

EXPANDING OUR UNDERSTANDING OF MARINE MICROBIOTA IN
THE NORTHWEST ATLANTIC AND ARCTIC SECTORS: FROM COASTAL
PHYTOPLANKTON TO DIAZOTROPH COMMUNITIES

by

Brent M. Robicheau

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
March 2023

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

© Copyright by Brent M. Robicheau, 2023

*This work is dedicated to Norbert and Anita Robicheau,
who are known to my family and myself as Grand-père and Grand-mère,
and who had the idea to bring me to the ocean in the first place.*

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	xii
LIST OF ABBREVIATIONS AND SYMBOLS USED	xiii
ACKNOWLEDGEMENTS	xvii
CHAPTER 1 Introduction	1
1.1 Chloroplast and Nitrogenase Molecular Markers	3
1.2 <i>Cand. Atelocyanobacterium thalassa</i> (UCYN-A) and its Ecotypes	4
1.3 The Bedford Basin Time Series Monitoring Program	5
1.4 Newer Areas of Diazotroph Research: Coastal and High-latitude Regions	6
1.5 Overview of Thesis	7
CHAPTER 2 Highly-resolved Interannual Phytoplankton Community Dynamics of the Coastal Northwest Atlantic	10
2.1 Abstract	10
2.2 Introduction	11
2.3 Materials and Methods	12
2.3.1 Sampling, Oceanographic Data, and Flow Cytometry	12
2.3.2 DNA Extraction and Sequencing	13
2.3.3 Data Analyses	15
2.4 Results	16
2.4.1 Temperature and Chlorophyll <i>a</i>	16
2.4.2 Broader Taxonomic Groups and Flow Cytometry	17
2.4.3 Individual ASV Profiles	21
2.4.4 Comparisons to the Scotian Shelf	24
2.4.5 Small Phytoplankton and their link to Atypical Temperature Conditions	26
2.5 Discussion	28
2.5.1 <i>Synechococcus</i> and <i>Eutreptiella</i> are Important Phytoplankton <i>16S</i> rRNA Gene Signatures in the Bedford Basin	28
2.5.2 Phytoplankton tracked by the Bedford Basin Time-series are Globally Relevant	29
2.5.3 Additional Insights gained from High-frequency DNA Sampling	32

2.6	Conclusions	33
2.7	Acknowledgements	34
2.8	Author contributions	34
2.9	Supplemental Information for Chapter 2	35
2.9.1	Supplemental Text for Chapter 2.....	35
2.9.2	Supplemental Tables for Chapter 2	42
2.9.3	Supplemental Figures for Chapter 2.....	45
CHAPTER 3	Microevolutionary Patterns in Ecotypes of the Symbiotic Cyanobacterium UCYN-A Revealed from a Coastal Time- series in the Northwest Atlantic	58
3.1	Abstract	58
3.2	Introduction.....	59
3.3	Materials and Methods.....	62
3.3.1	Sampling and Oceanographic Data	62
3.3.2	DNA extraction and amplicon sequencing.....	63
3.3.3	Quantitative-PCR Assays	65
3.3.4	Data Analysis.....	66
3.4	Results.....	67
3.4.1	Diversity and Evolutionary Dynamics.....	67
3.4.2	High-resolution Seasonal Patterns within the Coastal NW Atlantic ..	72
3.4.3	<i>In situ</i> Temperature affected the Dominance of A1 versus A2 Ecotypes.....	74
3.4.4	UCYN-A Algal Hosts are Amongst Co-occurring Phytoplankton Community	77
3.5	Discussion	79
3.5.1	Multiple UCYN-A Ecotypes Inhabit a Coastal Basin in the NWA ...	79
3.5.2	Phytoplankton Co-occurrences Suggest Presence of UCYN-A/ Host Consortium.....	79
3.5.3	UCYN-A Temporal Variability and Microevolutionary Patterns	81
3.6	Conclusions.....	84
3.7	Acknowledgements	85
3.8	Author Contributions	86
3.9	Supplemental Information for Chapter 3	86

3.9.1	Supplemental Text for Chapter 3.....	86
3.9.2	Supplemental Tables for Chapter 3	90
3.9.3	Supplemental Figures for Chapter 3	92
CHAPTER 4	Marine Nitrogen-fixers in the Canadian Arctic Gateway are Dominated by Biogeographically Distinct Non-Cyanobacterial Communities	106
4.1	Abstract	106
4.2	Introduction.....	107
4.3	Methods.....	110
4.3.1	Cruise Samples and their Environmental Data.....	110
4.3.2	Coastal NW Atlantic Time-series Samples and their Environmental Data	112
4.3.3	DNA Extractions, <i>nifH</i> Amplicon Sequencing, and UCYN-A Quantitative-PCRs	112
4.3.4	Data Analyses	114
4.4	Results.....	117
4.4.1	Diazotroph Biogeography within the Canadian Arctic Gateway	117
4.4.2	Dominant CAG Diazotrophs and their Environmental Links	124
4.4.3	UCYN-A within the Canadian Arctic Gateway	128
4.4.4	Searching for Dominant Canadian Arctic Gateway Diazotrophs Outside of the CAG & Polar Realm	130
4.5	Discussion	134
4.5.1	Taxonomy and Predicted Ecology of NCDs within the CAG.....	134
4.5.2	UCYN-A and other Cyanobacterial Diazotrophs within the Canadian Arctic Gateway	138
4.5.3	Biogeographic Division of Diazotrophs between Labrador Sea and Baffin Bay/CAA	139
4.6	Conclusions.....	140
4.7	Acknowledgements.....	141
4.8	Author Contributions	142
4.9	Supplemental Information.....	142
4.9.1	Supplemental Text.....	142
4.9.2	Supplemental Tables for Chapter 4	143
4.9.3	Supplemental Figures for Chapter 4.....	149

CHAPTER 5	Conclusions	161
5.1	Overview of Thesis Conclusions	161
5.2	Opportunities for Future Research.....	163
BIBLIOGRAPHY		169
Appendix A	Chapter 2 Copyright Permission	207
Appendix B	Supplemental Datasets	208

LIST OF TABLES

Table 2.1.	Small taxa associated with trends in 2016 (values are for 1–10m; unrarefied data).	28
Table 2.2.	Output of indicator species test.	42
Table 2.3.	List of dominant ASVs that were manually identified to genus/species level.	43
Table 2.4.	Additional ASVs dominant on the Scotian Shelf that were also manually identified.....	44
Table 3.1.	Z-test of selection for rarer UCYN-A ASVs.....	90
Table 3.2.	Significance of environmental predictors determined using the envfit function in <i>vegan</i> (Oksanen et al., 2022).	91
Table 4.1.	Summary of multi-level pattern analyses for dominant ASVs (or ASV groups) in CAG GN02 cruise based on sample groupings.	126
Table 4.2.	Significance of environmental parameters calculated using <i>envfit</i> (Oksanen et al., 2022).	143
Table 4.3.	Additional alignment scores for dominant <i>nifH</i> ASVs in the CAG based on comparison to <i>nr/nt</i> database in NCBI (Acland et al., 2014) via BLAST (Johnson et al., 2008) and Kapili & Dekas (2021) via standalone BLAST (Altschul et al., 1990).	144
Table 4.4.	Alignment scores for best matches between CAG <i>nifH</i> ASV sequences versus the dominant <i>nifH</i> OTUs reported for the western Arctic Ocean (Shiozaki et al., 2018), as well as to the <i>nifH</i> from low-latitude MAGs	146
Table 4.5.	Individual ASV scores for multi-level pattern analyses.....	148

LIST OF FIGURES

Figure 2.1.	Phytoplankton counts for 5-yr (1–10m) Bedford Basin (Halifax, NS) time-series.	17
Figure 2.2.	Weekly dynamics and seasonal patterns for major phytoplankton groups observed over the 4-yr timeseries in Bedford Basin (Halifax, NS).	18
Figure 2.3.	Individual temporal relative abundances profiles for Top 20 phytoplankton <i>16S</i> rRNA ASVs (chloroplast + cyanobacterial <i>16S</i> rRNA) in the Bedford Basin time-series from 2014–2017.	22
Figure 2.4.	Comparison of top Scotian Shelf phytoplankton ASVs versus ASVs that were observed in the Bedford Basin (BB) time-series.	25
Figure 2.5.	Warmer temperatures and trends for smaller phytoplankton during 2016.	27
Figure 2.6.	Diagram of flow cytometry gates used.	45
Figure 2.7.	Frequency distributions for the final number of chloroplast & cyanobacterial <i>16S</i> rRNA reads per sample.	46
Figure 2.8.	Rarefaction curves for the three <i>16S</i> rRNA datasets used in our study. ...	47
Figure 2.9.	Network analysis of microbial associations for Top Twenty Bedford Basin ASVs between V4-V5 and V6-V8.	48
Figure 2.10.	Weekly/seasonal phytoplankton trends in the Bedford Basin at 1, 5, & 10m depths as measured by unrarefied chloroplast and cyanobacterial <i>16S</i> rRNA relative abundances.	49
Figure 2.11.	Rarefied abundances for all three surface depths (1,5, and 10m).	50
Figure 2.12.	Analysis of Bray-Curtis Similarities between samples and then visualized according to the number of weeks between samples.	51
Figure 2.13.	Analysis of Bray-Curtis Similarities between samples and then visualized according to the number of weeks between samples.	52
Figure 2.14.	Oceanographic distributions (via <i>Tara</i> Oceans miTAGs) for cyanobacterial and Euglenozoa indicator species that are present in the Bedford Basin during the fall and spring, respectively.	53
Figure 2.15.	Relationship between temperature and <i>E. pomquetensis</i> as detected by the Bedford Basin molecular time-series.	54
Figure 2.16.	Salinity anomalies and nutrient anomalies.	54
Figure 2.17.	Top Twenty phytoplankton ASVs observed in the Bedford Basin time series are also detected seasonally in nearby Atlantic Zone Monitoring Program (AZMP) stations along the Halifax Line (HL).	55
Figure 2.18.	Phylogenetic assessment of <i>Synechococcus</i> ecotypes.	56

Figure 2.19.	Additional time-series image of Bedford Basin flow cytometry.	57
Figure 2.20.	Relationship between <math><3\mu\text{m}</math> cell densities and temperature separated into yearly plots.	57
Figure 3.1.	UCYN-A Ecotypes in the Bedford Basin.	68
Figure 3.2.	Analysis of mutations observed in UCYN-A ASVs within the Bedford Basin (BB). Main ASVs have >2 reads in dataset, other ASVs have 2 reads.	70
Figure 3.3.	Weekly distribution of UCYN-A ecotypes in the Bedford Basin over three years.	73
Figure 3.4.	The occurrence of UCYN-A ecotypes within Bedford Basin surface waters (1–10m) in relation to environmental parameters.....	76
Figure 3.5.	Network analysis of UCYN-A <i>nifH</i> ASVs versus phytoplankton <i>16S</i> rRNA ASVs (based on chloroplast and cyanobacterial V4-V5 <i>16S</i> rRNA gene signatures).	78
Figure 3.6	Summaries of raw sequencing data.	92
Figure 3.7.	Workflow for retrieving UCYN-A ASVs from total <i>nifH</i> ASVs list.....	93
Figure 3.8.	Modified UCYN-A2 quantitative PCR (qPCR) assay.	94
Figure 3.9.	Distributions for final individual quantitative-PCR efficiencies for UCYN-A1 and UCYN-A2 assays (expressed as percentages).	95
Figure 3.10.	References for UCYN-A1 and UCYN-A2 gBlock gene fragments (IDT) in FASTA format.	95
Figure 3.11.	Relative abundance distributions for each UCYN-A ASV.	96
Figure 3.12.	Maximum Likelihood tree of UCYN-A ASVs, UCYN-A oligotypes, and other cyanobacterial diazotrophs.	96
Figure 3.13.	Neighbor-joining tree of UCYN-A phylotypes.	97
Figure 3.14.	Minimum spanning nucleotide substitution network for fifty-two UCYN-A oligotypes identified by Turk-Kubo et al. (2017).	98
Figure 3.15.	Nucleotide differences between UCYN-A ASVs reported herein (bolded) and UCYN-A oligotypes reported in Turk-Kubo et al. (2017).	99
Figure 3.16.	UCYN-A <i>nifH</i> ASVs present in Bedford Basin between 2015–2017	100
Figure 3.17.	CLR values for each of the 65 UCYN-A ASVs over time along with the number of nucleotide differences (No.diff) each observation has relative to the major ASV within each group.	101
Figure 3.18.	<i>NifH</i> amino acid site conservation within UCYN-A ecotypes relative to amino acid sites conserved within other non-UCYN-A diazotrophs..	102
Figure 3.19.	Shelf water 60m intrusion events during 2015–2017 Bedford Basin time series.....	103

Figure 3.20.	Complete network analysis of <i>nifH</i> versus phytoplankton <i>16S</i> rRNA (cp/cy/ <i>16S</i> rRNA gene).	104
Figure 3.21.	Maximum likelihood phylogeny of Haptophyta cp/ <i>16S</i> rRNA ASVs with strong and weak copresences to UCYN-A <i>nifH</i> ASVs.	104
Figure 3.22.	Spatiotemporal patterns for phytoplankton <i>16S</i> rRNA ASVs with weak copresence and mutual exclusions in relation to UCYN-A <i>nifH</i> ASVs.	105
Figure 4.1.	Sites sampled along the Canadian Arctic Gateway during the ArcticNet1502 (GN02) GEOTRACES expedition.	111
Figure 4.2.	Major <i>nifH</i> Classes present in Canadian Arctic Gateway during GN02 Expedition.	119
Figure 4.3.	Multivariate analysis of <i>nifH</i> communities for Canadian Arctic Gateway during GN02 expedition.....	121
Figure 4.4.	Oceanographic data for Canadian Arctic Gateway during 2015 GN02 expedition.	123
Figure 4.5.	Maximum Likelihood tree of major <i>nifH</i> amplicon sequence variants (ASVs) present in the Canadian Arctic Gateway during the GN02 expedition.	125
Figure 4.6.	Cyanobacterial diazotrophs detected in Canadian Arctic Gateway during GN02 cruise.	129
Figure 4.7.	Dominant ASVs recovered from the CAG within a 5-yr time series in the NWA.	131
Figure 4.8.	Relationship between CLR values and relative abundances calculated for each ASV within each sample.	149
Figure 4.9.	Upset plot showing the number ASVs unique to each sample and shared across sample in the Arctic <i>nifH</i> GN02 dataset.	149
Figure 4.10.	Major diazotroph Orders present in Canadian Arctic Gateway during GN02 GEOTRACES Expedition.	150
Figure 4.11.	RDA analysis presented in Fig. 4.3, but instead samples are visualized by size fraction.	151
Figure 4.12.	Additional oceanographic data from the Canadian Arctic Gateway GN02 expedition.	151
Figure 4.13.	Individual spatial profiles (based on CLR values) for dominant <i>nifH</i> ASVs detected in the Canadian Arctic Gateway during GN02 expedition and present in ≥ 7 samples.	152
Figure 4.14.	Individual spatial profiles (based on CLR values) for dominant <i>nifH</i> ASVs detected in the Canadian Arctic Gateway during GN02 expedition and present in < 7 samples.	153

Figure 4.15.	Raw CLR data distributions for dominant ASVs in the CAG compared to each other according to oceanographic data. For ASVs in ≥ 7 samples.....	154
Figure 4.16.	Number of total <i>nifH</i> ASVs from the Canadian Arctic Gateway that are recovered and not-recovered within the Bedford Basin time series..	155
Figure 4.17.	All dominant ASVs recovered from the Canadian Arctic Gateway within a 5-yr time series in the coastal NW Atlantic organized by taxonomic Order.....	156
Figure 4.18.	Individual temporal profiles (for CLR values) for all dominant Unknown Desulfuromonadales (a) and Unknown Desulfobulbaceae (b) ASVs from the Canadian Arctic Gateway recovered within the Bedford Basin times series.....	157
Figure 4.19.	Dominant <i>Stutzerimonas stutzeri</i> in Bedford Basin and correlation with oxygen.	158
Figure 4.20.	Distributions for Aitchison Distances calculated between all Bedford Basin <i>nifH</i> samples and all Arctic <i>nifH</i> samples (both size fractions) and divided out by year (left panels) and by month (right panels).	158
Figure 4.21.	Branch lengths for Maximum Likelihood phylogeny presented in the main text.....	159
Figure 4.22.	Rarefaction curves for raw sequencing data.	160
Figure 5.1.	Example of using thesis results and <i>16S</i> rRNA datasets to inform additional culturing experiments.....	165

ABSTRACT

High-resolution ocean time series and strategic sampling in understudied environments provide unique opportunities to advance our knowledge of microbial life in the ocean. The goal of this thesis was to expand our understanding of the marine microbiome by focusing on phytoplankton and diazotrophs within the Northwest Atlantic (NWA) and Arctic Ocean sectors through the use of environmental DNA collected from: (i) a weekly multiyear time series located in Bedford Basin (N.S., Canada; samples from 2014–2019), and (ii) the understudied Canadian Arctic Gateway (CAG) during Jul–Aug 2015. This thesis has three main research chapters. In Chapter 2, I describe the dominant phytoplankton present in the coastal NWA over four years, providing the first detailed molecular picture of phytoplankton biodiversity across all seasons within this region of the Atlantic Ocean. In Chapter 3, I focus on the globally important diazotroph symbiont, *Candidatus Atelocyanobacterium thalassa* or ‘UCYN-A’ within the Bedford Basin over three years by describing the weekly dynamics of four ecotypes within this species (A1 to A4). Ultimately, findings from Chapter 3 advance our understanding of UCYN-A temporal dynamics within the coastal realm and further elucidate microevolutionary patterns demonstrated by ecotypes of this species. In Chapter 4, I move outside of the Bedford Basin and characterize UCYN-A and other diazotrophs within the CAG during the late summer and early fall of 2015. Importantly, findings from Chapter 4 provide an initial account of the possible biogeographic distributions that marine diazotrophs in the CAG can exhibit across the Labrador Sea, Baffin Bay, and Canadian Arctic Archipelago. Chapter 4 also increases the overall survey of marine diazotrophs within the Arctic (for which previous data are sparse) by elucidating the taxonomy and environmental conditions that were associated with diazotroph genetic signatures dominating the CAG diazotroph communities during 2015. Overall, Chapters 2 and 3 advance our understanding of phytoplankton and UCYN-A within the coastal NWA, while Chapter 4 advances our understanding of diazotroph diversity within the Arctic marine environment.

LIST OF ABBREVIATIONS AND SYMBOLS USED

# Pos. Obs.	Number of Positive Observations
°C	Degrees Celsius
ASV(s)	Amplicon Sequence (or sequencing) Variant(s)
ATP	Adenosine triphosphate
Ave	Average
Ave. Rel. Abun.	Average relative abundance
AZMP	Atlantic Zone Monitoring Program
BATS	Bermuda Atlantic Time-Series Study
BB	Bedford Basin (in Chapters 2–3) or Baffin Bay (Chapter 4)
BBMP	Bedford Basin Monitoring Program
BIO	Bedford Institute of Oceanography
BLAST	Basic Local Alignment Search Tool
bp	base-pairs
BSA	Bovine serum albumin
CAA	Canadian Arctic Archipelago
CAG	Canadian Arctic Gateway
Chl or chl	Chlorophyll
CLR	Centered log-ratio
cov.	Coverage or query coverage
cp16S rRNA	Chloroplast 16S rRNA gene
cpDNA	Chloroplast DNA
cy16S rRNA	Cyanobacterial 16S rRNA gene
D	Depth (see Table 4.1)
D-Fe	Dissolved iron
D-Mn	Dissolved manganese
DIN	Dissolved inorganic nitrogen
d_N	Number of nonsynonymous substitutions per nonsynonymous sites
dNTPs	Deoxyribonucleotide triphosphates
DOC/DOM	Dissolved organic carbon/dissolved organic matter

d_s	Number of synonymous substitutions per synonymous sites
dsDNA	Double stranded DNA
E1 and E2	Event 1 and Event 2
eDNA	Environmental DNA
Fluo or F	Fluorescence
FSC	Forward scatter
GTDB	Genome taxonomy database
H	High (see Table 4.1)
h	Hours
H_a	Alternative hypothesis
HCl	Hydrochloric acid
HL	Halifax Line
H_o	Null hypothesis
HOT	Hawaii Ocean Time-series
HS	High-sensitivity
I	Intrusion events of the Bedford Basin (e.g., as in I14)
L	Litres (when used as volume) or Low (see Table 4.1)
LAT or Lat	Latitude
LON or Lon	Longitude
LOQ	Limit of Quantification
Lr	Large size fraction
m	Meters
mg	Milligrams
miTAGs	Metagenomic <i>16S</i> rRNA Illumina tags
ML	Maximum Likelihood
mL	Millilitres
n	Number or sample size
nanoSIMS	Nanometer scale secondary ion mass spectrometry
NCBI	National Center for Biotechnology Information
NCD(s)	Non-cyanobacterial diazotroph(s)
NGDC	National Geophysical Data Center

Nitro or N	Nitrospirae (see Fig. 4.15)
NMDS	Non-metric multidimensional scaling
No.diff	Number of nucleotide differences
NOAA	National Oceanic and Atmospheric Administration
<i>nr/nt</i>	Nucleotide collection database of NCBI
NS	Nova Scotia
Nt. Diff.	Nucleotide differences
NWA	Northwest Atlantic
oligo(s) or OL	Oligotype(s)
OTU	Operational Taxonomic Unit
Oxygen or O ₂	Dissolved Oxygen
PE	Phycoerythrin
PI	Pair-wise identity
POC	Particulate organic carbon
PON	Particulate organic nitrogen
PSU	Practical Salinity Unit
qPCR	Quantitative PCR
RDA	Redundancy Analysis
Rel. Abun.	Relative abundance
rRNA	Ribosomal RNA
S	Size (see Table 4.1)
sal or Sa	Salinity
SCM	Subsurface chlorophyll maximum
SD	Standard Deviation
Sm	Small size fraction
Temp or T	Temperature in degrees Celsius (also see Table 4.1)
TP-Fe	Total particulate iron
TP-P	Total particulate phosphate
TP-V	Total particulate vanadium
UCYN-A	<i>Candidatus Atelocyanobacterium thalassa</i>
μL	Microlitres

μm	micrometers
μM	Micromolar
Unk. Bac.	Unknown Bacteria
Unk. Proteo.	Unknown Proteobacteria
UNKN	Unknown
V6-V8 or V4-V5	Variable regions 6 to 8 or 4 to 5 of rRNA gene

ACKNOWLEDGEMENTS

I would like to send my biggest thank you to my PhD supervisor, Dr. Julie LaRoche, for all the stellar advice, guidance, and opportunities that she has provided me while completing my degree. Her lessons and mentorship have turned me into a bona fide marine microbiologist and for that I will always be grateful. I am also thankful that during the course of this thesis Dr. Jenni Tolman (LaRoche lab manager) was a driving force in helping ensure that my data collection was completed with ease, which was no easy feat considering the 1000s of PCR reactions that have come together to form the heart of this thesis. Additionally, the research experiences that I have had at Dalhousie have been greatly enriched by all the wonderful LaRoche lab members that I worked with over the past few years, including Ian Luddington, Dr. Joerg Behnke, Sonja Rose, and Connor Mackie. From the LaRoche lab I would also like to thank Dr. Dhvani Desai and Dr. Jenni-Marie Ratten for providing me with advice early on regarding data analyses and bioinformatics.

