T5G2|E3T5G2_CROVB	contig_221	47.60%	195.667	3.55E-51
tr|E3T5S3|E3T5S3_CROVB	contig_689	39.00%	179.489	5.98E-48
tr|E3T4W5|E3T4W5_CROVB	contig_201	45.80%	174.096	7.09E-48
tr|E3T5F3|E3T5F3_CROVB	contig_217	42.40%	176.792	1.33E-44

tr|E3T5J8|E3T5J8_CROVB	contig_347	24.60%	172.555	6.62E-43
tr|E3T5H8|E3T5H8_CROVB	contig_1	47.10%	148.288	1.76E-40
tr|E3T5C1|E3T5C1_CROVB	contig_158	85.30%	138.272	3.25E-39
tr|E3T5Q7|E3T5Q7_CROVB	contig_483	25.70%	155.992	1.22E-37
tr|E3T5C0|E3T5C0_CROVB	scaffold_143	42.00%	138.272	6.50E-37
tr|E3T4F7|E3T4F7_CROVB	contig_275	33.50%	142.51	7.73E-37
tr|E3T5K6|E3T5K6_CROVB	contig_268	26.70%	150.599	5.41E-36
tr|E3T4R8|E3T4R8_CROVB	contig_327	39.90%	137.117	5.45E-36
tr|E3T4W8|E3T4W8_CROVB	contig_164	46.60%	142.895	6.09E-34
tr|E3T551|E3T551_CROVB	contig_239	47.80%	125.946	1.75E-33
tr|E3T4P8|E3T4P8_CROVB	contig_194	26.20%	133.265	3.92E-33
tr|E3T4S5|E3T4S5_CROVB	contig_131	25.20%	135.191	1.79E-31
tr|E3T4Q7|E3T4Q7_CROVB	contig_88	28.70%	135.576	2.34E-31
tr|E3T511|E3T511_CROVB	contig_686	27.50%	132.494	6.94E-31
tr|E3T4M2|E3T4M2_CROVB	contig_239	41.20%	117.857	1.01E-30
tr|E3T5B8|E3T5B8_CROVB	contig_136	40.70%	88.1965	2.86E-30
tr|E3T4P7|E3T4P7_CROVB	contig_413	31.60%	124.79	5.41E-30
tr|E3T4G9|E3T4G9_CROVB	contig_254	30.70%	121.324	2.10E-28
tr|E3T4H8|E3T4H8_CROVB	contig_94	43.40%	108.997	1.59E-27
tr|E3T4N3|E3T4N3_CROVB	contig_495	29.50%	114.775	1.93E-25
tr|E3T4T1|E3T4T1_CROVB	scaffold_143	32.10%	101.293	2.85E-24
tr|E3T5R5|E3T5R5_CROVB	contig_139	25.40%	110.538	1.19E-23
tr|E3T535|E3T535_CROVB	contig_560	29.20%	103.219	1.77E-22
tr|E3T5N2|E3T5N2_CROVB	contig_400	31.30%	99.3673	5.92E-22
tr|E3T5M8|E3T5M8_CROVB	contig_193	24.50%	99.3673	7.18E-22
tr|E3T5R8|E3T5R8_CROVB	contig_115	32.80%	105.916	8.10E-22
tr|E3T4K2|E3T4K2_CROVB	contig_131	33.60%	88.9669	2.29E-20
tr|E3T5N3|E3T5N3_CROVB	contig_595	22.80%	98.5969	2.48E-20
tr|E3T5R4|E3T5R4_CROVB	contig_164	57.40%	90.1225	6.67E-20
tr|E3T4N8|E3T4N8_CROVB	contig_88	25.80%	95.5153	1.46E-18
tr|E3T5C5|E3T5C5_CROVB	contig_339	33.30%	82.8037	1.67E-18
tr|E3T4S2|E3T4S2_CROVB	contig_310	44.40%	83.5741	2.04E-18
tr|E3T4F1|E3T4F1_CROVB	contig_147	24.50%	88.1965	9.43E-18
tr|E3T502|E3T502_CROVB	contig_2	25.60%	87.4261	4.33E-17

tr|E3T4V2|E3T4V2_CROVB	contig_415	25.30%	84.3445	8.72E-17
tr|E3T5K0|E3T5K0_CROVB	scaffold_550	36.20%	77.7962	9.87E-17
tr|E3T512|E3T512_CROVB	contig_689	28.10%	81.2629	1.80E-16
tr|E3T4R7|E3T4R7_CROVB	contig_327	23.90%	80.8777	1.17E-15
tr|E3T5P6|E3T5P6_CROVB	scaffold_143	35.20%	78.5666	1.41E-15
tr|E3T5J5|E3T5J5_CROVB	contig_316	26.30%	81.2629	2.57E-15
tr|E3T5H3|E3T5H3_CROVB	contig_679	34.00%	83.1889	2.81E-15
tr|E3T4Y9|E3T4Y9_CROVB	contig_177	23.10%	77.0258	6.24E-15
tr|E3T5U7|E3T5U7_CROVB	contig_537	27.60%	78.5666	5.38E-14
tr|E3T5Q5|E3T5Q5_CROVB	contig_686	33.60%	70.4774	6.50E-14
tr|E3T4N2|E3T4N2_CROVB	contig_238	31.80%	71.2478	7.32E-14
tr|E3T4H1|E3T4H1_CROVB	contig_415	35.30%	72.7886	3.24E-13
tr|E3T5P2|E3T5P2_CROVB	contig_467	29.80%	67.781	4.12E-13
tr|E3T4D5|E3T4D5_CROVB	contig_371	31.60%	75.8702	5.02E-13
tr|E3T4R3|E3T4R3_CROVB	contig_215	32.70%	67.3958	1.16E-12
tr|E3T4J2|E3T4J2_CROVB	contig_679	28.50%	72.4034	4.21E-12
tr|E3T4F8|E3T4F8_CROVB	contig_310	24.20%	71.633	6.03E-12
tr|E3T578|E3T578_CROVB	contig_157	50.00%	60.077	6.32E-12
tr|E3T4I5|E3T4I5_CROVB	contig_221	30.30%	64.3142	1.27E-11
tr|E3T539|E3T539_CROVB	contig_2	26.30%	65.4698	2.25E-11
tr|E3T4R4|E3T4R4_CROVB	contig_237	30.90%	63.5438	2.36E-11
tr|E3T587|E3T587_CROVB	contig_520	20.60%	68.1662	3.71E-11
tr|E3T4D0|E3T4D0_CROVB	contig_467	27.70%	65.0846	4.61E-11
tr|E3T4Y6|E3T4Y6_CROVB	contig_495	39.10%	68.5514	5.47E-11