With respect to sample collections, this thesis relies heavily on multiyear time series that are maintained by the Bedford Institute of Oceanography (BIO); from BIO, I would like to thank Andrew Cogswell and Dr. Emmanuel Devred for answering any and all questions that I had about the Bedford Basin time series. Likewise, Dr. Doug Wallace, Subhadeep Rakshit, and Dr. Chris Algar from Dalhousie's Oceanography Department are thanked for answering any and all questions that I had about the oceanographic chemical and physical properties of the Bedford Basin. I would also like to thank Dr. Andre Comeau at the IMR centre (Dalhousie) for being such a great wealth of knowledge when it came to sequencing our DNA samples. Furthermore, I am appreciative of all the excellent encouragement and advice that I received from my supervisory committee members Dr. Erin Bertrand and Dr. Robert Beiko throughout the course of my PhD research.

While the above individuals helped make my doctoral research a rewarding experience, there is an equally large number of individuals who also enriched my time as a PhD student and helped me navigate graduate school. This includes the biology office administration (Carolyn Young, Aileen Patterson, Julie Walker, and Chris MacNeil), Dr. Daniel Ruzzante, Dr. Sophia Stone, and Dr. Alastair Simpson. Outside of the LaRoche lab,

fellow graduate students in the Bentzen, Ruzzante, Simpson, and Bertrand labs were also a major support during my studies; especially, James Kho, Lisette Delgado, Sarah Salisbury, Loay Jabre, Cat Bannon, Megan Roberts, Scott McCain, Scott Pollar, Elden Rowland, Ellie Weise, Yana Eglit, Noor Youssef, and Beth Watson. I would also like to thank Debra Grantham and Andrew Schofield, who towards the end of my studies provided me with new opportunities to further my professional growth as an educator through a teaching assistantship in BIOL2030.

Finally, I would like to thank all of the those outside of Dalhousie that provided me with encouragement, including my family who offered their unwavering support throughout my entire time as graduate student, my friends (Shauna, Marc, Freya, Emily, & Mili), and my past supervisors who continue to share their wisdom and advice (Dr. M. Snyder, Dr. D. Stewart, H. d'Entremont, and Dr. A. Walker).

CHAPTER 1

Introduction

Phytoplankton and diazotrophs are critical for supporting life in the ocean through their respective roles in the carbon and nitrogen cycles. Phytoplankton grow photoautotrophically by using sunlight to sequester CO₂ via oxygenic photosynthesis—a process that represents a large fraction of carbon fixation globally (~40%; Falkowski, 1994, Falkowski et al., 2004). The fate of phytoplankton in the ocean (and therefore the fixed-C from these microorganisms) is multifaceted and can include grazing by larger plankton (e.g., zooplankton), the release of dissolved/particulate organic matter via processes such as cell death, exudation, and viral lysis, as well as contributions to the biological carbon pump via sinking organic particles (Calbet & Landry, 2004; Guidi et al., 2016; López-Sandoval et al., 2013; Resovsky et al., 1999; Wilhelm & Suttle, 1999). Additionally, organic materials from phytoplankton can make their way into the microbial loop, hence contributing to additional carbon recycling within the water column (Azam et al., 1983). In contrast, diazotrophs are important for the input of new nitrogen in the ocean. Specifically, diazotrophs contribute to the nitrogen cycle by adding new fixed-nitrogen via the conversion of atmospheric dinitrogen gas (N₂) to ammonia (NH₃)—a process termed biological nitrogen fixation (Capone et al., 2005; Zehr & Kudela, 2010). This energy intensive reaction (requiring sixteen ATP molecules and eight electrons) is carried out by the nitrogenase enzyme, which all diazotrophs possess (Kim & Rees, 1994; LaRoche & Breitbarth, 2005). New nitrogen from diazotrophy can not only be assimilated into organic nitrogen (i.e., into biomass), but it can also be oxidized into other forms of inorganic nitrogen (e.g., nitrite and nitrate via nitrification), and returned to dinitrogen gas via anammox and denitrification pathways (Dalsgaard et al., 2005; Pajares & Ramos, 2019; Thamdrup & Dalsgaard, 2002; Ward, 1996; Wuchter et al., 2006; Yool et al., 2007; Zehr & Kudela, 2010).

Phytoplankton are widespread in the ocean's euphotic (or sunlit) zone and include important groups such as cyanobacteria (e.g., *Synechococcus* and *Prochlorococcus*),

haptophytes (e.g., *Emiliania huxleyi* and *Chrysochromulina*), Bacillariophyta (or diatoms, e.g., *Minidiscus* and *Chaetoceros*), as well as dinoflagellates (e.g., *Dinophysis* and *Karenia*; Brand et al., 2012; Chisholm et al., 1988; Flombaum et al., 2013; Holligan et al., 1993; Leblanc et al., 2018; Liu et al., 2009; Luca et al., 2019; Myung et al., 2006; Olson et al., 1990). There is some overlap between the diversity encapsulated by both phytoplankton and diazotrophic groups in that some diazotrophs also photosynthesize and can fix both carbon and nitrogen. Such diazotrophs include: (i) cyanobacteria belonging to *Trichodesmium*, UCYN-A, *Cyanothece*, and *Crocospaera*, and (ii) diatom-diazotroph associations belonging to both *Richelia* (hosted by *Hemianulus* and *Rhizosolenia*) and *Calothrix* (hosted by *Chaetoceros*), as well as recently discovered *EpSB/EcSB* (found within *Epithemia* spp.; Luo et al., 2012; Schvarcz et al., 2022). Several bacterial and archaeal groups of diazotrophs also fall outside of the known cyanobacterial groups above and these are collectively referred to as ‘non-cyanobacterial diazotrophs’ or ‘NCDs’. Although many NCDs are still considered heterotrophic bacteria (Delmont et al., 2021), this diazotroph group also includes some archaea (e.g., ANME-2; Dekas et al., 2009), as well as microbes that are not strictly heterotrophic (for e.g., the photoheterotrophic *Marichromatium* sp. (Smith et al., 2014) and the chemoautotrophic bivalve symbiont *Cand. Thiodiazotropha endolucinida* (König et al., 2016)), hence the more all-encompassing NCD term can be used to describe this fraction of the microbiome (Moisander et al., 2017; Turk-Kubo et al., 2022; Zehr et al., 2003). Although microbes belonging to phytoplankton and (or) diazotroph categories have broad functional roles with respect to carbon and nitrogen fixation, the organisms within these groups can differ with respect to their carbon export contributions (Tréguer et al., 2018) and their nitrogen fixation rates (Montoya et al., 2004; Turk-Kubo et al., 2014), and can have additional functions (for e.g., denitrification; Turk-Kubo et al. 2022). Consequently, it is important to study the community dynamics of both phytoplankton and diazotrophs in ocean. For instance, spatiotemporal patterns of phytoplankton and diazotroph relative abundances within naturally occurring microbial communities have the potential to advance our fundamental understanding of taxon-specific ecologies within these groups, and they may also provide empirical observations that can contribute to modelling efforts by elucidating where and when certain taxa are likely to occur in the ocean in terms of their biogeographic

distributions and dominance (examples include Djurhuus et al. (2020), Meiler et al. (2022), and Tang & Cassar (2019)).

1.1 Chloroplast and Nitrogenase Molecular Markers

Respectively, the diversity of phytoplankton and diazotrophs were characterized in this dissertation through the analysis of cyanobacterial plus chloroplast *16S* ribosomal RNA (rRNA) and through the analysis of the *nifH* gene. While the *16S* rRNA gene can directly be used to resolve the relative proportion of cyanobacteria via their genomic DNA (Giovannoni et al., 1990; Shi & Falkowski, 2008), this same gene can also be used to examine the relative abundance of other eukaryotic phytoplankton due to the endosymbiotic origin of chloroplast-derived DNA (Keeling, 2013; Zablen et al., 1975) and to the fact that some phytoplankton photosynthesize via kleptoplasty (for e.g., certain dinoflagellates; Gast et al., 2007; Karlusich et al., 2020). Meanwhile, the *nifH* gene, which codes for the dinitrogenase reductase subunit of the nitrogenase enzyme is used to characterize diazotrophs (Kim & Rees, 1994; Rubio & Ludden, 2005)—this gene is also part of the *nifHDK* operon that encodes proteins for the entire nitrogenase enzyme (Zehr et al., 2003). *NifH* is typically amplified using a set of degenerate primers and conventional PCR to elucidate diazotrophic community compositions, or it is amplified using more targeted quantitative PCR (qPCR) assays to enumerate the *nifH* gene copies belonging to certain diazotrophs within a given sample (e.g., Langlois et al., (2008)). *NifH* genetic signatures (or phylotypes) are known to reflect deep branching Clusters I–IV based on chemical elements that are specific to the nitrogenase proteins of different taxa (versus conventional Molybdenum containing nitrogenases), as well as the taxonomic domain of phylotypes and the metabolic lifestyle of members within each cluster [Cluster I = Molybdenum and some Vanadium containing nitrogenases, Cluster Ib = cyanobacterial diazotrophs, Cluster II = Fe containing nitrogenases (instead of Mo) and some Archaea, Cluster III = many strict anaerobes, Cluster IV = closely related non-*nifH* genes (includes chlorophyllide reductase)](Chien & Zinder, 1996; Riemann et al., 2010; Zehr et al., 2003). The nitrogenase is irreversibly damaged in the presence of oxygen, therefore, the ability of diazotrophs to fix nitrogen under aerobic conditions has also been intensively studied (Fay, 1992; Gallon, 1981; Zehr & Capone, 2020). Research has shown that diazotrophs typically

separate oxygenic photosynthesis from nitrogen fixation either spatially (for e.g., *Nodularia* heterocysts; Braun et al., 2018) or temporally (for e.g., nitrogen fixation at night in *Crocospaera*; Mohr et al., 2010). Analyses of the *nifH* gene have shown a propensity for certain cyanobacterial diazotroph groups (*Trichodesmium*, *Cyanothece*, and *Crocospaera*) to occur predominantly within tropical zones (Fernández et al., 2010; Moisaner et al., 2010; Stal, 2009), while NCDs belonging to proteobacteria within Cluster I and III are now known to be particularly widespread but are still less understood (Farnelid et al., 2011; Turk-Kubo et al., 2022).

1.2 *Cand. Atelocyanobacterium thalassa* (UCYN-A) and its Ecotypes

Amongst the cyanobacterial diazotrophs, UCYN-A or *Cand. A. thalassa* is unique for several reasons including: (i) it is more broadly distributed into colder temperate regions (Martínez-Pérez et al., 2016; Shiozaki et al., 2015), (ii) it is the first known haptophyte-diazotroph symbiosis (Thompson et al., 2012), and (iii) it has lost its photosystem II (Tripp et al., 2010; Zehr et al., 2008). UCYN-A is a unicellular cyanobacterial diazotroph—hence the term ‘UCYN’ (e.g., Langlois et al., (2008), and Moisaner et al., (2010)). Although some partially successful attempts to obtain cultured isolates have recently been reported, UCYN-A was eventually lost from isolations leaving behind its host (Suzuki et al., 2021), therefore this diazotroph still remains largely uncultivated in laboratory studies. What is known ecologically about UCYN-A has mainly come from: genomic reconstructions of sorted cells (Tripp et al., 2010), diversity studies via amplicon sequencing to examine ecotypes (e.g., Turk-Kubo et al. (2017)), qPCR surveys for resolving abundances in the ocean (Farnelid et al., 2016), and nanometer scale secondary ion mass spectrometry or ‘nanoSIMS’ studies to determine UCYN-A nitrogen fixation rates and carbon-exchanges with host cells (Harding et al., 2018; Krupke et al., 2015; Martínez-Pérez et al., 2016; Mills et al., 2020). Paralleling other major marine microbes with phylogenetically distinct subgroups that encompass multiple ecotypes (for e.g., *Synechococcus* (Sohm et al., 2016), *Prochlorococcus* (Biller et al., 2015), and SAR11/Pelagibacterales (Kraemer et al., 2019)), eight ecotypes for UCYN-A have now been proposed (called ‘A1–A8’; Henke et al., 2018; Turk-Kubo et al., 2017). It is worth noting that recently the study of UCYN-A ecotypes has evolved in conjunction with changing methodologies for acquiring and processing

sequencing data. Although earlier cloning work showed separate ecotypes (A1 versus A2; e.g., Thompson et al., (2014)), next-generation *nifH* sequencing ushered in the use of oligotyping to describe UCYN-A ecotypes (Turk-Kubo et al., 2017). As an alternative to defining OTUs or ‘Operational Taxonomic Units’ using a 97% identity threshold for clustering similar sequences, oligotyping relies instead on grouping sequencing reads using nucleotide sites with maximum entropy (Eren et al., 2013). The use of oligotyping to study UCYN-A effectively predates the now accepted division of next-generation sequence data for microbial communities into amplicon sequencing variants or ‘ASVs,’ the latter strategy taking into account sequence error rates when processing nucleotide reads (Nearing et al., 2018). In line with recent studies (e.g., Ridame et al. (2022)), this dissertation uses ASVs instead of oligotyping and OTUs to characterize microbial communities. Although there is evidence to support the view that UCYN-A ecotypes have separate environmental preferences (for e.g., coastal A2 versus oligotrophic A1; Turk-Kubo et al., 2017), overlapping occurrences for ecotypes (e.g., Henke et al. (2018)) suggest that more spatiotemporal studies are still needed to fully resolve UCYN-A ecotype definitions.

1.3 The Bedford Basin Time Series Monitoring Program

Oceanographic time series represent a major counterpart to more sporadic expeditions for studying ocean processes and marine life, with examples including the Bermuda Atlantic Time-Series study (BATS) and the Hawaii Ocean Time-series (HOT) [since the 1980s], and even earlier time series such as Station M in the Norwegian Sea [1940s] (Gammelsrød et al., 1992; Karl & Lukas, 1996; Steinberg et al., 2001). The major strength of ocean time series is their ability to resolve events that occur on the scale of days-to-decades depending on the length of the time series; for example, shorter term events can include seasonal cycles for phytoplankton blooms, species successions, and nutrient changes in the water column, while longer decadal observations can help resolve baseline oscillations in the system versus the impacts of anthropogenetic changes (Benway et al., 2019; Ducklow et al., 2008; Steinberg et al., 2001). The Bedford Basin Monitoring Program or ‘BBMP’ (also referred to as Bedford Basin time series) has been sampling at weekly intervals (1992 to present day) in the coastal Northwest Atlantic (NWA; Li & Dickie, 2001). The BBMP is also situated within a region that is covered by biannual spring

and fall sampling along the nearby Scotian Shelf as part of the Atlantic Zone Monitoring Program or ‘AZMP’ (DFO Canada, 2006; Zorz et al., 2019). Since its inception, microbial cell counts retrieved through flow-cytometry, as well as core oceanographic parameters (for e.g., temperature, salinity, and nutrients) have been collected from the Bedford Basin (Li & Dickie, 2001), with the newer addition of environmental DNA (eDNA) sampling since 2014 (Raes et al., 2022). Earlier analyses of cell count data showed a strong relationship between phytoplankton size and season (larger phytoplankton in the spring, and smaller in the fall), hence pointing to the cell types attributed to chlorophyll increases during spring and fall from within the coastal NWA (Li & Dickie, 2001; Li et al., 2006). Although historical microscopy work gives some insight into the phytoplankton species underpinning the above trends (e.g., Conover & Mayzaud (1984)), these patterns have not been investigated for phytoplankton with respect to multi-year molecular sampling. Hence, weekly eDNA collections from the Bedford Basin represent a prime opportunity to continue advancing our understanding of the identity and inter/intraspecific patterns exhibited by members of the NWA phytoplankton community. Others and I have also recently demonstrated that eDNA from the Bedford Basin is a key resource for studying microbial processes, showing that nitrification at the bottom of the water column was delayed due to the dilution of nitrifier groups via water column mixing (Haas et al. 2021). In this dissertation eDNA datasets from the Bedford Basin will be accessed repeatedly in each research chapter to examine microbes present within the coastal NWA.

1.4 Newer Areas of Diazotroph Research: Coastal and High-latitude Regions

Recent oceanographic studies have challenged the conventional views on the biogeographical distribution of marine diazotrophs that described their preferred habitat as warmer photic waters of low dissolved inorganic nitrogen (DIN) (Zehr & Capone, 2020, 2021a). Features that challenge this view include diazotrophs being identified in coastal areas (Tang, Wang, et al., 2019), in colder high-latitude waters (e.g., Shiozaki et al., 2018), and below the photic zone (Benavides et al., 2018). In this regard, UCYN-A has especially typified a breakdown of the accepted principles of diazotrophy, not only does it fix nitrogen when DIN is available (Mills et al., 2020), but it also occurs in the Arctic (Harding et al.,

2018; Shiozaki et al., 2018) and has been attributed to high coastal fixation rates in the NWA (Tang, Wang, et al., 2019). As indicated by Tang, Wang, et al., (2019) and Turk-Kubo et al. (2021), UCYN-A's key importance to high nitrogen fixation within nearshore environments points to a critical need for more spatiotemporal measurements of diazotrophs within coastal regions of the ocean. It is important to note that the Bedford Basin may be a prime area to study diazotrophs given that the spring phytoplankton bloom and subsequent draw-down of inorganic nitrogen that occurs during and following the bloom could logically create a temporal niche for diazotrophs (Fonseca-Batista et al., 2019; Li & Dickie, 2001; Shi & Wallace, 2018).

Regarding the polar realm, reasons for marine diazotrophy being overlooked within the Arctic Ocean include: (i) diazotrophic microbes were once thought to be mainly located in warmer oligotrophic waters where dissolved inorganic nitrogen is chronically low (Shiozaki et al., 2018), (ii) high fixed nitrogen concentrations in the Arctic Ocean were assumed unselective for diazotrophy (Zehr & Capone, 2021a), (iii) higher dissolved oxygen levels in Arctic seawater were considered less favourable for marine diazotrophs given that the nitrogenase enzyme is damaged by oxygen (Fernández-Méndez et al., 2016; Gallon, 1992; Stal, 2017), and (iv) relatively higher temperature preferences were attributed to some marine diazotrophs in the ocean with the well-studied *Trichodesmium* sp. mainly dominating in tropical/subtropical regions (Breitbarth et al., 2007; Zehr & Capone, 2020, 2021a). Although more recent studies have begun to unravel community patterns associated with diazotrophs in the Arctic Ocean (for e.g., Fernández-Méndez et al., (2016) and Shiozaki et al., (2018)), basic spatiotemporal data are still critically needed for diazotrophs within this region (reviewed by von Friesen & Riemann (2020)).

1.5 Overview of Thesis

The overarching objective of this thesis was to advance our understanding of marine microbes with respect to their diversity and community dynamics within the NW Atlantic and Arctic. Due to their relationships to carbon and nitrogen cycles, and hence their intersectionality with respect to influencing macronutrients in the ocean, I chose to focus on the phytoplankton and diazotrophic fractions of the marine microbiome. In Chapter 2, I describe using a molecular approach the phytoplankton community dynamics within the

coastal NW Atlantic. Specifically, Chapter 2 provides an in-depth look at the weekly patterns exhibited by the cyanobacterial and eukaryotic phytoplankton present in the region via the Bedford Basin time series. These data were further explored within Chapter 2 by: (a) establishing the degree of overlap between the phytoplankton present within the Bedford Basin and the nearby Scotian Shelf, and (b) examining any annual variability with respect to the known relationship between temperature and small phytoplankton abundance within the region (Li et al., 2006). With prior knowledge that UCYN-A is generally found at higher detection levels within the coastal NWA (Tang, Wang, et al., 2019), Chapter 3 sought to characterize the weekly dynamics of UCYN-A ecotypes over the course of 3-yrs from within this region. For Chapter 3, I developed an improved qPCR assay that separates the UCYN-A2 ecotype from the -A1 ecotype allowing me to search for any temporal variability that is exhibited by these two UCYN-A ecotypes within the coastal realm of the NWA. Furthermore, in Chapter 3, I studied individual UCYN-A ASVs at very high weekly resolution to describe microevolutionary patterns associated with UCYN-A. Phytoplankton data from Chapter 2 were also used in Chapter 3 to assess whether the UCYN-A/host consortium was present within our study site by using a network analysis of cyanobacterial and chloroplast *16S* rRNA versus *nifH* relative abundance data. In Chapter 4, I move just outside of the coastal NW Atlantic and focus on characterizing diazotrophs that occurred within the Canadian Arctic Gateway (CAG; namely, Labrador Sea, Baffin Bay, and the Canadian Arctic Archipelago) during a 2015 Canadian GEOTRACES research expedition (Anderson et al., 2014; Lehmann et al., 2019). For Chapter 4, I assessed whether there was any broad biogeographic separation evident for diazotroph communities across the CAG, and I also describe the ASVs that represent the dominant diazotrophs captured during the expedition. Chapter 4 also uses qPCR assays to enumerate *nifH* gene copy numbers belonging to UCYN-A, hence assessing whether UCYN-A is more broadly distributed into the eastern Canadian Arctic Ocean. As an additional component, I further determined if dominant diazotrophs observed in Chapter 4 could be recovered outside of the CAG (via comparison to the Bedford Basin and to other published *nifH* datasets (Delmont et al., 2021; Shiozaki et al., 2018; Turk-Kubo et al., 2022)). Finally, in Chapter 5, I provide an overview of the major thesis conclusions and

provide examples of how the findings presented herein can inform future studies on phytoplankton and diazotrophs within the NW Atlantic and Arctic sectors.

CHAPTER 2

Highly-resolved Interannual Phytoplankton Community Dynamics of the Coastal Northwest Atlantic

This chapter was previously published in the journal *ISME Communications* (Robicheau et al., 2022) and has been reproduced here with permission (Appendix A).

2.1 Abstract

Microbial observatories can track phytoplankton at frequencies that resolve monthly, seasonal, and multiyear trends in environmental change from short-lived events. Using 4-years of weekly flow cytometry along with chloroplast and cyanobacterial *16S* rRNA sequence data from a time-series station in the coastal Northwest Atlantic (Bedford Basin, Nova Scotia, Canada), we analyzed temporal observations for globally-relevant genera (e.g., *Bolidomonas*, *Teleaulax*, *Minidiscus*, *Chaetoceros*, *Synechococcus*, and *Phaeocystis*) in an oceanic region that has been recognized as a likely hotspot for phytoplankton diversity. Contemporaneous Scotian Shelf data also established that the major phytoplankton within the Bedford Basin were important in the Scotian Shelf during spring and fall, therefore pointing to their broader significance within the coastal Northwest Atlantic (NWA). Temporal trends revealed a subset of indicator taxa along with their DNA signatures (e.g., *Eutreptiella* and *Synechococcus*), whose distribution patterns make them essential for timely detection of environmentally-driven shifts in the NWA. High-resolution sampling was key to identifying important community shifts towards smaller phytoplankton under anomalous environmental conditions, while further providing a detailed molecular view of community compositions underpinning general phytoplankton succession within the coastal NWA. Our study demonstrates the importance of accessible coastal time-series sites where high-frequency DNA sampling allows for the detection of shifting baselines in phytoplankton communities.

2.2 Introduction

Marine phytoplankton contribute ~40% of global carbon fixation and their impact for higher trophic levels, biological carbon uptake, and hence climate, is well recognized (Bonachela et al., 2016; Boyce et al., 2010; Falkowski, 1994). In the Northwest Atlantic (NWA) primary production is characterized by spring and fall phytoplankton blooms (Longhurst, 1995), and sampling at coastal time-series stations has demonstrated a pattern of few larger phytoplankton species (often diatoms) dominating the spring bloom with progressive shifts to higher cell density of smaller phytoplankton species as temperature increases throughout the summer months (Li et al., 2006). However, based on early winter and spring latitudinal transects in the central NWA, Bolaños et al. (2020) recently challenged the broadly accepted view that larger diatoms dominate the spring bloom, proposing instead that small phytoplankton species are important members of spring blooms within the North Atlantic.

Long-term microbial observatories are vital for tracking marine microbes (Buttigieg et al., 2018) and are an important counterpart to the remote sensing of phytoplankton (Hirata et al., 2008). The Bedford Basin Monitoring Program (BBMP), located in Halifax Harbour, Nova Scotia, Canada (Li & Dickie, 2001), represents one of >70 microbial observatories that exist globally (Buttigieg et al., 2018; see (Harris, 2010; Hunter-Cevera et al., 2016; Karl & Lukas, 1996; Steinberg et al., 2001) for other examples). Bedford Basin (71m deep, 8km long) is connected to the Scotian Shelf (Shi & Wallace, 2018) and displays characteristic nutrient and phytoplankton annual cycling for the temperate NWA, including annual spring and fall blooms separated by strong summer stratification (Li & Dickie, 2001; Li et al., 2006). As a fjord with a long narrow entrance (Crawford et al., 2022), Bedford Basin experiences limited freshwater input (Kerrigan et al., 2017; Shi & Wallace, 2018) with an approximately three-month flushing time and a net outward flow for surface waters (Kerrigan et al., 2017; Shan & Sheng, 2012). On a global scale, the NWA coastal waters near the BBMP, in connection to the Gulf Stream, have also been predicted as a hotspot for phytoplankton richness (Barton et al., 2010; Clayton et al., 2013; Dutkiewicz et al., 2020), and in addition, this region is of special interest because higher latitudes/temperate waters are likely to display a higher turnover of species due to high monthly variability in environmental conditions (Righetti et al., 2019).

The phytoplankton community of the Bedford Basin has been documented using flow cytometry at a basic level since the 1990s (Cullen et al., 2007; DFO Canada, 2006; Li & Dickie, 2001; Li, 1998), however, molecular work has been limited and has focused mainly on seasonal non-photosynthetic bacteria, as well as a subset of metaproteomes (El-Swais et al., 2015; Georges et al., 2014). Previous studies reported increased diatom cell density in spring phytoplankton blooms in the fjord (Conover, 1975; Lehman, 1981; Li et al., 2006); while flow cytometry has shown that *Synechococcus* growth coincides with increases in chlorophyll *a* (chl *a*) concentrations in the late summer and fall (Li et al., 2006).

Here we present four years of weekly sampling in the Bedford Basin that characterizes the phytoplankton communities using chloroplast and cyanobacterial *16S* rRNA gene metabarcodes paired with phytoplankton cell concentrations obtained by flow cytometry. Using these data, we compare the latest cell concentrations and phytoplankton community characteristics to historical records and report on novel phytoplankton diversity trends observed within this coastal system. Indicator species—organisms associated with a specific set of environmental conditions (Siddig et al., 2016)—were also identified and may be important in tracking environmental changes in the NWA in general. Using a comparable metabarcoding dataset from a transect across the Scotian Shelf towards the Gulf Stream, we also show that >80% of the major phytoplankton identified in the Bedford Basin time-series were also present in phytoplankton communities of the Scotian Shelf during spring and fall, and that there was a general overlap in the dominant phytoplankton present at both the Scotian Shelf and inshore basin. Finally, we use our high-resolution multi-parameter sampling to highlight atypical phytoplankton community shifts that occurred in relation to anomalous nutrient and temperature events.

2.3 Materials and Methods

2.3.1 Sampling, Oceanographic Data, and Flow Cytometry

Water samples were collected using Niskin bottles from 1, 5, 10, and 60m depths weekly from Bedford Basin (BB; 44.6936 LAT, -63.6403 LON [or 44° 41' 37" N, 63° 38' 25" W]; Halifax, Nova Scotia, Canada), and transported in dark bottles kept in a cooler to a laboratory at Dalhousie University (NS, Canada) and processed immediately upon arrival.

For flow cytometry, 2.5mL of seawater per depth was prefiltered using 35µm cell strainers and autofluorescent cell counts were recorded on a CSampler-equipped BD Accuri™ C6 Flow Cytometer (BD Biosciences, USA) with optical filters for Chlorophyll [>670nm] and Phycoerythrin [585/40nm] detection. The flow cytometry approach implemented targeted ~1–35µm cells, and cell counts were corrected using blanks (0.2µm-filtered seawater) measured concurrently each week (see Supplemental Methods 2.S1 herein for cytometry gate details; raw cell counts can be accessed through the Supplemental Data file S1 provided in Robicheau et al. (2022)).

DNA samples were filtered using acid-washed tubing and bottles and a peristaltic pump. Seawater (500mL) was prefiltered using mesh (at 160µm in 2014–2015, and at 330µm in 2016–2017) and then filtered onto 0.2µm polycarbonate Isopore filters (Millipore, Ireland). Samples from four depths were processed simultaneously, and individual filters were flash frozen in cryovials and stored at -80°C until processing. Samples for DNA were also collected from the Scotian Shelf along the Halifax Line (HL) transect as part of the annual spring and fall Atlantic Zone Monitoring Program (AZMP; cruise codes: HUD2014004, HUD2014030, HUD2016003, HUD2016027, COR2017001, and EN2017606). Cells for AZMP DNA were collected by sequential filtration of water through 3µm and 0.2µm polycarbonate membrane filters using either a vacuum pump (2014) or a peristaltic pump (2016–2017) with coarse prefilters of 160µm (2014) and 330µm (2016–2017) (see Zorz et al. (2019) for full details).

Temperature, chl *a*, and nutrient data for the BB were provided by the Bedford Institute of Oceanography (BIO) (<http://www.bio.gc.ca/science/monitoring-monitorage/bbmp-pobb/bbmp-pobb-en.php>). Temperature and oceanographic data for the AZMP are also available by request from BIO. Overall, flow cytometry and molecular data covered Jan 2014–Dec 2018 and Jan 2014–Dec 2017, respectively.

2.3.2 DNA Extraction and Sequencing

DNA was extracted using a DNeasy Plant Mini kit & protocol (Qiagen, Germany) using the enhanced lysis procedure described by Zorz et al. (2019), and then checked for amount/purity on a NanoDrop 2000c (Thermo Scientific, USA). Illumina MiSeq 300 bp paired-end sequencing of the *16S* ribosomal RNA (rRNA) gene was subsequently carried

out at the Integrated Microbiome Resource at Dalhousie University as in Zorz et al. (2019) and using an established microbiome amplicon sequencing workflow (Comeau et al., 2017). Dual-indexed Illumina fusion primers were used to target variable regions for bacterial *16S rRNA* V6-V8 (primers B969F & BA1406R; Comeau et al., 2011) and for universal *16S rRNA* V4-V5 (primers 515FB & 926R; Parada et al., 2016; Walters et al., 2016). V6-V8 was used for AZMP samples because a partial dataset for this marker was already available for 2014 and 2016 (reported in Zorz et al. (2019) and Willis et al. (2019)). V6-V8 sequences for 2017 have not been published elsewhere.

Amplicon sequence variants (ASVs) were determined using QIIME 2 version 2019.7 (Bolyen et al., 2019) as implemented in the Microbiome Helper pipeline (Comeau et al., 2017). Final taxonomies were derived from a PhytoREF-trained classifier (Decelle et al., 2015) after an initial taxonomic assignment via a full length *16S*-trained SILVA-based classifier (Quast et al., 2013; see Supplemental Methods 2.S2 for further details on ASV selection and characterization). BB chloroplast and cyanobacterial ASVs were rarefied to 200 reads (see Supplementary Fig. 2.7 for frequency distributions of reads per sample and Supplementary Fig. 2.8 for rarefaction curves). Excluding samples with zero reads, at a rarefaction threshold of 200 reads there was a sample loss of ~13.5% (for V4-V5) and ~33% (for V6-V8) for surface samples (1–5m), and furthermore, molecular datasets were skewed towards having a smaller number of final reads per sample (Supplementary Fig. 2.7). To avoid the further exclusion of samples, we did not increase the rarefaction threshold beyond 200 reads. Unless specified, rarefied data were used in BB sample comparisons. AZMP ASV data were not rarefied to preserve all reads and thereby enable identification of any BB-dominant ASVs that were present but rare on the Scotian Shelf; reads for both datasets were converted to percent relative abundance scores prior to any statistical analysis or data visualization. Relative abundance = (number of reads per ASV in a sample ÷ total chloroplast and cyanobacterial reads in said sample) × 100. ASV tables and accompanying reference sequences are available as Supplementary Data files S2–S4 in Robicneau et al. (2022); furthermore, reference sequences for all dominant ASVs were deposited in GenBank (NCBI Resource Coordinators, 2018) under accession codes MZ541860–MZ541862, MZ542324–MZ542326, MZ542548–MZ542554, and MZ571675–MZ571759. Raw sequencing data are also available as Sequence Read

Archives listed under NCBI BioProjects PRJNA785606, PRJNA785859, and PRJNA785872 (NCBI Resource Coordinators, 2018). The number of sequencing reads lost during the sequence processing pipeline can be viewed in Appendix B Supplementary Data A1–A3.

2.3.3 Data Analyses

Statistical analyses, data visualizations, and maps were generated in R version 4.0.0 (R Core Team, 2021) via RStudio version 1.2.5042 (RStudio Team, 2020) using various R packages (listed in Supplemental Methods 2.S3).

In some instances, we focused on a subset of ASVs to facilitate an in-depth assessment of species-level diversity, in such cases we report on the top twenty ASVs found in BB samples per year and representing >80% of the dataset. “Cyanobacteria” and “Phylum” (for other non-cyanobacterial ASVs) was chosen as the highest taxonomic identifiers for plotting large-scale patterns in the BB and AZMP Halifax Line (HL). For clarity, we subdivided the Ochrophyta into Bacillariophyta (i.e., diatoms), Bolidophyceae, Silicoflagellates (i.e., the Dictyochophyceae), and Pelagophyceae. The Bacillariophyta comprised diatom ASVs identified to either a specific taxonomic class (for e.g., Bacillariophyceae, Coscinodiscophyceae, etc.) or simply to the phylum Bacillariophyta. The top twenty ASVs for the AZMP were also selected but on a per sample basis given the biannual nature of this dataset. The list of top AZMP ASVs was also limited to those that reached $\geq 20\%$ relative abundance in at least one sample.

PhytoREF-specified taxonomic assignments (Decelle et al., 2015) for the top ASVs were further refined manually with online BLAST (Altschul et al., 1990; Johnson et al., 2008) using the NCBI nucleotide (*nr/nt*) collection (NCBI Resource Coordinators, 2018). Matches closest to 100% coverage and 100% pair-wise identity (PI) were retained as the final taxonomic identification; ambiguity was resolved following priority for matches to complete genomes > complete genes > partial chloroplast *16S* rRNA (*cp16S* rRNA) gene or partial cyanobacterial *16S* rRNA gene.

Indicator species tests were run using a multi-level pattern analysis via the *multipatt* function in the *indicspecies* R package (De Cáceres & Legendre, 2009) using the point biserial correlation coefficient function “*r.g*” therein with 9,999 permutations.

The *MUSCLE* algorithm (Edgar, 2004) was used to build sequence alignments to calculate in *MEGA* (Kumar et al., 2016) the number of pair-wise nucleotide differences between dominant ASVs with identical or near-identical taxonomies (typically down to the species level). For NMDS plots, Bray-Curtis dissimilarity scores were calculated for Hellinger standardized sample data and then NMDS was run on these data using the *vegan* package (Oksanen et al., 2019). Environmental vectors were fit onto ordinations using the *envfit* function in *vegan* (Oksanen et al., 2019) using 999 permutations. NMDS species scores along with environmental vectors were visualized using *ggplot2* (Wickham, 2016).

Maximum Likelihood trees that assessed the putative placement of dominant *Synechococcus* ASVs into known ecotypes for this genus (Ahlgren & Rocap, 2012; Sohm et al., 2016) were built using *MEGA* (Kumar et al., 2016). The distribution of dominant cyanobacterial and Euglenozoa ASVs were also compared to publicly available *Tara* Oceans datasets (Logares et al., 2014; Sunagawa et al., 2015). See Supplemental Methods 2.S4 for a more detailed explanation of how our ASVs were compared to *Tara* Oceans *mi*TAGS (Logares et al., 2014).

Network analysis comparing V4-V5 and V6-V8 ASVs was carried out using *CoNet* (Faust & Raes, 2016) between identical sample sets, and the resulting network was visualized in *Cytoscape* (Cline et al., 2007; parameter settings are given in Supplemental Methods 2.S5).

Vegan (Oksanen et al., 2019) was used for calculating the Bray-Curtis similarities (by subtracting dissimilarity values from 1) to assess the degree of periodicity in community similarities across the 4-year time-series (see Fuhrman et al. (2015) and Cram et al. (2015) for further information on this approach).

2.4 Results

2.4.1 Temperature and Chlorophyll *a*

The mean temperature in BB surface water from Jan 2014–Dec 2018 ranged from ~0°C (winter months) to 18°C (late summer) [mean minimum for surface depths (1–5m) = -0.261°C ± 0.15 SD; mean maximum for surface depths (1–5m) = 18.73°C ± 1.38 SD] (Fig. 2.1a). The mean surface chl *a* peaked during Mar–May in the spring and Sept–Nov in the fall (Fig. 2.1a). The largest chl *a* maximum (46.5 mg/m³) was in fall 2016 (Fig. 2.1a).

Annual increases in chl *a* during fall and spring blooms were similar between the two seasons, except for 2016, when the fall increase was markedly larger [Fall 2016 = 46.5 mg/m³ at 5m versus Spring 2016 = 21 mg/m³ at 1m] (Fig. 2.1a).

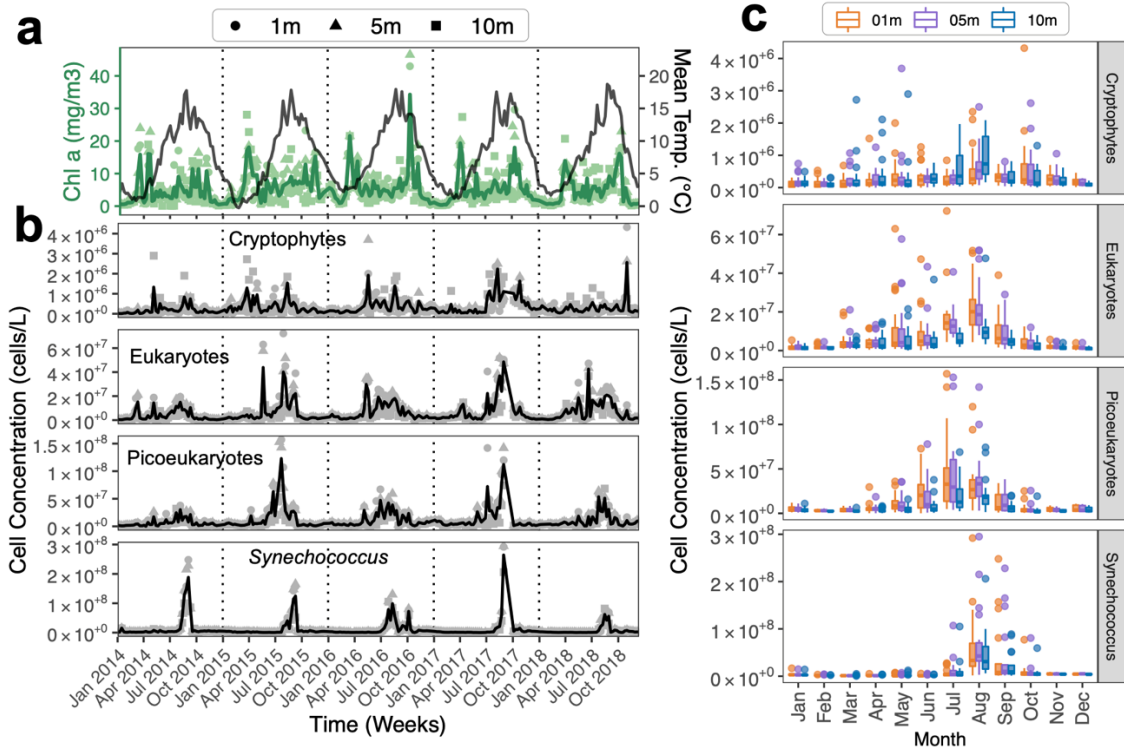


Figure 2.1. Phytoplankton counts for 5-yr (1–10m) Bedford Basin (Halifax, NS) time-series. (a) mean chlorophyll *a* (chl *a*) and mean temperature, (b) mean flow cytometry cell counts for Cryptophytes, Eukaryotes, Picoeukaryotes, and *Synechococcus*. Raw data points are shown as shapes, mean is shown as black lines. (c) Monthly cell count distributions by water depth and all five years combined. The flow cytometry size range is approximately 1–35 μ m. See Supplemental Methods S1 for gate descriptions. The maximum average cell density for each group was: *Synechococcus* = 2.64×10^8 cells/L [mainly bloomed late August and prior to fall chl *a* peaks], Picoeukaryotes = 1.23×10^8 cells/L [increased mainly during summer], Eukaryotes = 4.85×10^7 cells/L [mainly increased starting late spring, then occurred throughout the summer with highest counts in August], and Cryptophytes = 2.56×10^6 cells/L [bloomed mainly during summer/August and values were higher at 10m]. Eukaryotes and Picoeukaryotes also reached higher cell counts in shallower depths (1–5m). Highest mean chl *a* reached during the five years = $34.25 \text{ mg/m}^3 \pm 18.23 \text{ SD}$.

2.4.2 Broader Taxonomic Groups and Flow Cytometry

ASVs belonged to eight broader taxonomic groups (Fig. 2.2a), which generally overlapped between the two genetic markers used. The most notable exception was the Bolidophyceae (mainly observed for V4-V5 and at lower percentages for V6-V8; Fig.

2.2a). We also assessed the similarity for taxonomic assignments between V4-V5 and V6-V8 using a network analysis; for ~64% of top ASVs that co-occurred with another top ASV

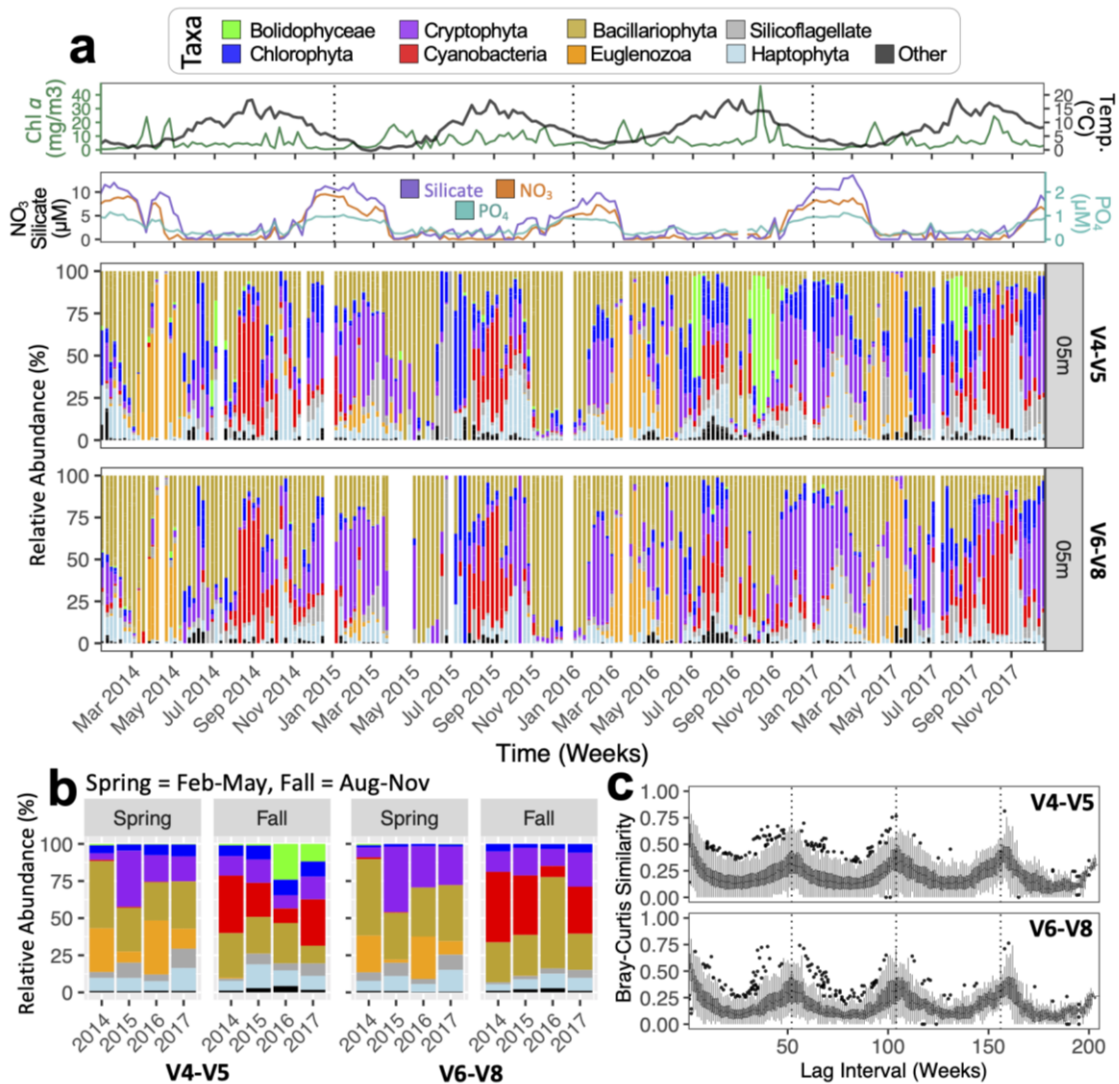


Figure 2.2. Weekly dynamics and seasonal patterns for major phytoplankton groups observed over the 4-yr timeseries in Bedford Basin (Halifax, NS). Sequences are chloroplast *16S* rRNA plus cyanobacterial *16S* rRNA. **(a)** weekly relative abundances for all ASVs at 5m based on their assignment to a major taxonomic group; values shown relative to nutrients, chl *a* and temperature [see Supplementary Fig. 2.10 for 1 & 10m depths]. Unrarefied data (shown in figure) had identical trends to rarefied data, yet unrarefied data retained more samples [compare Supplementary Figs. 2.10 and 2.11]. White columns represent missing or unsuccessfully sequenced samples, or those that only had bacterial *16S* rRNA reads; for our 1–60m Bedford Basin datasets 9 (V4-V5) and 23 (V6-V8) samples only yielded bacterial *16S* rRNA data, and 2 (V4-V5) and 5 (V6-V8) samples did not yield any final sequence reads post processing (Appendix B Supplementary Data A1–A3). **(b)** Summary bar-plots comparing total relative abundance of broader groupings per season and year [going from Feb-01 to May-01 and Aug-01 to

Nov-01], data shown are rarefied to 200 reads and includes all ASVs and 1–10m depths. (c) Distributions of Bray-Curtis similarities between 5m samples plotted against the number of weeks separating samples [relative abundances used from rarefied data; values were also Hellinger transformed (Oksanen et al., 2019)].

in a one-to-one relationship (between opposite markers) the two variable regions led to the same species name (Supplementary Fig. 2.9 and Supplemental Results 2.S1). Note the ‘top 20’ (referred to as ‘top’) BB ASVs are those exhibiting the highest annual relative abundances [82% and 80% of all BB V4-V5 and V6-V8, respectively (for chloroplast and cyanobacterial *16S* rRNA reads)]. Molecular interannual comparisons at 5m for all ASVs (Fig. 2.2) indicates that: (i) Bolidophyceae displayed higher relative abundances in 2016 & 2017; (ii) Haptophyta displayed higher relative abundances primarily in winter/preceding the spring chl *a* maxima and sometimes near the fall chl *a* maximum (e.g. 2015)—the winter period also showed low chl *a*, as well as low Eukaryote and Picoeukaryote cells via flow cytometry (gates that would include haptophytes; Figs. 2.1 and 2.2a), suggesting that colonial haptophytes too large to be captured by flow-cytometry (e.g. *Phaeocystis*; Schoemann et al., 2005) may have been present; (iii) Chlorophyta, Bacillariophyta, and Silicoflagellates displayed less consistent trends in relative abundances corresponding to seasonal phytoplankton blooms; (iv) Cryptophytes were consistently present during months with higher nutrient concentrations; and (v) cyanobacteria and Euglenozoa usually dominated the phytoplankton community during/near the fall and spring bloom periods, respectively. For top BB ASVs *Synechococcus* was the only genus within the cyanobacteria. Unrarefied relative abundances between V4-V5 and V6-V8 were strikingly similar apart from the Bolidophyceae (Fig. 2.2a). Corresponding data for 1m and 10m BB depths also showed similar results (Supplementary Fig. 2.10) and rarefied data versus nonrarefied data showed near identical trends (Supplementary Figs. 2.10 and 2.11).

Several groups of phytoplankton (e.g., Bacillariophyta) did not display peak relative abundances consistently corresponding to seasonal blooms, but rather displayed higher relative abundances intermittently throughout the time-series (Fig. 2.2a). The relative abundances for each major taxonomic group for all ASVs from 1–10m depths with respect to spring and fall months are shown in Fig. 2.2b. While the dominant taxonomic group in the spring months tended to differ annually, cyanobacteria were typically the

dominant group in the fall (Fig. 2.2b), suggesting that the phytoplankton groups dominating in the spring in BB may be less predictable than in the fall. Flow-cytometry counts further showed that the BB phytoplankton community displayed an increase in several microalgal groups after the spring bloom and into the summer months, with *Synechococcus* increasing prior to late fall and winter months (Fig. 2.1b). Specific flow cytometry estimates of cell densities for Cryptophytes, Eukaryotes, Picoeukaryotes, are also shown (Figs. 2.1b and 2.1c). While we did not observe recurrent increases in cell density for any particular group prior to the increases in chl *a* corresponding to spring blooms (Fig. 2.1), we did observe peaks in Eukaryotes (2014) and several peaks in Cryptophytes (2015) that correspond to chl *a* increases in those respective years (Fig. 2.1). Flow-cytometry results were remarkably consistent with historical records (see <http://www.bio-iob.gc.ca/science/monitoring-monitorage/bbmb-pobb/bbmb-pobb-en.php>; last accessed 7-May-2020) with maximum values for *Synechococcus*, Picoeukaryotes, and Cryptophytes, as well as temperature and chl *a* maxima ranges being comparable to previous studies (Cullen et al., 2007; Li et al., 1998; Li et al., 2006; Li & Dickie, 2001). As previously observed (Li & Dickie, 2001), the spring bloom may have been populated by larger phytoplankton cells that were not recorded by flow cytometry. As our DNA sequencing indicates, the phytoplankton community dynamics during the spring period appear to be better resolved using molecular approaches (Fig. 2.2).

Although weekly trends suggested that finer scale changes in phytoplankton community compositions even for very broad taxonomic groups could occur quite rapidly from week-to-week (Fig. 2.2a), an analysis of the Bray-Curtis similarities between samples relative to the number of weeks between samples (i.e. the lag interval) verified an underlying year-to-year community stability/cyclicality for the phytoplankton of Bedford Basin (observe similarity increases at ~52-week intervals; Fig. 2.2c; Fuhrman et al., 2015). This trend was also consistent for all three surface depths, for both *16S* rRNA markers, and regardless of whether sequence data was rarefied or not (Supplementary Figs. 2.12 and 2.13).

The Euglenozoa reoccurred near/during the spring bloom and showed an increase in relative abundance that generally paralleled the increase in chl *a* in 2014, 2016, and 2017 (they also remained present for several weeks after spring chl *a* peaks; Fig. 2.2a).

Interestingly, the lower relative abundances of Euglenozoa in 2015 was balanced by a larger relative abundance and cell counts for Cryptophyta (Figs. 2.2a and 2.1b, respectively). Given that *Synechococcus* and Euglenozoa were the only broader taxonomic groups that displayed clear temporal profiles linked to fall and spring bloom periods, respectively, we designated these two groups as indicator species of seasonal BB phytoplankton blooms (Supplementary Table 2.2). Given their regional importance, we searched for the top ASVs belonging to these two groups within $_{mi}$ TAGs from *Tara* Oceans data (Bork et al., 2015; Logares et al., 2014; Sunagawa et al., 2015). BB *Synechococcus* (V4-V5: $n = 2$) and Euglenozoa (V4-V5: $n = 1$) have matches to the *Tara 16S* rRNA $_{mi}$ TAGs (Supplementary Fig. 2.14); these trends suggest that the two top BB *Synechococcus* ASVs are found globally (i.e. they are likely cosmopolitan), while the top Euglenozoa ASV was only detected at two *Tara* sites, the North Atlantic (39.2305, -70.0377) and the Southeast Atlantic shelf waters (-32.2401, 17.7103) suggesting a potential preference for coastal (or near coastal) regions (Supplementary Fig. 2.14; Logares et al., 2014; Sunagawa et al., 2015).

2.4.3 Individual ASV Profiles

Weekly relative abundance profiles for the top twenty BB ASVs are shown (Fig. 2.3a; for V4-V5 $n = 37$ ASVs and V6-V8 $n = 39$ ASVs [n can be >20 due to yearly differences]). Comparison to *PhytoREF* (Decelle et al., 2015) provided broader taxonomy (i.e., typically class-level); however, 59% of V4-V5 and 74% of V6-V8 ASVs required further comparison to the NCBI *nr/nt* database to obtain a species or genus assignment (NCBI Resource Coordinators, 2018). For consistency, the classification for each top ASV was confirmed via BLAST in NCBI (Johnson et al., 2008).

Some ASVs occurred in select years and sporadically throughout the year (e.g. ASV4 *Micromonas pusilla* and ASV7 *Bathycoccus prasinus*; Fig. 2.3a), while others occurred at high relative abundances and displayed a consistent annual pattern (e.g. ASV11 *Teleaulax amphioxeia*, *Synechococcus* sp. [ASVs 12, 14, 43, & 45], *Chaetoceros* sp. + *Chaetoceros simplex* [bASVs 17, 18, 55, 56 & ASV54], *Eutreptiella pomquetensis* [ASV32/ASV69–71], and *Phaeocystis* spp. [ASV36/75]) (Fig. 2.3a). Seasonally early (Spring + Summer) and late (Summer + Fall) *Minidiscus trioculatus* ASVs were also

observed (Fig. 2.3a; Supplementary Table 2.2). We consider that indicator taxa (Supplementary Table 2.2) that returned frequently and regularly during the 4-yr likely provide the optimal metric for overall phytoplankton community change in our temperate coastal waters.

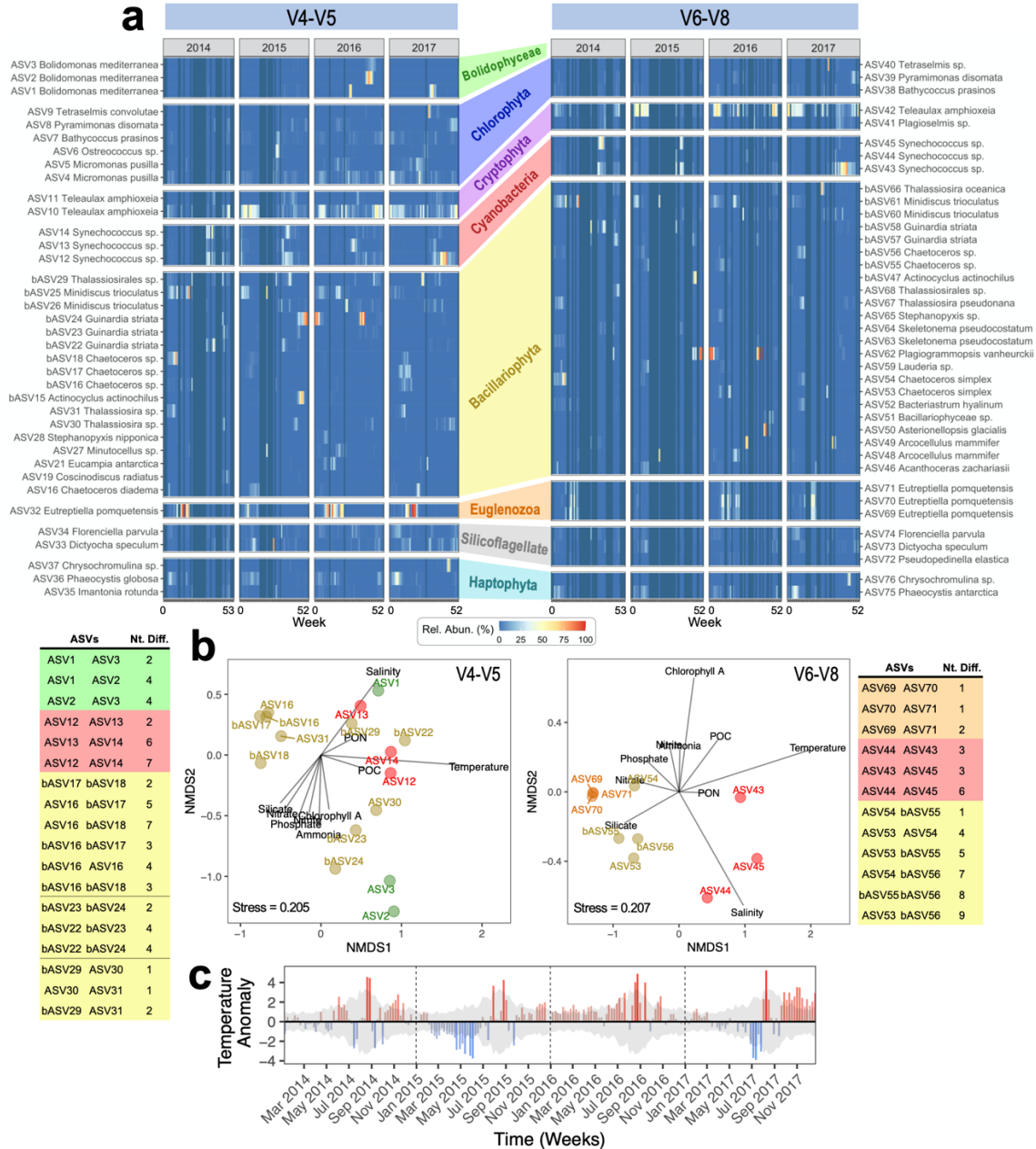


Figure 2.3. Individual temporal relative abundances profiles for Top 20 phytoplankton *16S* rRNA ASVs (chloroplast + cyanobacterial *16S* rRNA) in the Bedford Basin time-series from 2014–2017. Data is rarefied. **(a)** Weekly relative abundances (Rel. Abun.) at 5m water depth for V4-V5 and V6-V8. Rel. Abun. (%) = (number of reads per ASV in a sample / total chloroplast and cyanobacterial reads in said sample) × 100. **(b)** Plot of species scores after 5m samples subjected to NMDS analysis. Only groups of ASVs with >3 identical

genus and/or species names are shown along with the number of nucleotide differences (Nt. Diff.) between these ASVs. Environmental variables are overlaid as lines. (c) Weekly temperature anomalies at 5m [values are $\pm^{\circ}\text{C}$ difference relative to the 1992–2017 weekly mean temperature]. Shading in anomaly plots show standard deviations of the weekly 1992–2017 means. Anomalies for nutrients and salinity also given (Supplementary Fig. 2.16). Also see the indicator species test of Supplementary Table 2.2 for significant seasonal associations of top ASVs.

Multiple top ASVs were often identified within the same species or genus. NMDS analyses were used to assess the preferred environmental conditions for closely related ASVs to determine whether sequences with only 1–9bp differences (Nt. Diff.; Fig. 2.3b) represented true biological variants or perhaps resulted from sequencing errors (Callahan et al., 2017). The distribution similarity of closely related ASVs over the 4-yr period at 5m was plotted using NMDS ordination overlaid with environmental variables to determine whether the ASVs co-varied temporally (Fig. 2.3b). This analysis revealed that small V4-V5 and V6-V8 chloroplast and cyanobacterial *16S* rRNA nucleotide differences could represent true interspecific differences with ecological relevance (see Supplemental Results 2.S2 for more specific trends).

Changes in temperature appeared to influence the temporal patterns of several ASVs designated as key indicator species (Fig. 2.3). In particular, *Synechococcus* ASV12 & ASV43 had higher relative abundance values during 2017 when there were consistently high temperature anomalies during the late fall/early winter (Fig. 2.3c). The opposite was seen for *E. pomquetensis*, whereby its temporal patterns were consistent with the laboratory-determined narrow growth range of 0–10 $^{\circ}\text{C}$ for this species (Supplementary Fig. 2.15; McLachlan et al., 1994). *E. pomquetensis* (ASV32 & 69–71) had especially low relative abundances during 2015, which was the only year with 5m temperature down to 0 $^{\circ}\text{C}$ (Figs. 2.1a, 2.2a, and Supplementary Fig. 2.15). We propose that the sub-zero temperatures reached at 5m during spring 2015 led to the observed shift from Euglenozoa to Cryptophytes (Fig. 2.2a). The patterns above also lend their support to the use of *Synechococcus* ASV43 and *E. pomquetensis* as indicator species in BB, given that changes in the relative abundance of these two phytoplankton groups paralleled temperature anomalies (warmer and cooler conditions, respectively; Fig. 2.3).

For the ASVs identified to at least genus-level we provide summary stats and reference accessions for *BLAST* matches (Supplementary Table 2.3; Johnson et al., 2008).

Nearly all the dominant phytoplankton identified to species-level were marine (according to www.algaebase.org, last accessed 17-May-2021), except for *Acanthoceras zachariasii* (freshwater; Edlund & Stoermer, 1993) and *Pseudopedinella elastica* (brackish; Tomas, 1993). Hence, freshwater input appeared to have little influence on shaping the dominant phytoplankton observed.

2.4.4 Comparisons to the Scotian Shelf

Using spring and fall AZMP data, we found that the vast majority [85% or 33/39] of the top V6-V8 BB ASVs were present on the Scotian Shelf (Supplementary Fig. 2.17). The six top BB ASVs in the fjord that were not found on the Scotian Shelf during our study were ASV40 *Tetraselmis* sp., ASV69-ASV71 *Eutreptiella pomquetensis*, ASV73 *Dictyocha speculum*, and bASV60 *Minidiscus trioculatus*. At the shelf there were a total of 36 top AZMP ASVs: 66% of these were also recovered in the fjord, 39% were dominant in both regions, 28% were dominant on the shelf but still found in the fjord, and 33% were dominant at the shelf but absent in the fjord (Fig. 2.4).

Two BB ASVs [ASV42 + ASV75] had especially high relative abundances across samples from nearly all shelf stations in the spring (Fig. 2.5 and Supplementary Fig 2.17), confirming the importance of *T. amphioxeia* and *Phaeocystis* sp. to phytoplankton communities inside the fjord, as well as beyond the shelf break during spring periods (Figs. 2.3 and 2.4, and Supplementary Table 2.2). Similarly, ASV45 *Synechococcus* sp. was observed in the small size fraction at consistently high relative abundances across nearly all AZMP HL stations in the fall season (Fig. 2.4). Maximum-Likelihood trees indicate that the BB *Synechococcus* likely belong to *Synechococcus* clades I (ASV13 and ASV44) and IV (ASV14; Supplementary Fig. 2.18; Ahlgren & Rocap, 2012; Sohm et al., 2016).

Analysis of top AZMP ASVs associated with the shelf (Fig. 2.4) revealed that: (i) cASV20 *Braarudosphaera bigelowii* and cASV13 *Fragilariopsis* sp. were consistently dominant in the >3 μ m fraction at all HL stations in the fall and spring, respectively (these variants were also identified in the fjord but outside its list of top ASVs), (ii) *Pelagomonas* sequence variants (cASV22 & 21) are important throughout the shelf, and (iii) dominant AZMP ASVs, which are also absent from the fjord, were mainly found during the fall season beyond the shelf break (for e.g., *Trichodesmium* and *Prochlorococcus*). This last

class of ASVs also included an off-shelf fall-associated *E. pomquetensis* variant (cASV18) that was not observed in the fjord (Fig. 2.4c); this ASV likely represents a warm-water associated ecotype given that it occurred in $20.2 \pm 1.7^\circ\text{C}$ waters at station HL08.

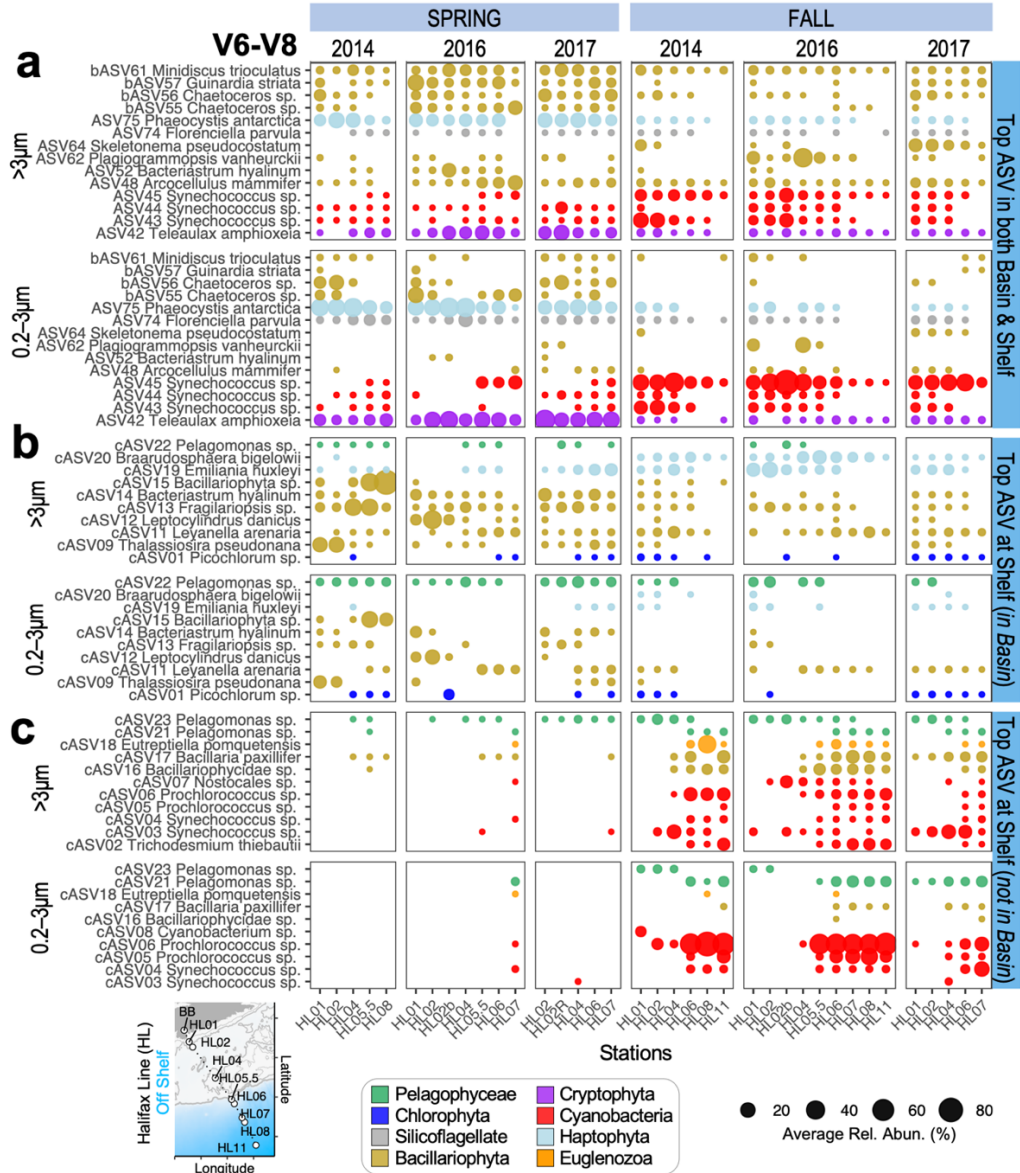


Figure 2.4. Comparison of top Scotian Shelf phytoplankton ASVs versus ASVs that were observed in the Bedford Basin (BB) time-series. Stations were sampled from the Halifax Line (HL) transect during Spring and Fall Atlantic Zone Monitoring Program (AZMP) expeditions. **(a)** ASVs that are within the top twenty list for both Basin and Shelf. **(b)** Top AZMP ASVs that were also detected in the Bedford Basin; however, they are not among the top basin-specific ASVs. **(c)** Top AZMP ASVs that were not detected in the Bedford Basin. Note that AZMP ASVs are limited to the top twenty most relatively abundant ASVs per sample, and to those ASVs that reached $\geq 20\%$ in at least one sample. No AZMP cruise data available for 2015. Individual relative abundances or ‘Rel. Abun.’ as % = (number of

reads per ASV in a sample / total chloroplast and cyanobacterial reads in said sample) × 100; an average was then taken per station. Data are from V6-V8 metabarcoding between 1–80m water depths (surface & photic zone; Zorz et al., 2019); data unrarefied. Summary stats for BLAST results given in Supplementary Table 2.4. Stations HL02b and HL02R were second samplings of HL02 (during return trips to shore). Note stations are organized left-to-right by increasing distance from the Bedford Basin.

From a broader perspective, AZMP data revealed that numerous phytoplankton species observed by the BBMP are also found on the Scotian Shelf and often during both spring and fall seasons (Fig. 2.4 and Supplementary Fig. 2.17). Hence, these taxa are key phytoplankton beyond the Bedford Basin and into the more expansive coastal NWA shelf waters.

2.4.5 Small Phytoplankton and their link to Atypical Temperature Conditions

Based on the example of a temperature related community shift from *Eutreptiella* to cryptophytes (Figs. 2.1 and 2.2) and historical observations linking temperature and cell density (Li et al., 2006), we examined the relationship between temperature and <3µm cells. Temperature versus cell densities from 2014–2018 demonstrated that in 2016 abnormally high temperatures throughout winter, summer, and early fall months (Figs. 2.3c and 2.5a) coincided with higher densities for <3µm cells throughout the summer and fall (Fig. 2.5c and Supplementary Fig. 2.19). Nitrate levels were low during the winter mixing of 2015/2016, amounting to a period of consistently low nitrate anomalies (Fig. 2.5b and Supplementary Fig. 2.16). This same year also displayed an increase in the correlation between temperature and density for <3µm cells (Fig. 2.5d). Note that another study has already proposed weaker mixing in the Bedford Basin during Winter 2015/2016 (see Haas et al. (2021)). In addition to flow cytometry counts, cp16S rRNA and cyanobacterial 16S rRNA data for 2016 further indicated the presence of smaller phytoplankton (Table 2.1). Lastly, an examination of shelf data also hinted that the unique dynamics of 2016 may not have been restricted to the fjord alone; for instance, ASV62 *Plagiogrammopsis vanheurckii* occurred at every station along the HL transect in Fall 2016 (Fig. 2.4 and Supplementary Fig. 2.17).

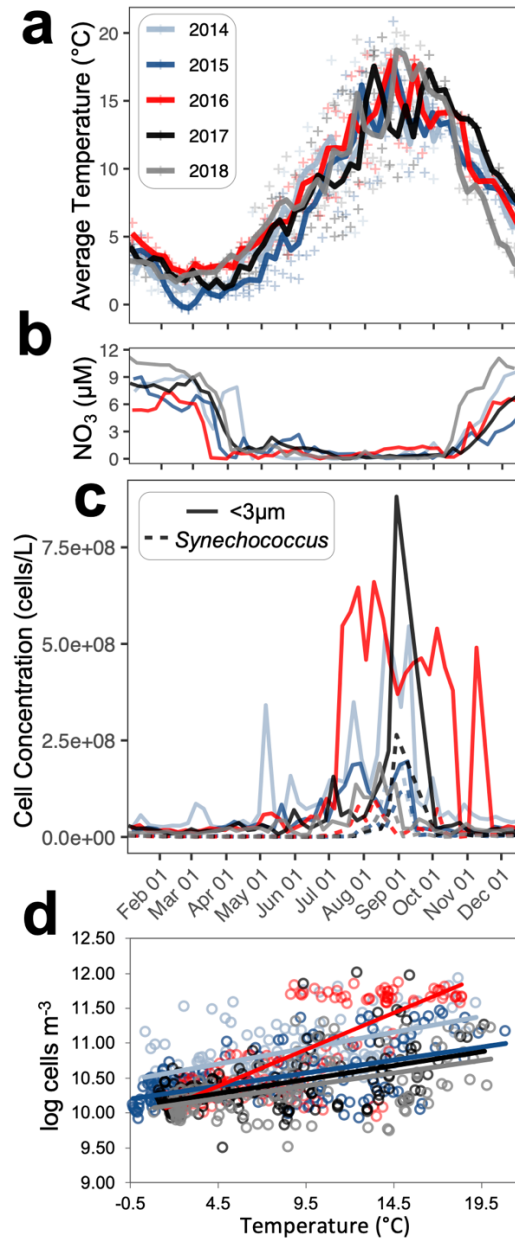


Figure 2.5. Warmer temperatures and trends for smaller phytoplankton during 2016. (a) average surface temperatures by year [raw data shown as plus symbols], (b) surface nitrate levels by year suggestive of weaker winter mixing in 2016, (c) average surface <3µm cell concentrations (plotted by Month/Day), and (d) their yearly correlations with temperature (for the <3µm group). Surface depths were 1m, 5m, and 10m samples. Trendlines in (c) assumed a linear relationship, and the equations: 2014: $y = 0.0454x + 10.495$; 2015: $y = 0.0369x + 10.222$; 2016: $y = 0.1051x + 9.9062$; 2017: $y = 0.0401x + 10.099$; and 2018: $y = 0.0366x + 10.04$. For comparison, the 1993–2005 cell density vs. temperature relationship for BB phytoplankton (not just <3µm) is reported as $y = 0.097x + 9.47$ by Li et al. (2006). Individual years are shown in Supplemental Fig. 2.20.

Table 2.1. Small taxa associated with trends in 2016 (values are for 1–10m; unrarefied data). Ave. Rel. Abun. stands for average relative abundance.

ASV	Ave. Rel. Abun. (% \pm SD) [Period]	Approximate Cell Sizes (μm^3)
ASV49 <i>Arcocellulus mammifer</i>	62 \pm 20 [Early Jul]	\sim 11 ^{a,*}
ASV44 <i>Synechococcus</i> sp.	21 \pm 18 [Late Jul] 16 \pm 11 [Early Aug]	\sim 0.38–1.15 ^b
ASV42 <i>T. amphioxeia</i>	21 \pm 10 [Late Jul] 10 \pm 8 [Early Aug]	\sim 109 ^c
ASV62 <i>Plagiogrammopsis vanheurckii</i>	61 \pm 8 [Last Week of Aug] 56 \pm 35 [Sept]	\sim 372 ^d
ASV50 <i>Asterionellopsis glacialis</i>	68 \pm 18 [Oct]	\sim 1,492 ^c

a Using dimensions in (Percopo et al., 2011) and formula in (Sun & Liu, 2003);

b see (Agawin et al., 2004; Bertilsson et al., 2003);

c median of volumes reported at nordicmicroalgae.org (last accessed 14-Apr-2021);

d see (Harrison et al., 2015; Tomas, 1997);

* Network (Supplemental Fig. 2.9) suggested ASV49 *A. mammifer* may be ASV1 *B. mediterranea*, however, even if the latter is correct, cells would still be small given *B. mediterranea*'s diameter of 1–1.7 μm (Guillou et al., 1999). Furthermore, *A. mammifer* is reported as \lesssim 3 μm in its shorter axes (Percopo et al., 2011).

2.5 Discussion

2.5.1 *Synechococcus* and *Eutreptiella* are Important Phytoplankton *16S* rRNA Gene Signatures in the Bedford Basin

Photosynthetic organelles along with their plastid genomes (Falkowski et al., 2004) have garnered attention as targets for characterizing phytoplankton communities (Choi et al., 2017; Decelle et al., 2015; Needham et al., 2017). Overall, our chloroplast plus cyanobacterial *16S* rRNA metabarcoding approach revealed a coherence in the multiyear phytoplankton community composition, i.e., some similarity to previous microscopy records, while also providing higher resolution for species- (and in some cases) ecotype-level taxonomy for smaller phytoplankton (Supplemental Discussions 2.S1 & 2.S2 provide further context and information on this topic).

Synechococcus cyanobacterial patterns were the most consistent feature present between flow cytometry and chloroplast/cyanobacterial *16S* rRNA data during our study. Although some *Synechococcus* ecological patterns for the Bedford Basin were previously

known (Li et al., 2006), we expanded this knowledge by demonstrating that the dominant BB *Synechococcus* ecotypes belong to clades I and IV. Others have recently demonstrated a shift from *Synechococcus* near the fjord to *Prochlorococcus* off-shelf (Zorz et al., 2019).

In contrast to *Synechococcus*, our molecular data demonstrated Euglenozoa, particularly *E. pomquetensis*, had a distinct spring/early summer occurrence pattern at our study site. Although the basic biology of *E. pomquetensis* is known (McLachlan et al., 1994), to the best of our knowledge, the striking patterns of this species' association with the spring period has not previously been described. This is possibly due to sampling design and/or issues with morphological identification, which may in some cases obfuscate this trend (Supplemental Discussion 2.S3). Given that Euglenozoa was observed at high relative abundance even after chl *a* maxima (e.g. 2016; Fig. 2.2) and that 16S rRNA data is compositional, one can suggest that Euglenozoa may retain a presence after the bloom via a mixotrophic lifestyle that would include grazing in addition to photosynthesis, thereby explaining why Euglenozoa remained present weeks after peaks in spring chl *a* when inorganic nutrient availability decreases. Confirmed mixotrophy in another *Eutreptiella* species (Yoo et al., 2018) suggests that *E. pomquetensis* could have a mixotrophic lifestyle.

Overall, we propose ASV43 *Synechococcus* (Clade I) and ASV32/69-71 *E. pomquetensis* as indicator species having special importance for detecting environmentally driven change in the fjord for fall and spring seasons, respectively, as the former was linked to warmer summer temperatures, while the latter appeared adversely affected by colder winter/spring temperatures (Fig. 2.3; McLachlan et al., 1994). Changes in these particular ASVs might possibly be a preamble to trends expected from ongoing climate change. Ultimately the molecular identification of seasonally specific indicator species provides a framework and baseline from which to assess (through DNA sampling) any future effects that extreme environmental change may have on the typically reoccurring primary producers of the coastal NW Atlantic.

2.5.2 Phytoplankton tracked by the Bedford Basin Time-series are Globally Relevant

Although there has been interest in the bacterial communities along the Scotian Shelf (Willis et al., 2019; Zorz et al., 2019), the phytoplankton communities remain

relatively unexplored via molecular analysis except for recent reports by Zorz et al. (2019) and Willis et al. (2019), as well as previous *I8S* sequencing by Dasilva et al. (2014).

Recent literature suggests that many of the phytoplankton we identified within the Bedford Basin, and nearby at the Scotian Shelf, are globally significant. *Synechococcus* (of clades I & IV in the fjord), were previously identified as important in the NWA and were especially dominant in the subpolar region during winter (Bolaños et al., 2020); members of these clades have also been detected in colder waters with elevated nutrients in the North Pacific Ocean (Sohm et al., 2016). Several phytoplankton inhabiting the fjord were similarly reported on the Scotian Shelf by others (Dasilva et al., 2014) and in the off-shore NWA (Bolaños et al., 2020). Those in common include: *Bathycoccus*, *Micromonas*, *Chaetoceros*, *Phaeocystis*, *Teleaulax*, and *Thalassiosira*, which were all identified as key phytoplankton found in the off-shore NWA with both *Micromonas* and *Bathycoccus* especially relevant to the NWA subtropic zone (Bolaños et al., 2020, 2021). Members of these two taxa are also widespread (Monier et al., 2016), and *Micromonas* and *P. antarctica* both occur in the Southern Ocean (Irion et al., 2021). In addition to genera mentioned above, *Ostreococcus*, *Dictyocha*, *Florenciella*, *Fragilariopsis*, *Minidiscus*, and *Braarudosphaera bigelowii* were also observed on the Scotian Shelf (Dasilva et al., 2014). *Florenciella parvula* has been reported as an important component of the dictyochophyte fraction in mesotrophic surface samples from the eastern North Pacific (Choi et al., 2020), while the shelf-wide distributed *Minidiscus trioculatus*, was recently proposed as a small diatom of great importance due to its widespread occurrence and likely contribution to carbon export (Leblanc et al., 2018). *Fragilariopsis* and *B. bigelowii* are also of special note; the former is common in polar environments (Lundholm & Hasle, 2010), while the latter is associated with the symbiotic nitrogen-fixing unicellular cyanobacteria, *Candidatus Atelocyanobacterium thalassa* (Martínez-Pérez et al., 2016; Zehr et al., 2016). Furthermore, the *Pelagomonas* sp. identified herein matched known (and geographically wide-ranging) wildtype *Pelagomonas* (Worden et al., 2012). Interestingly, *T. amphioxeia* was recently reported to have two morphotypes with differing ploidy and winter/spring versus summer distributions (Altenburger et al., 2020). BBMP data may reflect this novel ecology; however, additional data/confirmation is needed. Finally, *Arcocellulus mammifer*, one of the small Bacillariophyta we recognised as important in 2016, responded positively

to increased temperature in field incubation experiments conducted at the San Pedro Ocean Time-series (from NE Pacific; Kling et al., 2020). Like *A. mammifer*, several small taxa (Table 2.1) were linked to the unique dynamics of 2016; future experimental work should investigate these taxa from the perspective of community-based growth responses under warming *in situ* temperatures.

Cell counts and community composition-based analyses for $<3\mu\text{m}$ cells versus temperature demonstrated that in the coastal NWA small phytoplankton are particularly responsive to increases in *in situ* temperatures during the same year. This trend was likely not a sole consequence of temperature, but rather temperature's effect on the water column during earlier parts of 2016, and therefore nutrient availability. This is supported by Haas et al. (2021) who concluded weaker BB winter mixing during 2015/16, as well as by observations herein of negative nitrate anomalies (Supplementary Fig. 2.16). While the generally accepted view of higher small phytoplankton cell densities in late summer and fall for the BB/NWA was upheld during the span of our study (Li et al., 2006), the unique pattern of 2016 (effectively a higher temperature and lower nutrient scenario) highlights the variable nature of the cell density versus temperature relationship within the coastal NWA (particularly under warmer conditions; Li et al., 2006). This variability ultimately points to high-frequency sampling and datasets as being essential for identifying changing trends within coastal environments of the NWA.

Collectively, our results established a baseline for seasonal variation in phytoplankton *16S* rRNA gene diversity over a period of several years in the Bedford Basin, a coastal fjord that has been sampled for several decades (Li & Dickie, 2001). Many of the phytoplankton tracked by the BBMP time-series are globally relevant, hence our observations provide highly resolved data for some of the most important oceanic primary producers. In essence, the phytoplankton community of the BBMP is a continuum of the phytoplankton in the NWA and shows important weekly trends for species that are dominant in the NWA, including, *Arcocellulus*, *Bolidomonas*, *Teleaulax*, *Minidiscus*, *Chaetoceros*, *Phaeocystis*, as well as multiple ecotypes of *Synechococcus*. As such, the results presented herein contribute to our known understanding of the biota within the NWA—a region of global significance for marine productivity, sustainable marine

fisheries (Chassot et al., 2010; Gentry et al., 2017), and predicted global phytoplankton richness (Barton et al., 2010).

2.5.3 Additional Insights gained from High-frequency DNA Sampling

Ocean time-series continue to be a key resource for the study of ocean microbiomes and their community dynamics (Benway et al., 2019; Fuhrman et al., 2015); for instance, our ability to track phytoplankton with weekly frequency provided additional insights into: i) the existence of potentially novel/unknown ecotypes (for e.g., a warm-water associated *Eutreptiella* ASV at the edge of the Scotian Shelf), ii) the extent to which weekly community transitions can occur for dominant phytoplankton within the region (for e.g., rapid transitions in phyla-level community compositions were often evident even within monthly timeframes), and iii) the general utility of both V4-V5 and V6-V8 within the *cp16S* rRNA gene for tracking phytoplankton (as corroborated by our network analysis between the two markers). As these points suggest, the ecological knowledge that can be gained from high-resolution molecular sampling of the ocean microbiome using a stationary time-series is multifaceted and can range from the characterization of basic species distributions to the collection of *in situ* observational data that can reveal ocean variability on multiple time scales (Benway et al., 2019).

Another major advantage of weekly DNA sampling is that our final time-series provided insight into the phytoplankton successional trends (that is the restructuring of community compositions across time; Rigosi et al., 2010) that occurred over four complete annual cycles. Classically, phytoplankton succession in the NWA has been defined by reoccurring yearly cycles of pico-phytoplankton (prior to spring bloom), diatoms (during the spring bloom), followed by other phytoplankton (e.g., coccolithophores post spring-bloom), with further succession towards small phytoplankton during the fall bloom (Bolaños et al., 2020; Daniels et al., 2015; Li et al., 2006). When we examined Bray-Curtis similarities between samples, we observed a clear cyclical relationship for phytoplankton communities, indicating that there is indeed an underlying reoccurring cycle with respect to the *in situ* phytoplankton diversity of the Bedford Basin (Fig. 2.2c; Fuhrman et al., 2015). Another feature of this pattern, however, was that peaks in Bray-Curtis similarities typically only approached 0.5 (as opposed to 1.0 for 100% identical communities; Fuhrman

et al., 2015); therefore, despite the phytoplankton community displaying cyclicality over multi-year scales, the patterns were not entirely deterministic (i.e., non-random; Masuda et al., 2017) in the sense that the community composition was not exactly the same each year (Fuhrman et al., 2015). This point, along with the various individual temporal profiles we presented for dominant phytoplankton ASVs within the Bedford Basin (Fig. 2.3) demonstrates the rather complex nature of the *in situ* phytoplankton diversity that exists within the classical succession patterns mentioned earlier. For example, obvious and repeatable patterns in *16S* rRNA gene relative abundances were observed for *Phaeocystis* (appearing in early spring), *Eutreptiella* (appearing in spring and early summer), and *Synechococcus* (appearing in fall). While in contrast, the successional patterns for diatoms were more semi-predictable in that dominant diatom ASVs appeared year-after-year, but their temporal reoccurrence patterns were less clearly defined. Given that detailed *in situ* molecular observations can be lacking for key transitional periods such as during spring blooms (Daniels et al., 2015) and during winter periods (Bolaños et al., 2020), in future, it will be worthwhile to use our molecular time-series of phytoplankton diversity to inform additional studies regarding phytoplankton succession within the coastal NWA.

2.6 Conclusions

In this study we presented a detailed time-series of phytoplankton occurring at a coastal site within the Northwest Atlantic, along with coincident phytoplankton observations at a nearby transect along the Scotian Shelf. Together these datasets: i) provided a comprehensive and broad survey of the dominant phytoplankton within the coastal NWA across all four seasons, ii) revealed the identity of key indicator species and novel ecotypes within the region, iii) pointed towards the contribution of smaller cells under anomalous nutrient and temperature conditions, and iv) validated the use of two *16S* rRNA gene variable regions (V4-V5 & V6-V8) for phytoplankton tracking and for investigating intraspecific (e.g. ecotype) patterns in the context of time-series molecular data. Collectively, our analyses amount to a more detailed molecular picture of both the cumulative and seasonal phytoplankton biodiversity within the coastal NWA. Our ability to link regionally-specific taxa to the phytoplankton present at the Scotian Shelf and within the global ocean via literature further supports the view that the Bedford Basin Monitoring

Program (BBMP) is especially well suited for identifying seasonal and interannual trends for a variety of key temperate phytoplankton. Therefore, with its ease of access and long-term high-resolution set of observations, the BBMP is an initiative that lends itself as a perfect backdrop for further manipulative field experiments and process studies to assess the future effects of climate change on primary productivity in the NWA.

2.7 Acknowledgements

The Bedford Institute of Oceanography is acknowledged for assistance in sample collection, allowing participation in AZMP expeditions, and for providing physiochemical data. Also thanked are I. Luddington, J-M. Ratten, C. Mackie, C. Willis, J. Zorz, and S. Rose for contributing to the weekly time-series, J. Cherian for assisting in literature compilation, and M. Segura Guzman for help with anomaly calculations. Funding was provided by the National Sciences and Engineering Research Council of Canada (NSERC) in the form of Discovery Grants to J. LaRoche and E. Bertrand, as well as a Canada Graduate Scholarship–Doctoral Award to B. M. Robicheau. Additional funding was provided by the Ocean Frontier Institute to J. LaRoche and E. Bertrand, a Canada Foundation for Innovation Grant to J. LaRoche, and from a Nova Scotia Graduate Scholarship and Killam Predoctoral Award to B. Robicheau.

2.8 Author contributions

B. Robicheau and J. Tolman performed data collection and measurements. B. Robicheau performed all graphical and statistical analyses and wrote the entire first draft of the manuscript. J. Tolman performed data processing and organized sampling logistics. J. LaRoche provided overall experimental design and worked with B. Robicheau and E. Bertrand to generate data interpretations.

2.9 Supplemental Information for Chapter 2

2.9.1 Supplemental Text for Chapter 2

2.9.1.1 Supplemental Methods 2.S1: Flow-cytometry Gates

Flow cytometry gates for autofluorescence and size (forward scatter or FSC) were as in Li & Dickie (2001), and then also refined using both in-house axenic phytoplankton cultures with known autofluorescence/FSC profiles and Fluoresbrite microspheres (Polysciences, USA). An event exclusion threshold was set at <800 for Chlorophyll. Gates were as follows: Cryptophytes [high Chl, high PE, high FSC], Eukaryotes [high FSC, high Chl], Picoeukaryotes [low Chl, low PE, & lower FSC], and *Synechococcus* [low Chl, high PE, and low FSC] (Supplementary Fig. 2.6). Less than 3 μ m, 3–10 μ m, and >10 μ m flow cytometry gates were approximated using fluorescent microspheres at 0.7, 3, 6, and 10 μ m (Polysciences, USA; Spherotech, USA; BD, USA; Polysciences, USA, respectively).

2.9.1.2 Supplemental Methods 2.S2: Amplicon Sequence Variants (ASVs)

Raw read qualities were checked with *FASTQC* (Andrews, 2010) and *MultiQC* (Ewels et al., 2016). Primer sequences were removed using *cutadapt* (Martin, 2011). Paired-end reads were stitched together using *PEAR* (Zhang et al., 2014), then imported into *QIIME2* (Bolyen et al., 2019). Low-quality reads were removed, and the remaining reads denoised into ASVs using *deblur* (Amir et al., 2017). As part of the user-specified *deblur* trim settings V6-V8 reads and V4-V5 reads were trimmed to 380bp and 350bp, respectively. ASVs with sequence read frequencies less than $[0.001 \times \text{mean sample depth}]$ were attributed to sequencer bleed-through and removed. Initial taxonomy was assigned via a Naïve-Bayes approach using the *classify-sklearn* command (Bolyen et al., 2019; Pedregosa et al., 2011) and a full-length 16S rRNA trained classifier (based on *SILVA* database v132; Quast et al., 2013). Phytoplankton taxonomy was further refined with a *PhytoREF*-trained classifier (Decelle et al., 2015) to reclassify any ASVs initially designated as either ‘cyanobacteria’ or ‘chloroplast’ by the *SILVA* database (Quast et al., 2013).

2.9.1.3 Supplemental Methods 2.S3: R Packages and Maps

The following packages were used for data visualizations/statistical analyses unless otherwise specifically indicated in the main text: *ggplot2* (Wickham, 2016), *ggrepel* (Slowikowski, 2020), *reshape2* (Wickham, 2007), *scales* (Wickham & Seidel, 2020), and *cowplot* (Wilke, 2019). Any ‘packages’ mentioned within the text are also R packages (R Core Team, 2021). Note that the *stat_summary* function in *ggplot2* (Wickham, 2016) is used in various instances to plot the mean as a line across the time-series.

The data in Supplementary Figure 2.17, were organized using hierarchical clustering via the *hclust* function in the *stats* package (R Core Team, 2021) and through the *ggdendro* (de Vries & Ripley, 2016) and the *scale* function in R (R Core Team, 2021).

Maps were generated using the *mapdata* (Brownrigg, 2018), *ggrepel* (Slowikowski, 2020), and *ggnewscale* (Campitelli, 2020) packages in R. Bathymetry data were retrieved from the online *ERDDAP* (Simons, 2019) server via its *griddap* protocol/data access form and the *ETOPO1* topography dataset [ID = *etopo180*] (Amante & Eakins, 2009); for this dataset the institution/creator is listed as the National Oceanic and Atmospheric Administration (NOAA) & the National Geophysical Data Center (NGDC).

2.9.1.4 Supplemental Methods 2.S4: Comparison to Tara Oceans *miTAGs* for *Synechococcus* and Euglenozoa

For the distribution of cyanobacterial and Euglenozoa ASVs across the *TARA* Oceans samples, we retrieved metagenomic Illumina tag (*miTAG*) count data from the companion website to Sunagawa et al. (2015)[see: ocean-microbiome.embl.de/companion.html; Last Accessed 15-Jun-2020], which also contains reference data for Logares et al. (2014). Our ASV sequences were locally aligned to the SILVA *16S* rRNA sequences from the *TARA* website above using the online *BLAST* server (Altschul et al., 1990; Johnson et al., 2008), thus facilitating the retrieval of count data for *miTAGs* that shared similarity to our own ASVs (Logares et al., 2014). Matches were limited to ASVs with 100% pair-wise identity (PI) and 100% coverage.

2.9.1.5 Supplemental Methods 2.S5: V4-V5 versus V6-V8 Network Analysis

For the network, rarified data (all ASVs & all depths) were converted to relative abundance, and a minimum occurrence of 20 reads across all samples was required (removes rarer ASVs). Using the ensemble approach (Faust et al., 2012) the following methods were used to calculate V4-V5 ASV versus V6-V8 ASV network associations: Pearson and Spearman correlations, Mutual Information similarities, as well as Bray Curtis and Kullback-Leibler dissimilarities, with 1000 top and bottom edges retained for each (Faust et al., 2012). Initial p -values were calculated through permutation [100 iterations] and randomization via row shuffling, in addition, the renormalize parameter was selected and final p -values also incorporated bootstrapping [100 iterations] (Faust et al., 2012). Brown's method was selected for merging p -values with a Benjamini-Hochberg multiple test correction setting (Brown, 1975; Faust et al., 2012; Faust & Raes, 2016). The significance threshold was set at $\alpha = 0.05$. These settings reflect parameters provided by https://psbweb05.psb.ugent.be/conet/microbialnetworks/conet_new.php (last accessed 5-Aug-2020).

2.9.1.6 Supplemental Results 2.S1: Network Analysis Comparing V4-V5 versus V6-V8

The sparsity of full-length *cp16S* rRNA reference sequences for uncultivated phytoplankton often precluded linking ASVs from the two variable regions to a common reference sequence. We conducted a network analysis to determine if any of the top ASVs could be directly correlated between the two *16S* rRNA markers using their temporal abundance profiles (Supplementary Fig. 2.9). Fourteen one-to-one ASV co-occurrences between the two *16S* rRNA markers were found. The majority (64.3%) of these had the same species name. Although symbiotic and/or mutualistic interactions cannot be excluded when interpreting the network, ASVs with significant network correlations yet different identification could have originated from the same taxon. One example is bASV24 *Guinardia striata* versus ASV62 *Plagiogrammopsis vanheurckii* (Supplementary Fig. 2.9 and Fig. 2.3).

2.9.1.7 Supplemental Results 2.S2: Trends for NMDS of Closely Related ASVs

NMDS plots indicate that for 2014–2017: (i) there were likely two ecotypes/strains of *Bolidomonas mediterranea* observed in the Bedford Basin [ASV1 associated with high salinity]; (ii) two ASVs of *Synechococcus* for each variable region that correlated with higher temperature [V4-V5 = ASV12 & 14; V6-V8 = ASV43] while another ASV was correlated with higher salinity [V4-V5 = ASV13; V6-V8 = ASV44 & 45]; (iii) the *Eutreptiella pomquetensis* ASVs observed in the V6-V8 dataset showed matching temporal distributions questioning the biological/ecological significance of the three ASVs; (iv) the detection of different ecotypes for several diatom species is suggested based on their preference for different environmental conditions, for example two ecotypes/strains of *Chaetoceros simplex* were detected by V6-V8 (with ASV53 somewhat associated with high salinity).

2.9.1.8 Supplemental Discussion 2.S1: Discordance Between Historical and New Molecular Data (This Study)

Amongst the predominant diatoms in our dataset were species from *Chaetoceros*, *Coscinodiscus*, *Eucampia*, *Pyramimonas*, *Skeletonema*, and *Thalassiosira*. Each of these genera have been reported in the Bedford Basin previously, with *Chaetoceros*, *Skeletonema*, and *Thalassiosira* consistently reported by multiple researchers (Conover & Mayzaud, 1984; Conover, 1975; Kepkay et al., 1997; Kranck & Milligan, 1988; Lehman, 1981; Mayzaud & Taguchi, 1979) While none of the species-level identifications from these studies matched those found through our *cp16S* rRNA analysis, it is possible that this discordance could have arisen from the use of molecular data (herein) versus morphological data (previous research) for species assignment. For example, although we identified *Skeletonema pseudocostatum* as the dominant *Skeletonema* species, *S. costatum* has been most often reported in the Bedford Basin (Conover & Mayzaud, 1984; Lehman, 1981; Mayzaud & Taguchi, 1979). Medlin et al. (1991) propose that these two species can be readily distinguished morphologically, but the ability to do so depends on rather detailed knowledge of diatom morphological traits and potentially the use of electron microscopy (for e.g., see Kooistra et al. (2008) and Sarno et al. (2005)). One should note, however, that there was indeed overlap for some species-level identifications; for instance, the

Silicoflagellate, *Dictyocha speculum*, was the same species reported by Li et al. (1998). Taxonomic discordance may also arise from species under-representation in reference sequence datasets (discussed further below in Discussion 2.S2); although alignment hits were >97% for *Chaetoceros*, *Skeletonema* and *Thalassiosira*, in several cases matches still had <100% PI (see Supplementary Table 2.3).

One should also recognize that the study of Willis et al. (2019) also recovered Bolidophyceae primarily by V4-V5, and that the results we presented in our study for a much larger dataset agrees with this earlier finding.

2.9.1.9 Supplemental Discussion 2.S2: Coherence in the Multiyear Phytoplankton Community Composition

Our taxonomic identifications based on cp*16S* rRNA metabarcoding were generally consistent with previous microscopy records for larger phytoplankton, while providing a much more detailed taxonomic identification of smaller phytoplankton. It was evident that taxonomic discordance could have arisen due to differences in molecular data (herein) versus historical morphological data, for example, genera seemed to match well but species often did not (see Discussion 2.S1); as mentioned above in Discussion 2.S1, another possible source of discordance may have been a lack of necessary reference sequences in GenBank, which would understandably hinder one's ability to arrive at a variant's true taxonomy. For several of the dominant ASVs detected by both variable regions, similar reference accession codes were returned as top BLAST matches; this congruency may be due to the *nr* database (NCBI Resource Coordinators, 2018) lacking sufficient sequences for the true species corresponding to each ASV, or may be a product of characterizing the same dominant microbial population regardless of marker choice (albeit with some differences in sequence conservation, as the pair-wise identity matches were not always 100%). More full length cp*16S* rRNA sequences are needed for type materials if researchers hope to further optimize the investigation of phytoplankton diversity via chloroplast/plastid subsets generated through *16S* rRNA metabarcoding. In future, large metabarcoding datasets of time series samples (such as the one presented herein) will also be useful in identifying and characterising novel cryptic species/strains that may currently lack the sampling frequency needed to discover their underlying ecologies (for instance,

see Choi et al. (2017)). For example, the taxonomic resolution afforded in this study allowed for the identification of closely related ecotypes from the same species that thrive under different environmental conditions (Fig. 2.3b).

In general, there was also a lack of phototropic dinoflagellates observed in our study and this is mostly likely explained by low cell densities for this group. We did observe four Dinophyta ASVs belonging to *Karlodinium veneficum* (for V4-V5 at 0.002% of entire dataset) and *Karenia mikimotoi* (for V4-V5 at 0.024% of dataset, and for V6-V8 at 0.005% of dataset) at very low relative abundances. It is possible that some of the dinoflagellates in the Bedford Basin may have been non-photosynthetic (i.e., lacking chloroplasts; Schnepf & Elbrächter, 1999), although Dasilva et al. (2014) have shown through 18S rRNA cloning that many of the dinoflagellates identified in April and October 2009 along the nearby Scotian Shelf were mixotrophic. DNA pre-filtration would have also theoretically allowed for the capture of dinoflagellates given that Lehman (1981) reported an average BB dinoflagellate cell volume of $54,444 \pm 3,120 \mu\text{m}^3$; for one of the larger dinoflagellates they identify, *Dinophysis norvegica*, its size is upwards of $70\mu\text{m}$ in one dimension (Carpenter et al., 1995). Given that metabolism and cell size cannot explain low dinoflagellate detection, we turn instead to cell densities. Historical microscopy observations show that annual average ratios in this region are on the order of 1055:10:1 for *Synechococcus* : Diatoms : Dinoflagellates (Li et al., 2006). Based on these ratios, there exists a relatively low expected probability of recovering dinoflagellate cpDNA gene sequences compared to diatoms in this fjord, hence, low dinoflagellate concentrations seem the most plausible explanation for dinoflagellate rarity in our molecular data.

2.9.1.10 Supplemental Discussion 2.S3: *Eutreptiella* and Historical Reporting

Reasons for low *Eutreptiella* reporting in the past may include selective exclusion during flow cytometry, water sample treatment, and morphology scoring. Unlike *Synechococcus*, which has an average cell size of $0.9\mu\text{m}$ (Morel et al., 1993), *E. pomquetensis* (previously *Tetreutreptia pomquetensis*) has an average cell length of $\geq 70\mu\text{m}$ (McLachlan et al., 1994) and would be selectively excluded by pre-filtration of flow cytometry samples. McLachlan et al. (1994) also indicate that unless fixed in glutaraldehyde or Lugol's solution, *E. pomquetensis* is very sensitive to increased

temperatures, with exposure to $\geq 10^{\circ}\text{C}$ generally lethal. Thus, one can assume that unless Bedford Basin water samples are rapidly fixed at near *in situ* temperatures, there is a likelihood of biasing microscopy samples against *E. pomquetensis*. Furthermore, microscopic identification could have misclassified *E. pomquetensis* cells or placed them into a broader taxonomic group (e.g., flagellate). Intriguingly, another Euglenozoa, *Euglena* sp., is listed as occurring in the Bedford Basin during the 1990s (Li et al., 1998).

2.9.2 Supplemental Tables for Chapter 2

Table 2.2. Output of indicator species test. 1–10m depths and all weeks used in analysis; data were rarified and converted to percent relative abundance. Significance codes: ‘***’ significant at 0.001, ‘**’ significant at 0.01, ‘*’ significant at 0.05

V4-V5			V6-V8		
Association function: r.g			Association function: r.g		
Significance level (alpha): 0.05			Significance level (alpha): 0.05		
Total number of species: 37			Total number of species: 39		
Selected number of species: 36			Selected number of species: 37		
Number of species associated to 1 group: 22			Number of species associated to 1 group: 24		
Number of species associated to 2 groups: 14			Number of species associated to 2 groups: 13		
Number of species associated to 3 groups: 0			Number of species associated to 3 groups: 0		
List of species associated to each season(s)			List of species associated to each season(s)		
	Stat	p-value		Stat	p-value
Group Fall #sps. 5			Group Fall #sps. 5		
ASV12.Synechococcus.sp.	0.504	1e-04 ***	ASV43.Synechococcus.sp.	0.587	1e-04 ***
ASV11.Teleaulax.amphioxieia	0.443	1e-04 ***	ASV41.Plagioselmis.sp.	0.466	1e-04 ***
ASV14.Synechococcus.sp..CC9902	0.387	1e-04 ***	ASV45.Synechococcus.sp.	0.409	1e-04 ***
ASV2.Bolidomonas.mediterranea	0.256	1e-04 ***	ASV64.Skeletonema.pseudocostatum	0.294	1e-04 ***
bASV22.Guinar dia.striata	0.212	1e-04 ***	ASV50.Asterionellopsid.glacialis	0.227	2e-04 ***
Group Spring #sps. 6			Group Spring #sps. 9		
ASV36.Phaeocystis.globosa	0.412	1e-04 ***	bASV56.Chaetoceros.sp.	0.443	0.0001 ***
bASV17.Chaetoceros.sp.	0.39	1e-04 ***	ASV75.Phaeocystis.antarctica	0.417	0.0001 ***
bASV18.Chaetoceros.sp.	0.316	1e-04 ***	ASV59.Lauderia.sp.	0.349	0.0001 ***
ASV31.Thalassiosira.sp.	0.311	1e-04 ***	ASV54.Chaetoceros.simplex	0.344	0.0001 ***
ASV28.Stephanopyxis.nipponica	0.241	1e-04 ***	ASV67.Thalassiosira.pseudonana	0.316	0.0001 ***
ASV21.Eucampia.antarctica	0.173	3e-04 ***	bASV55.Chaetoceros.sp.	0.316	0.0001 ***
Group Summer #sps. 4			Group Summer #sps. 3		
ASV5.Micromonas.pusilla	0.473	0.0001 ***	ASV52.Bacteriastrium.hyalinum	0.307	0.0001 ***
ASV27.Minutocellus.sp.	0.304	0.0001 ***	ASV65.Stephanopyxis.sp.	0.247	0.0001 ***
ASV6.Ostreococcus.sp.	0.29	0.0001 ***	ASV53.Chaetoceros.simplex	0.166	0.0063 **
ASV9.Tetraselmis.convolutae	0.126	0.0125 *	Group Summer #sps. 3		
Group Winter #sps. 7			Group Winter #sps. 7		
bASV24.Guinar dia.striata	0.311	1e-04 ***	ASV48.Arcocellulus.mammifer	0.367	1e-04 ***
ASV8.Pyramimonas.disomata	0.295	1e-04 ***	bASV66.Thalassiosira.oceanica	0.221	1e-04 ***
ASV35.Imantonia.rotunda	0.261	1e-04 ***	ASV40.Tetraselmis.sp.	0.171	3e-04 ***
bASV15.Actinocyclus.actinochilus	0.252	1e-04 ***	Group Winter #sps. 7		
ASV19.Coscinodiscus.radiatus	0.222	1e-04 ***	ASV38.Bathycoccus.prasinos	0.349	1e-04 ***
bASV23.Guinar dia.striata	0.206	1e-04 ***	ASV39.Pyramimonas.disomata	0.297	1e-04 ***
ASV37.Chrysochromulina.sp.	0.193	1e-04 ***	bASV57.Guinar dia.striata	0.257	1e-04 ***
Group Fall+Summer #sps. 4			Group Fall+Summer #sps. 4		
ASV13.Synechococcus.sp.	0.228	2e-04 ***	ASV68.Thalassiosirales.sp.	0.251	1e-04 ***
bASV26.Minidiscus.triocolatus	0.222	1e-04 ***	bASV47.Actinocyclus.actinochilus	0.234	1e-04 ***
ASV1.Bolidomonas.mediterranea	0.201	3e-04 ***	ASV51.Bacillariophyceae.sp.	0.208	2e-04 ***
bASV29.Thalassiosirales.sp.	0.173	3e-04 ***	ASV76.Chrysochromulina.sp.	0.195	6e-04 ***
Group Fall+Winter #sps. 2			Group Fall+Summer #sps. 4		
ASV30.Thalassiosira.sp.	0.272	1e-04 ***	ASV49.Arcocellulus.mammifer	0.224	0.0002 ***
ASV3.Bolidomonas.mediterranea	0.175	7e-04 ***	bASV60.Minidiscus.triocolatus	0.22	0.0001 ***
Group Spring+Summer #sps. 4			Group Fall+Winter #sps. 1		
ASV32.Eutreptiella.pomquetensis	0.4	0.0001 ***	ASV62.Plagiogrammopsis.vanheurckii	0.31	1e-04 ***
bASV25.Minidiscus.triocolatus	0.256	0.0001 ***	Group Spring+Summer #sps. 6		
bASV16.Chaetoceros.sp.	0.209	0.0002 ***	ASV71.Eutreptiella.pomquetensis	0.379	0.0001 ***
ASV16.Chaetoceros.diadema	0.159	0.0026 **	ASV69.Eutreptiella.pomquetensis	0.371	0.0001 ***
Group Spring+Winter #sps. 2			Group Spring+Summer #sps. 6		
ASV10.Teleaulax.amphioxieia	0.244	1e-04 ***	ASV70.Eutreptiella.pomquetensis	0.349	0.0001 ***
ASV34.Floren ciella.parvula	0.233	1e-04 ***	bASV61.Minidiscus.triocolatus	0.291	0.0001 ***
Group Summer+Winter #sps. 2			Group Spring+Winter #sps. 2		
ASV7.Bathycoccus.prasinos	0.216	2e-04 ***	ASV72.Pseudopedinella.elastica	0.232	0.0001 ***
ASV4.Micromonas.pusilla	0.208	1e-04 ***	ASV46.Acanthoceras.zachariasii	0.167	0.0064 **
			Group Summer+Winter #sps. 2		
			ASV74.Floren ciella.parvula		
			0.269		
			1e-04 ***		
			ASV42.Teleaulax.amphioxieia		
			0.233		
			2e-04 ***		

Table 2.3. List of dominant ASVs that were manually identified to genus/species level. Also provided are reference accessions and *BLAST* pairwise-identities (PI) to the NCBI nucleotide (*nr/nt*) collection (Altschul et al., 1990; Johnson et al., 2008; NCBI Resource Coordinators, 2018). Note that when multiple reference accessions were found to match at equivalent PI and query coverage only a few are given as example. Please see methods section for further details on the approach used for resolving *BLAST* matches. NCBI sequences likely of endobiont origin were excluded during taxonomy assessment (e.g. sequences in (Tsuchiya et al., 2015, 2018)).

Taxon	Accessions (PI % Query Coverage %)
ASV46 <i>Acanthoceras zachariasii</i>	NC_038009.1 (97.11 100)
bASV15 <i>Actinocyclus actinochilus</i>	FJ002163.1* (100 100)
bASV47 <i>Actinocyclus actinochilus</i>	FJ002163.1* (99.21 100)
ASV48 <i>Arcocellulus mammifer</i>	FJ002193.1 (100 100)
ASV49 <i>Arcocellulus mammifer</i>	FJ002193.1 (98.43 100)
ASV50 <i>Asterionellopsis glacialis</i>	FJ002233.1 (98.68 100)
ASV51 Bacillariophyceae sp.	FJ002233.1, AF514850.1 (98.68 100)
ASV52 <i>Bacteriastrum hyalinum</i>	FJ002166.1 (97.89 100)
ASV7 <i>Bathycoccus prasinus</i>	LN735275.2, FO082259.2 (100 100)
ASV38 <i>Bathycoccus prasinus</i>	FN563099.1 (100 100)
ASV1 <i>Bolidomonas mediterranea</i>	LN735367.3, AY702144.1 (98.29 100)
ASV2 <i>Bolidomonas mediterranea</i>	KC509524.1, LN735367.3, AY702144.1 (98.29-98.57 100) [May be <i>Leptocylindrus danicus</i> , however only 1bp more similar than other <i>Bolidomonas</i> and <i>PhytoREF</i> also suggests <i>Bolidomonas</i>]
ASV3 <i>Bolidomonas mediterranea</i>	LN735367.3, AY702144.1 (98.6 100)
ASV16 <i>Chaetoceros diadema</i>	MH011755.1, LN735283.2 (99.14 100)
ASV53 <i>Chaetoceros simplex</i>	KJ958479.1 (99.21 100)
ASV54 <i>Chaetoceros simplex</i>	LC088209.1, KJ958479.1 (99.21-99.74 100)
bASV17 <i>Chaetoceros</i> sp.	MH011755.1, LN735388.3, LN735283.2, AJ319825.1 (99.43 100)
bASV18 <i>Chaetoceros</i> sp.	LN735300.2 (100 100)
bASV16 <i>Chaetoceros</i> sp.	MH011753.1, JN207225.1 (99.43-99.71 99-100)
bASV55 <i>Chaetoceros</i> sp.	LC088209.1, FJ002204.1 (99.47-100 100)
bASV56 <i>Chaetoceros</i> sp.	NC_053621.1, FJ002215.1, FJ159135.1 (98.42 100)
ASV37 <i>Chrysochromulina</i> sp.	LN735342.3, LN735328.3, LN735326.3, AB196967.1 (99.71 100) – Matches several homotypic synonyms for <i>Chrysochromulina</i> genus
ASV76 <i>Chrysochromulina</i> sp.	AB196966.1 (98.42 100)
ASV19 <i>Coscinodiscus radiatus</i>	AJ536462.1 (98.86 100), possibly <i>C. granii</i> (new sequence Apr-2021 MW561225.1 (99.14 100))
ASV33 <i>Dictyocha speculum</i>	NC_043929.1* (100 100)
ASV73 <i>Dictyocha speculum</i>	NC_043929.1* (99.74 100)
ASV21 <i>Eucampia antarctica</i>	FJ002159.1 (99.71 100)
ASV32 <i>Eutreptiella pomquetensis</i>	KY706202.1*, EU750699.1 (100 100)
ASV69 <i>Eutreptiella pomquetensis</i>	KY706202.1* (99.47 100)
ASV70 <i>Eutreptiella pomquetensis</i>	KY706202.1* (99.74 100)
ASV71 <i>Eutreptiella pomquetensis</i>	KY706202.1* (100 100)
ASV34 <i>Florenciella parvula</i>	NC_044407.1*, LN735277.2 (100 100)
ASV74 <i>Florenciella parvula</i>	NC_044407.1* (100 100)
bASV22 <i>Guinardia striata</i>	LN735412.3, NC_037998.1* (99.42-98.85 99-100)
bASV23 <i>Guinardia striata</i>	NC_037998.1* (99.43 100)
bASV24 <i>Guinardia striata</i>	NC_037998.1* (99.43 100)
bASV57 <i>Guinardia striata</i>	NC_037998.1* (98.95 100)
bASV58 <i>Guinardia striata</i>	NC_037998.1* (98.16 100)
ASV35 <i>Imantonia rotunda</i>	LN735489.3, AY702150.1 (100 100) – new species name is <i>Dicrateria rotunda</i>
ASV59 <i>Lauderia</i> sp.	FJ002202.1, AJ536459.1 (100 100)
ASV4 <i>Micromonas pusilla</i>	LN735276.2, MT136879.1, EF051748.1 (100 100)
ASV5 <i>Micromonas pusilla</i>	LN735344.3 (99.14 100)

Taxon	Accessions (PI % Query Coverage %)
bASV26 <i>Minidiscus trioculatus</i>	FJ002231.1* (100 100)
bASV25 <i>Minidiscus trioculatus</i>	FJ002231.1* (99.71 100)
bASV60 <i>Minidiscus trioculatus</i>	FJ002231.1* (99.74 100)
bASV61 <i>Minidiscus trioculatus</i>	FJ002231.1* (100 100)
ASV27 <i>Minutocellus</i> sp.	LN735501.3, FJ002193.1, FJ002199.1 (99.15-99.43 100) – was also similar to <i>Arcocellulus mammifer</i>
ASV6 <i>Ostreococcus</i> sp.	LN735433.3, LN735218.2, AY702141.1 (100 100)
ASV75 <i>Phaeocystis antarctica</i>	JN117275.2 (100 100)
ASV36 <i>Phaeocystis globosa</i>	KC900889.1, MT471334.1 (100 100)
ASV62 <i>Plagiogrammopsis vanheurckii</i>	NC_037998.1, NC_037997.1 (98.42-98.43 100) – also had high similarity to <i>Guinardia striata</i>
ASV41 <i>Plagioselmis</i> sp.	AB164406.1 (98.95 100)
ASV72 <i>Pseudopedinella elastica</i>	NC_044408.1 (99.21 100)
ASV8 <i>Pyramimonas disomata</i>	FN563101.1* (100 100)
ASV39 <i>Pyramimonas disomata</i>	FN563101.1* (100 100)
ASV63 <i>Skeletonema pseudocostatum</i>	MK372941.1 (100 100)
ASV64 <i>Skeletonema pseudocostatum</i>	MK372941.1 (99.74 100)
ASV28 <i>Stephanopyxis nipponica</i>	LN735463.3, AJ536465.1 (100 100) – was also similar to <i>Stephanopyxis turris</i> , however, species name not given in sequence description for LN735463.3
ASV65 <i>Stephanopyxis</i> sp.	FJ002176.1, AJ536465.1 (100 100)
ASV12 <i>Synechococcus</i> sp.	CP047959.1 (100 100)
ASV13 <i>Synechococcus</i> sp.	JX530065.1, CP011941.1 (99.71-100 100)
ASV14 <i>Synechococcus</i> sp.	MH358353.1, MT994359.1 (100 100)
ASV44 <i>Synechococcus</i> sp.	JX529910.1, JX477000.1 (99.47 100)
ASV43 <i>Synechococcus</i> sp.	CP047959.1, CP047942.1 (100 100)
ASV45 <i>Synechococcus</i> sp.	LN847356.1 (100)
ASV10 <i>Teleaulax amphioxeia</i>	KP899713.1* (100 100)
ASV11 <i>Teleaulax amphioxeia</i>	KP899713.1* (99.43 100)
ASV42 <i>Teleaulax amphioxeia</i>	KP899713.1*, EU123322.1 (100 100)
ASV9 <i>Tetraselmis convolutae</i>	LN735267.2 (98.29 100)
ASV40 <i>Tetraselmis</i> sp.	HE610165.1, KU167097.1 (93.63-96.25 98) – matches more closely to <i>Tetraselmis cordiformis</i> , however, alignments for these accessions only have 98% query coverage and <97% PI.
bASV66 <i>Thalassiosira oceanica</i>	GU323224.1 (98.95 100)
ASV67 <i>Thalassiosira pseudonana</i>	EF067921.1 (99.74 100)
ASV30 <i>Thalassiosira</i> sp.	KT956318.1, LN735461.3, LN735460.3 (100 100)
ASV31 <i>Thalassiosira</i> sp.	MH011825.1, KT956312.1 (100 100)
bASV29 <i>Thalassiosirales</i> sp.	MK372941.1, GU323224.1 (100 100)
ASV68 <i>Thalassiosirales</i> sp.	Now possibly <i>Coscinodiscus walesii</i> * (new sequence Apr-2021 MW561224.1 (98.42 100))

Table 2.4. Additional ASVs dominant on the Scotian Shelf that were also manually identified. Information is reported using the same approach described in the caption of Supplementary Table 2.3 (see above).

Taxon	Accessions (PI % Query Coverage %)
cASV1 <i>Picochlorum</i> sp.	MN647759.1 (98.69 100)
cASV2 <i>Trichodesmium thiebautii</i>	AF091321.1 (99.74 100); MT478931.1 (100 100)
cASV3 <i>Synechococcus</i> sp.	JX477000.1 (100 100)
cASV4 <i>Synechococcus</i> sp.	CP047954.1 (100 100); CP047949.1 (100 100)
cASV5 <i>Prochlorococcus</i> sp.	CP007754.1 (100 100)
cASV6 <i>Prochlorococcus</i> sp.	CP018346.1, CP018345.1, CP018344.1 (all 100 100)
cASV7 <i>Nostocales</i> sp.	AB491868.1 (95.2 98); MT488223.1 (93.39 99); AM230674.1 (93.39 99)
cASV8 <i>Cyanobacterium</i> sp.	None

Taxon	Accessions (PI % Query Coverage %)
cASV9 <i>Thalassiosira pseudonana</i>	FJ002218.1, EF067921.1 (both 99.21 100)
cASV11 <i>Leyanella arenaria</i>	FJ002242.1 (99.47 100)
cASV12 <i>Leptocylindrus danicus</i>	KC509524.1 (99.74 100)
cASV13 <i>Fragilariopsis</i> sp.	NC_045244.1, LR812620.1, FJ002238.1 (all 97.64 100)
cASV14 <i>Bacteriastrum hyalinum</i>	FJ002166.1 (98.42 100)
cASV15 <i>Bacillariophyta</i> sp.	None
cASV16 <i>Bacillariophycidae</i> sp.	KM218905.1, FJ002217.1, FJ002224.1, GU591328.1 (all 98.43 100)
cASV17 <i>Bacillaria paxillifer</i>	AJ536452.1 (98.42 100)
cASV18 <i>Eutreptiella pomquetensis</i>	KY706202.1 (99.74 100)
cASV19 <i>Emiliania huxleyi</i>	X82156.1 (100 100); JN022705.1 (99.74 100); AY741371.1 (99.74 100)
cASV20 <i>Braarudosphaera bigelowii</i>	AB847986.2 (100 100)
cASV21 <i>Pelagomonas</i> sp.	JX297813.1 (100 100)
cASV22 <i>Pelagomonas</i> sp.	JX297813.1 (99.74 100)
cASV23 <i>Pelagomonas</i> sp.	JX297813.1 (98.68 100)

2.9.3 Supplemental Figures for Chapter 2

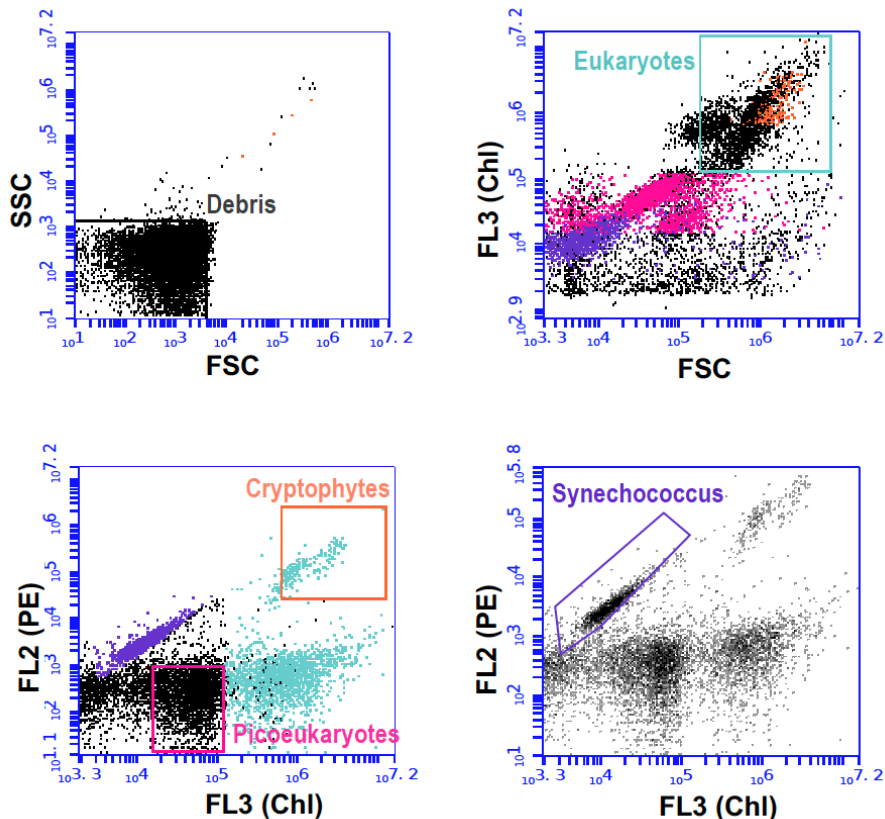


Figure 2.6. Diagram of flow cytometry gates used. See supplemental methods for further details about gate descriptions. Debris gate shown using a negative control.

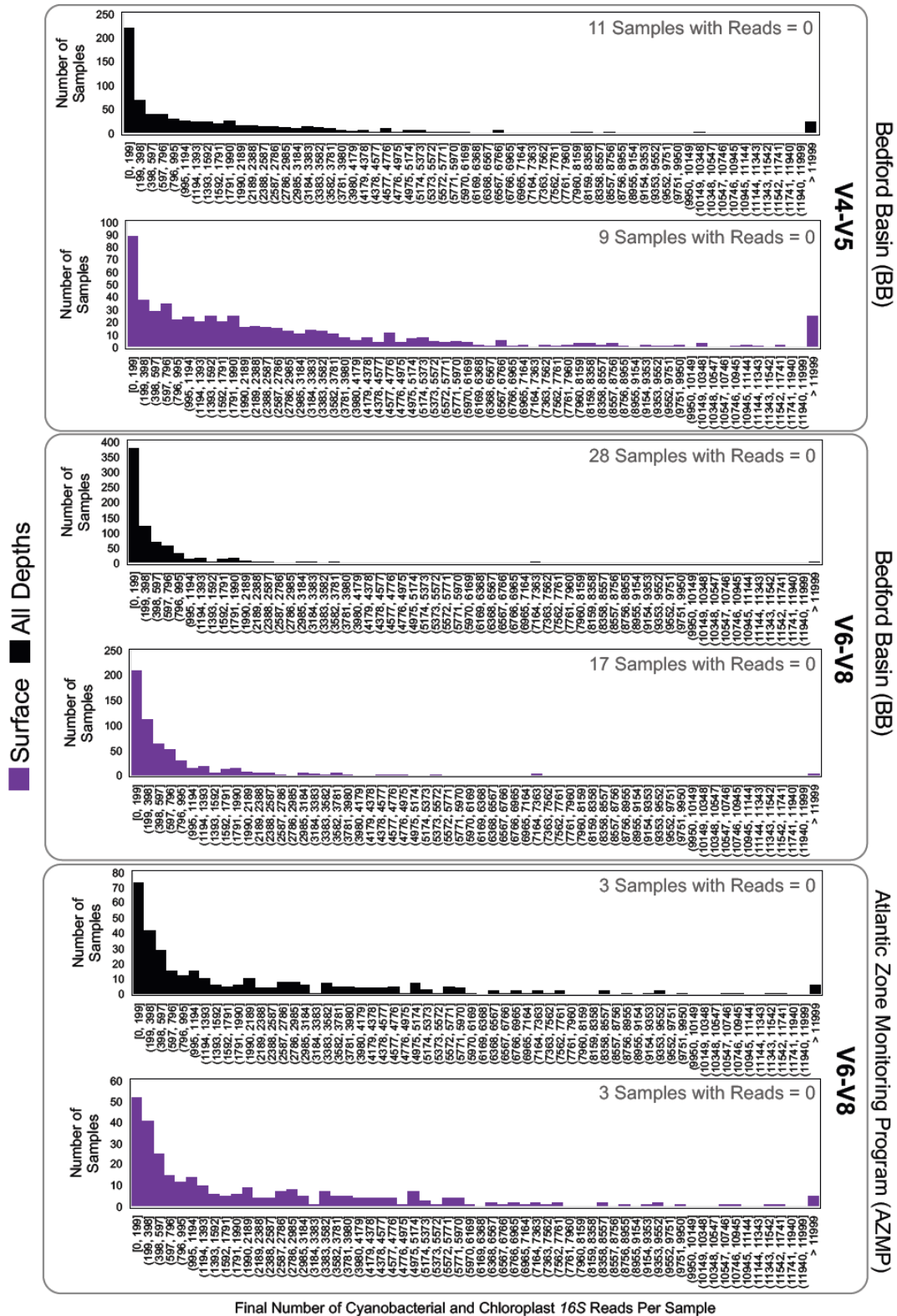


Figure 2.7. Frequency distributions for the final number of chloroplast and cyanobacterial *16S* rRNA reads per sample. BB surface = 1–10m; AZMP surface = 1–80m (includes photic zone).

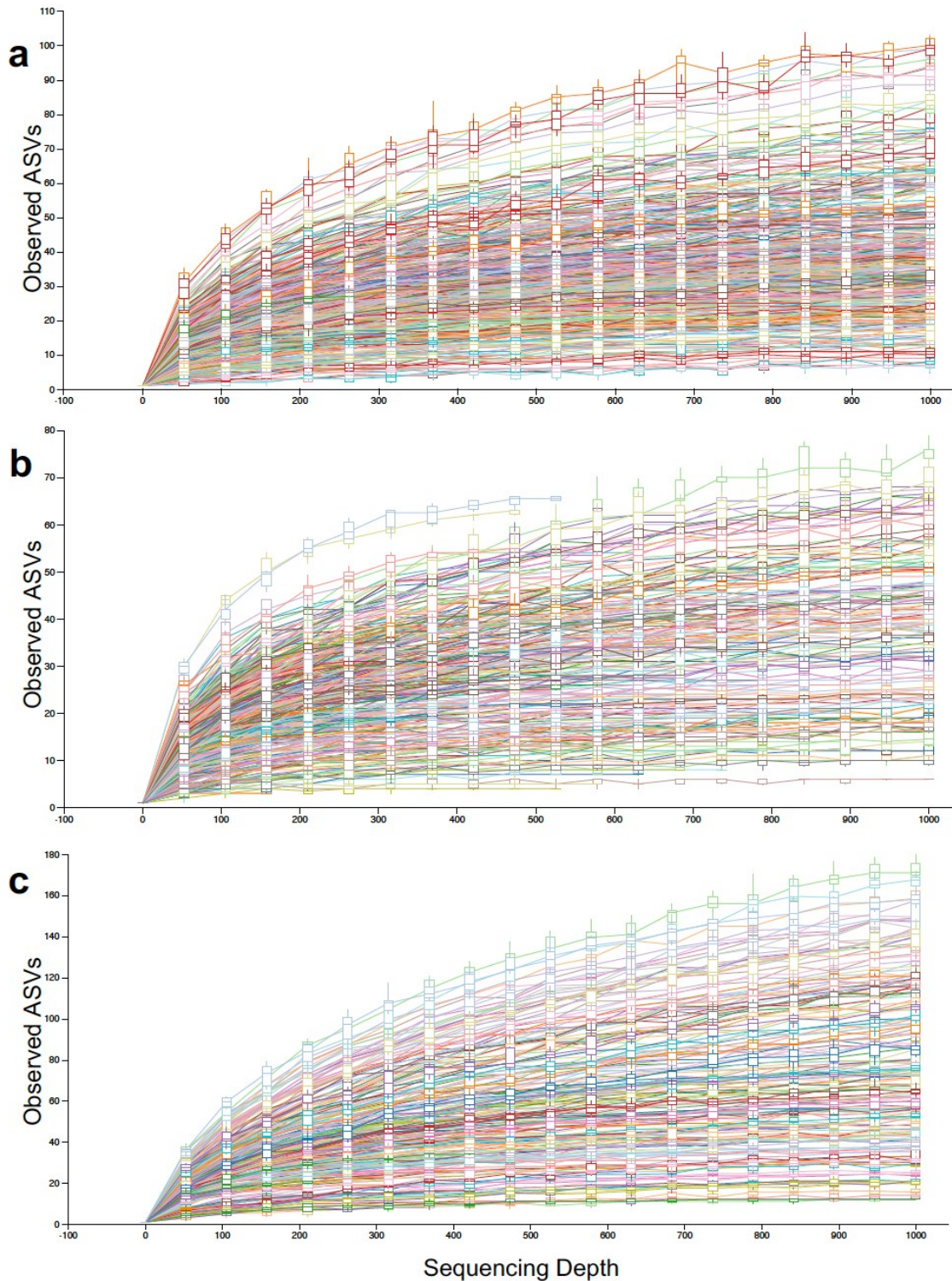


Figure 2.8. Rarefaction curves for the three *16S* rRNA datasets used in our study. Panels corresponding to datasets for: **(a)** Bedford Basin (BB) V4-V5, **(b)** BB V6-V8, and **(c)** Atlantic Zone Monitoring Program V6-V8. Colors represent samples; no legend shown as colors can repeat for different samples. Plots generated using *QIIME2 View* (Bolyen et al., 2019).

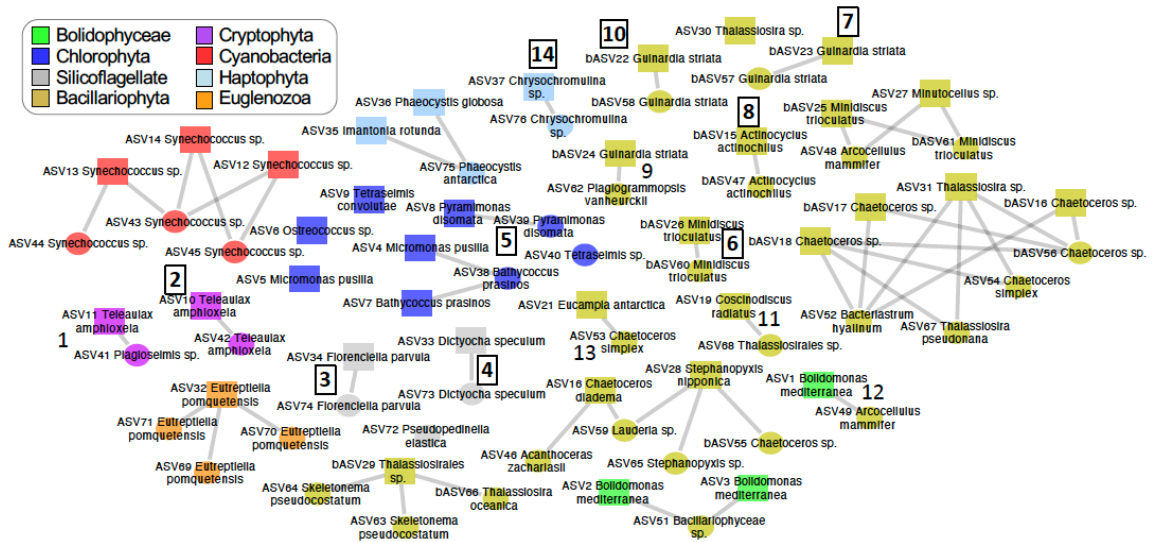


Figure 2.9. Network analysis of microbial associations for Top Twenty Bedford Basin ASVs between V4-V5 and V6-V8. The network shown is based on constructing an original network using all rarefied Bedford Basin samples (1, 5, 10, & 60m) and all ASVs and then reduced this larger network to only show statistically significant copresence relationships ($\alpha = 0.05$) for the top ASVs between V4-V5 (squares) and V6-V8 (circles). One-to-one relationships are numbered with boxes denoting equivalent taxonomic assignments. Only statistically significant positive correlations between V4-V5 and V6-V8 markers for each taxonomic group are shown.

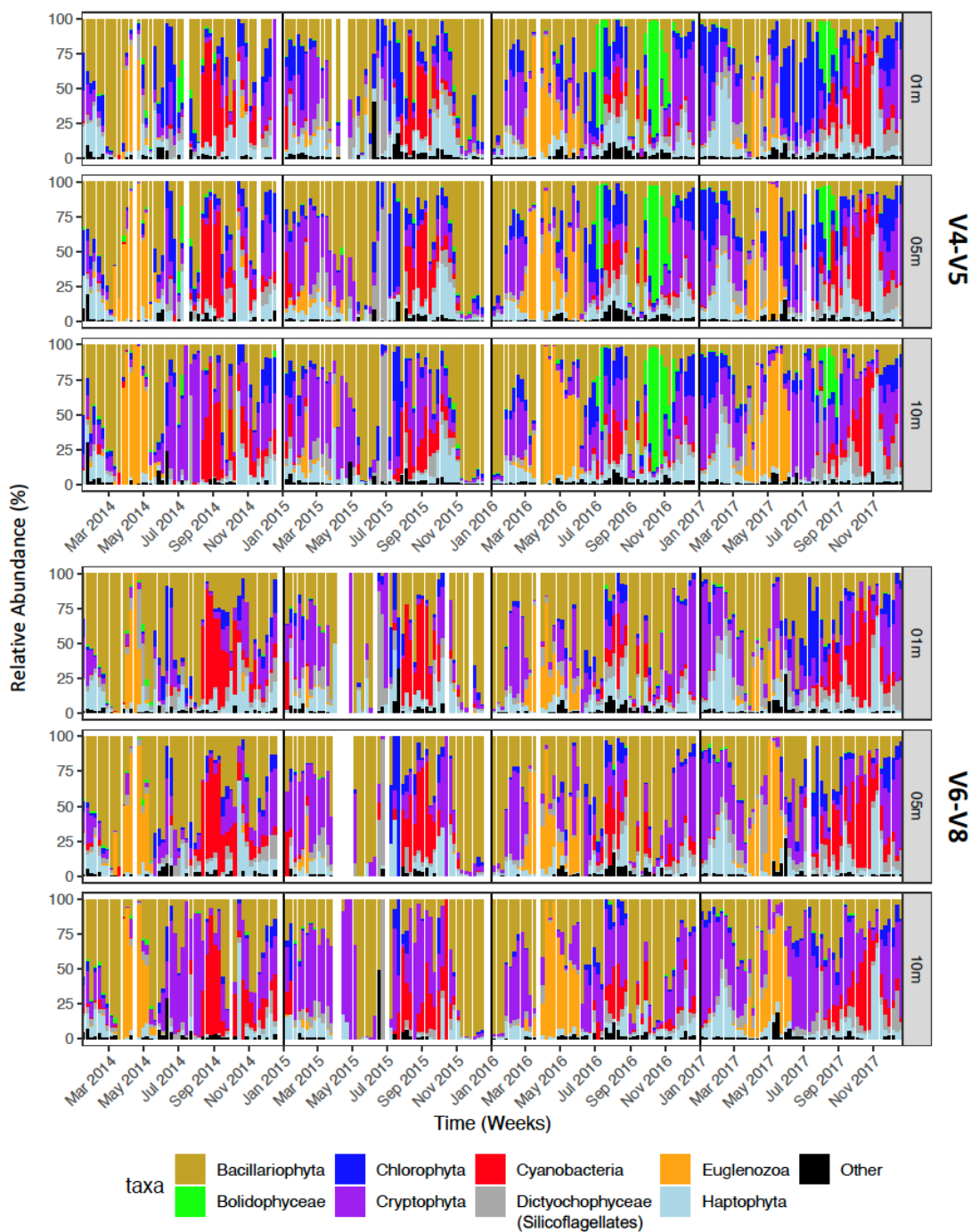


Figure 2.10. Weekly/seasonal phytoplankton trends in the Bedford Basin at 1, 5, & 10m depths as measured by unrarefied chloroplast and cyanobacterial *16S* rRNA relative abundances. Colors represent major phytoplankton groups. Note analysis uses all phytoplankton ASVs in dataset. Blank columns represent either missing samples, those with only bacterial *16S* rRNA reads, or those with unsuccessful sequencing (see Supplemental Data A1-A3 provided herein (Appendix B) for more information).

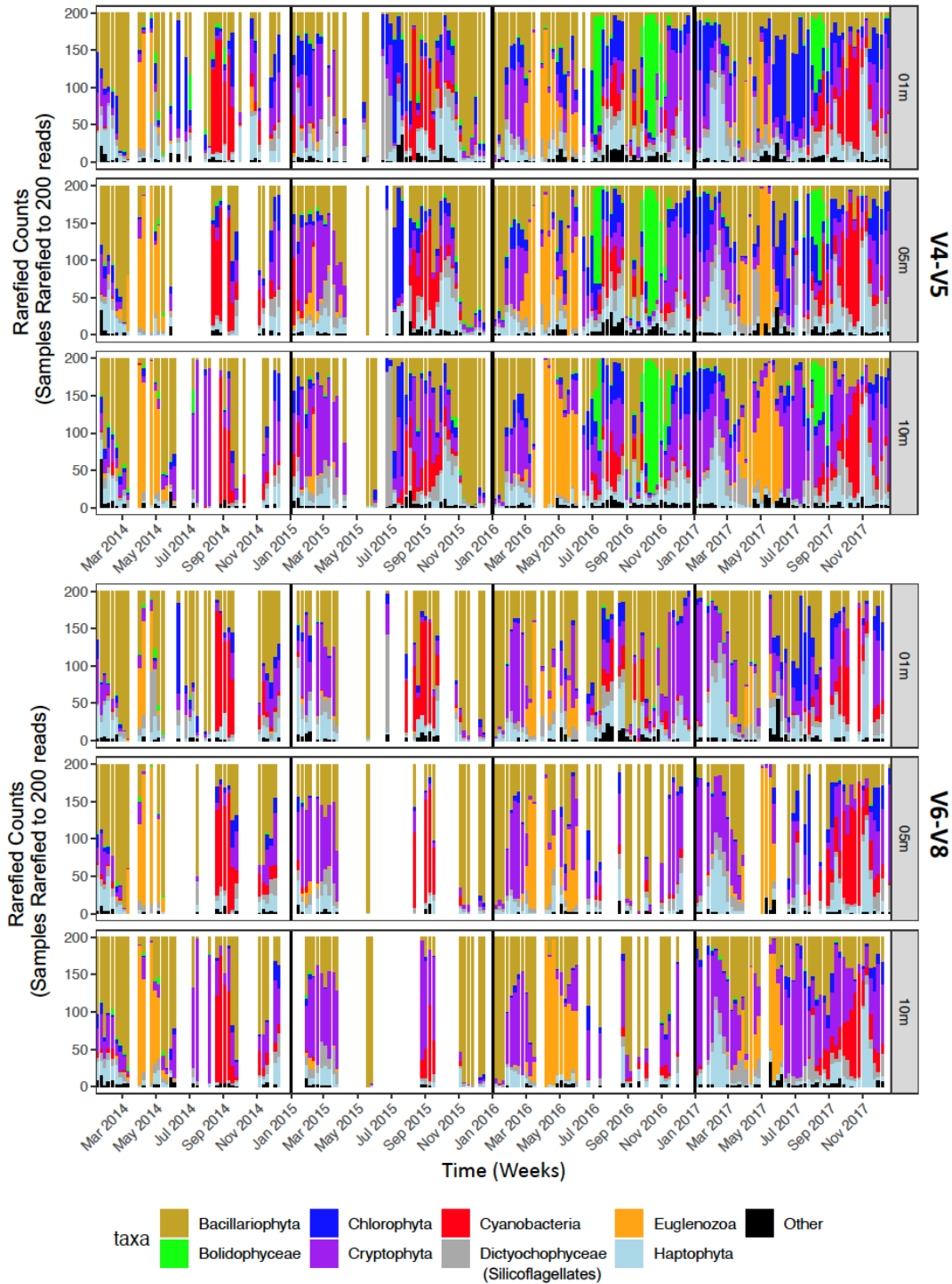


Figure 2.11. Rarefied abundances for all three surface depths (1,5, and 10m). Note that while various samples were removed during the rarefaction procedure, the full dataset does help to fill in any such gaps by capturing the diversity for at least one of the three surface depths (compare 1, 5, & 10m).

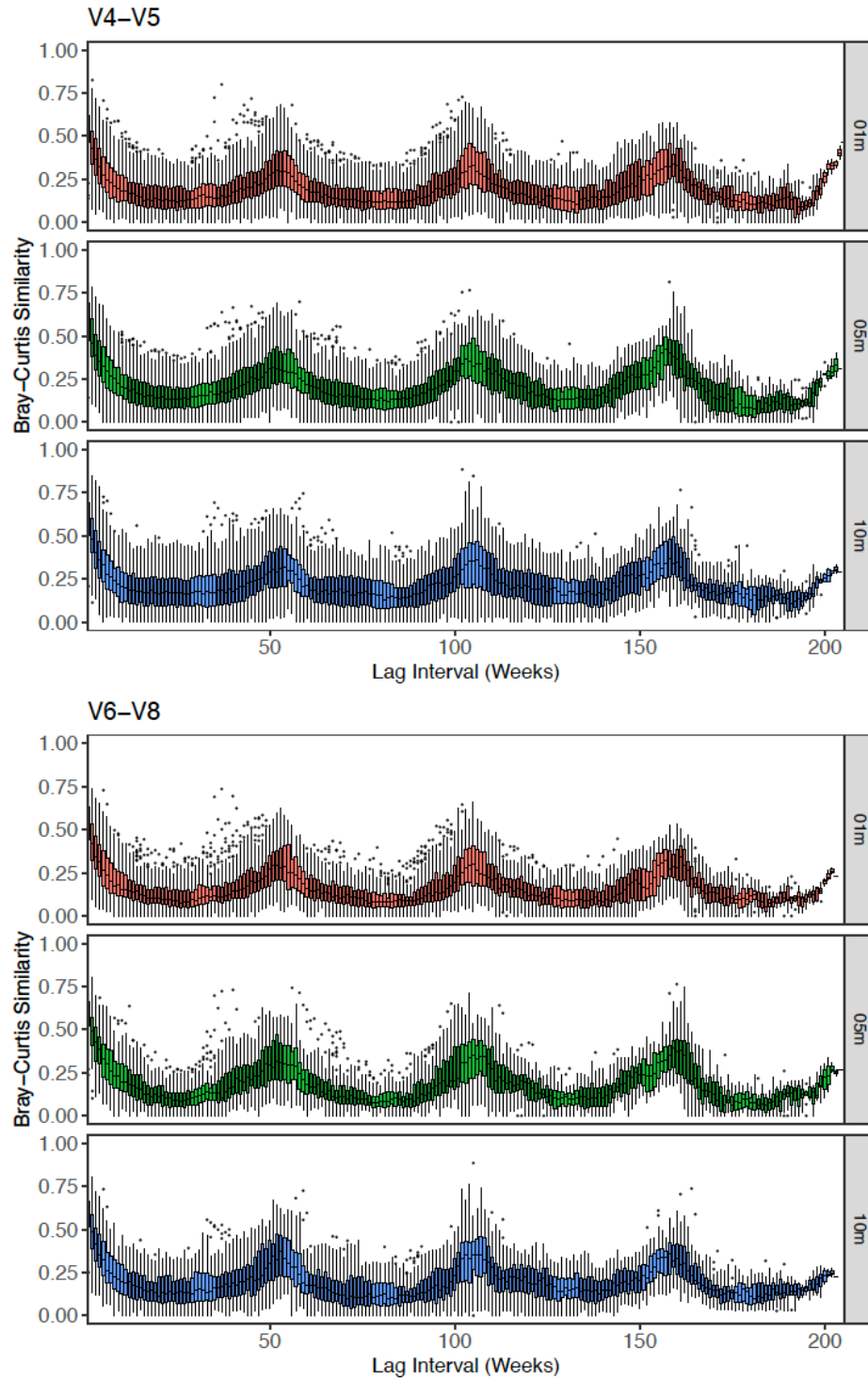


Figure 2.12. Analysis of Bray-Curtis Similarities between samples and then visualized according to the number of weeks between samples. Data rarefied to 200 reads then converted to relative abundance & Hellinger transformed (Oksanen et al., 2019). Major trend is that similarities peak at annual intervals. Compare to Supplementary Fig 2.13 for unrarefied dataset.

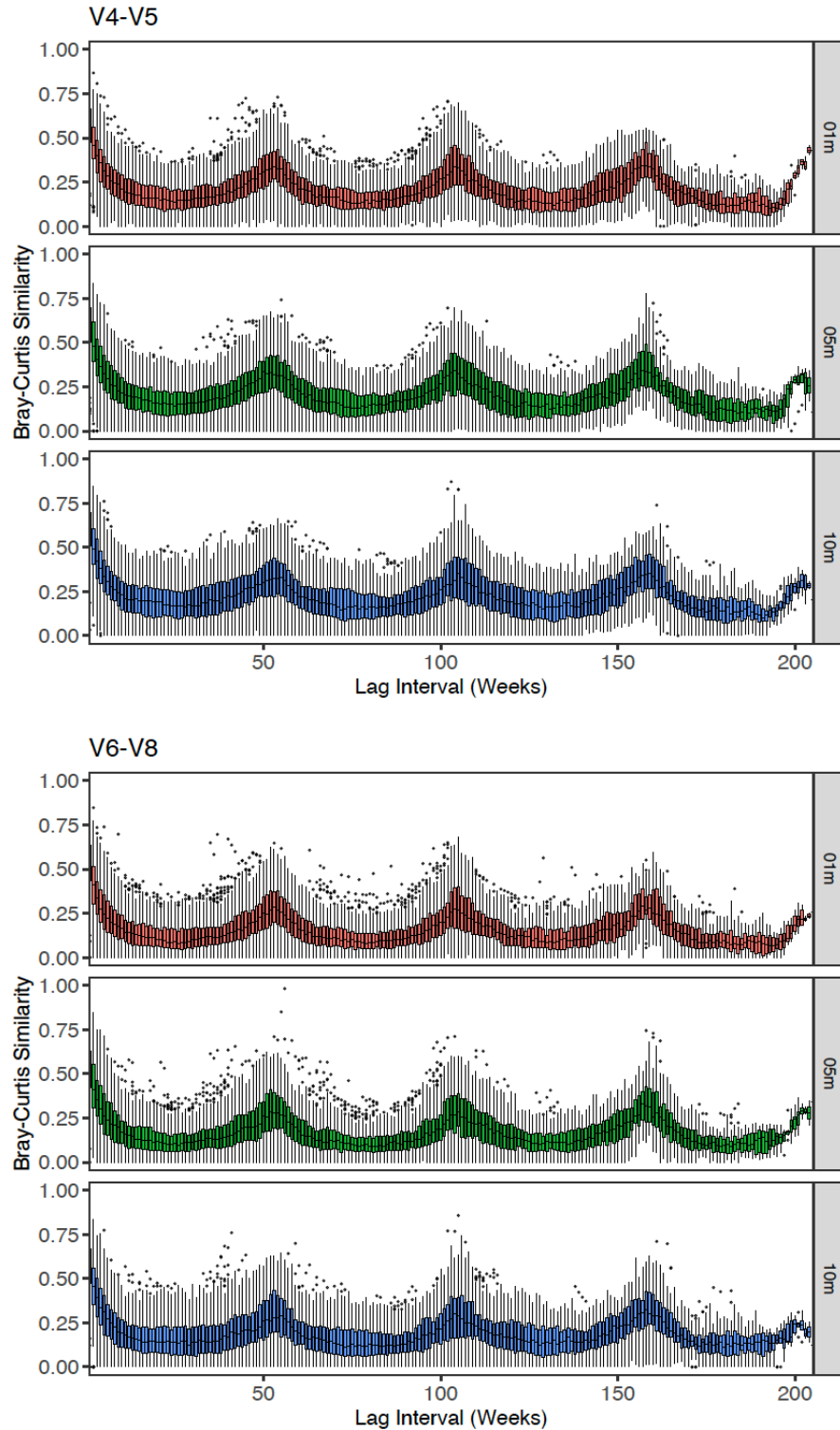


Figure 2.13. Analysis of Bray-Curtis Similarities between samples and then visualized according to the number of weeks between samples. Data are unrarefied, converted to relative abundance, and Hellinger transformed (Oksanen et al., 2019). Plot can be used to assess the effect of rarefaction on Supplementary Fig. 2.12.

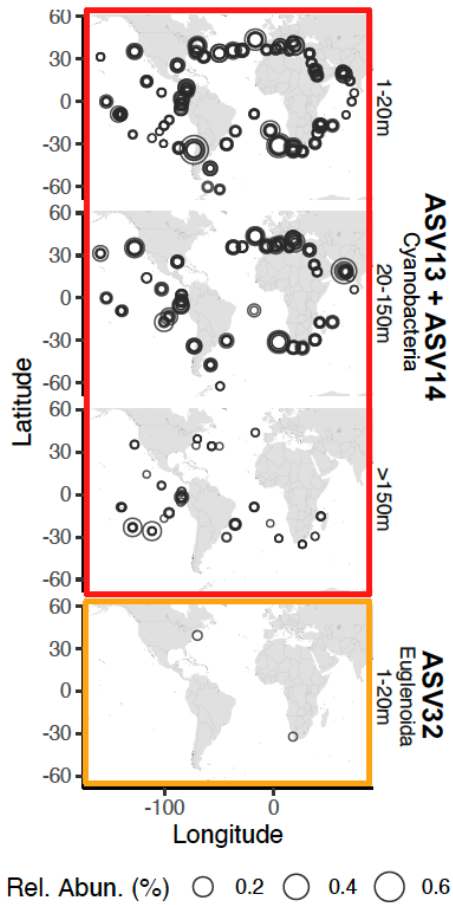


Figure 2.14. Oceanographic distributions (via *Tara* Oceans miTAGs) for cyanobacterial and Euglenozoa indicator species that are present in the Bedford Basin during the fall and spring, respectively. Relative abundances & distributions are based on *Tara* Oceans data provided by (Logares et al., 2014; Sunagawa et al., 2015). Only *Tara* miTAGs with 100% pair-wise similarity and 100% coverage to our ASVs plotted.

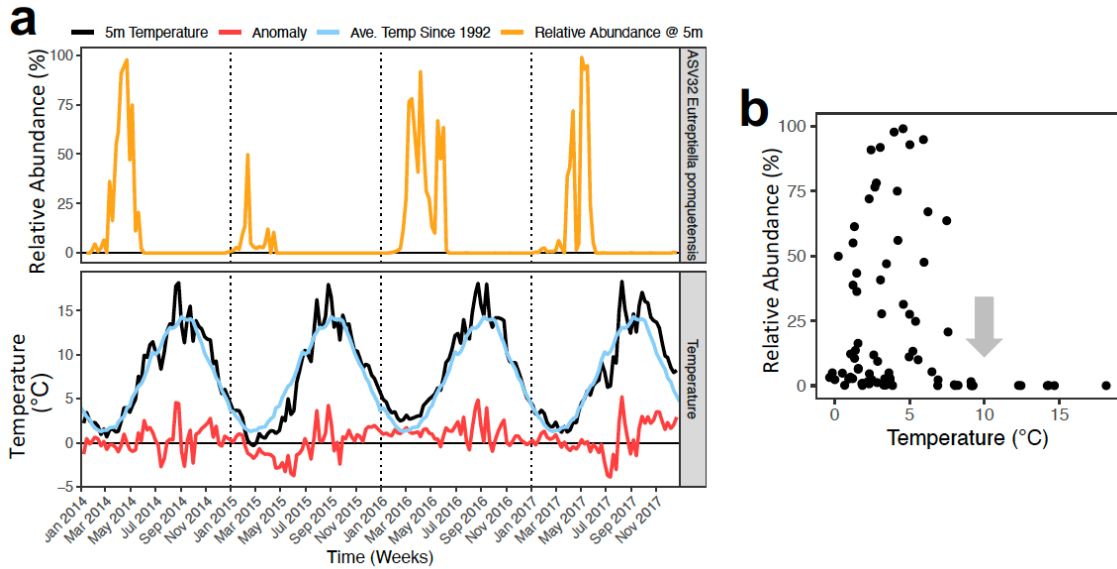


Figure 2.15. Relationship between temperature and *E. pomquetensis* as detected by the Bedford Basin molecular time-series. **(a)** Change in unrarefied relative abundance and temperature versus time; shows how sub-zero temperature in 2015 corresponded to decrease in relative abundance. **(b)** Unrarefied Relative abundance values versus temperature; shows how higher relative abundance were found generally above 0°C and less than 10°C (arrow).

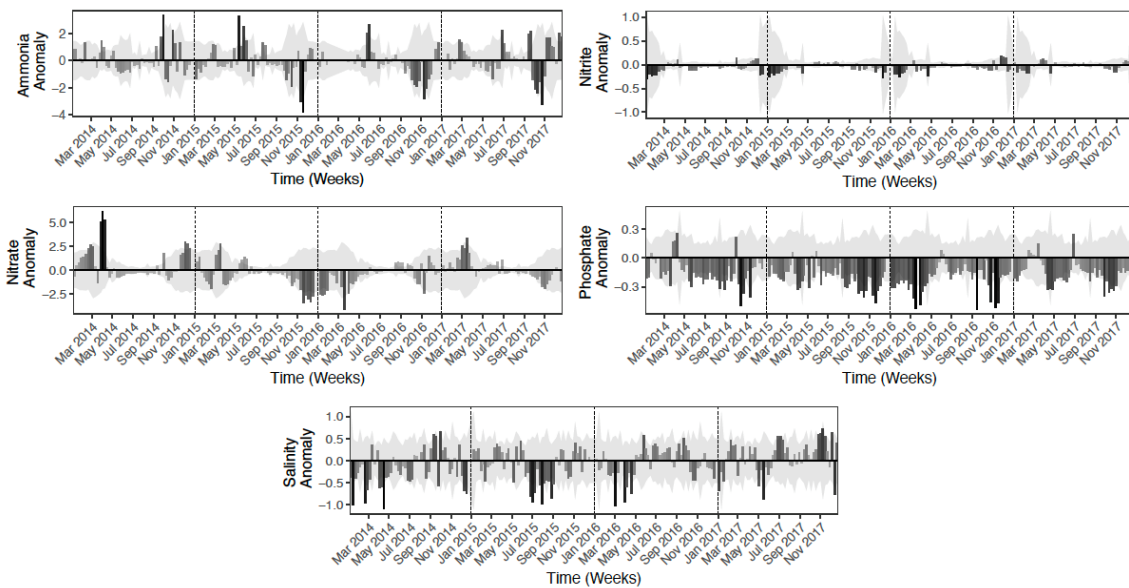


Figure 2.16. Salinity anomalies and nutrient anomalies. Values are the positive or negative difference relative to the 5m weekly mean for 1992–2017 (shading = standard deviations).

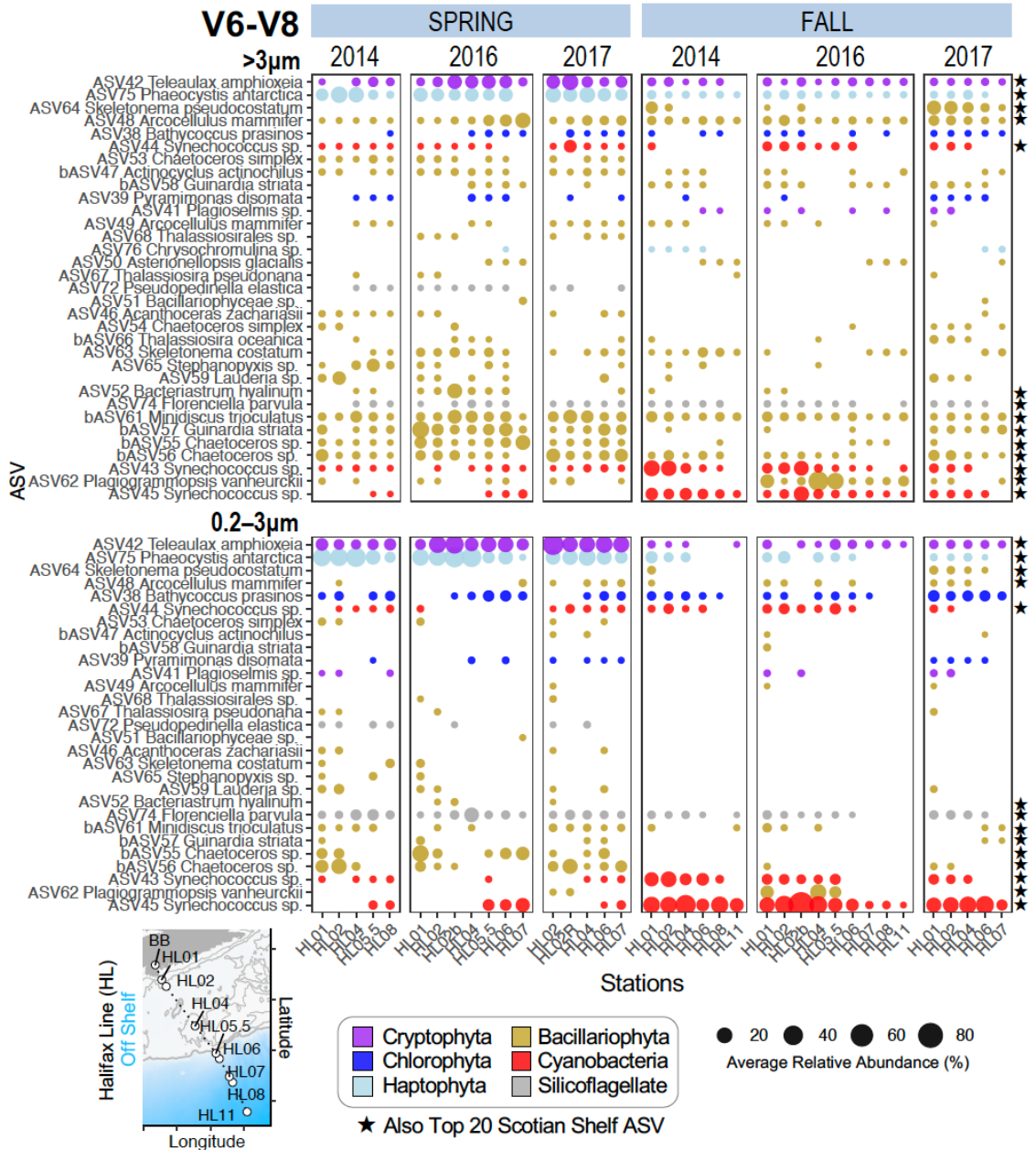


Figure 2.17. Top Twenty phytoplankton ASVs observed in the Bedford Basin time series are also detected seasonally in nearby Atlantic Zone Monitoring Program (AZMP) stations along the Halifax Line (HL). V6-V8 average relative abundances for individual ASVs throughout the water column for $>3\mu\text{m}$ and $0.2\text{--}3\mu\text{m}$ fractions (top and bottom panels, respectively; data unrarefied). To help visualize presence/absence trends, ASVs are organized along the y-axis according to hierarchical clustering using the complete linkage method, a Euclidean distance matrix, and scaled total relative abundance per ASV. Starred ASVs (★) are sequence variants that also appear among the top twenty AZMP ASVs in each sample. Dominant ASVs from the Bedford Basin without a counterpart in the AZMP are not shown. The relative abundance calculation is the same as that reported in the caption of Figure 2.4.

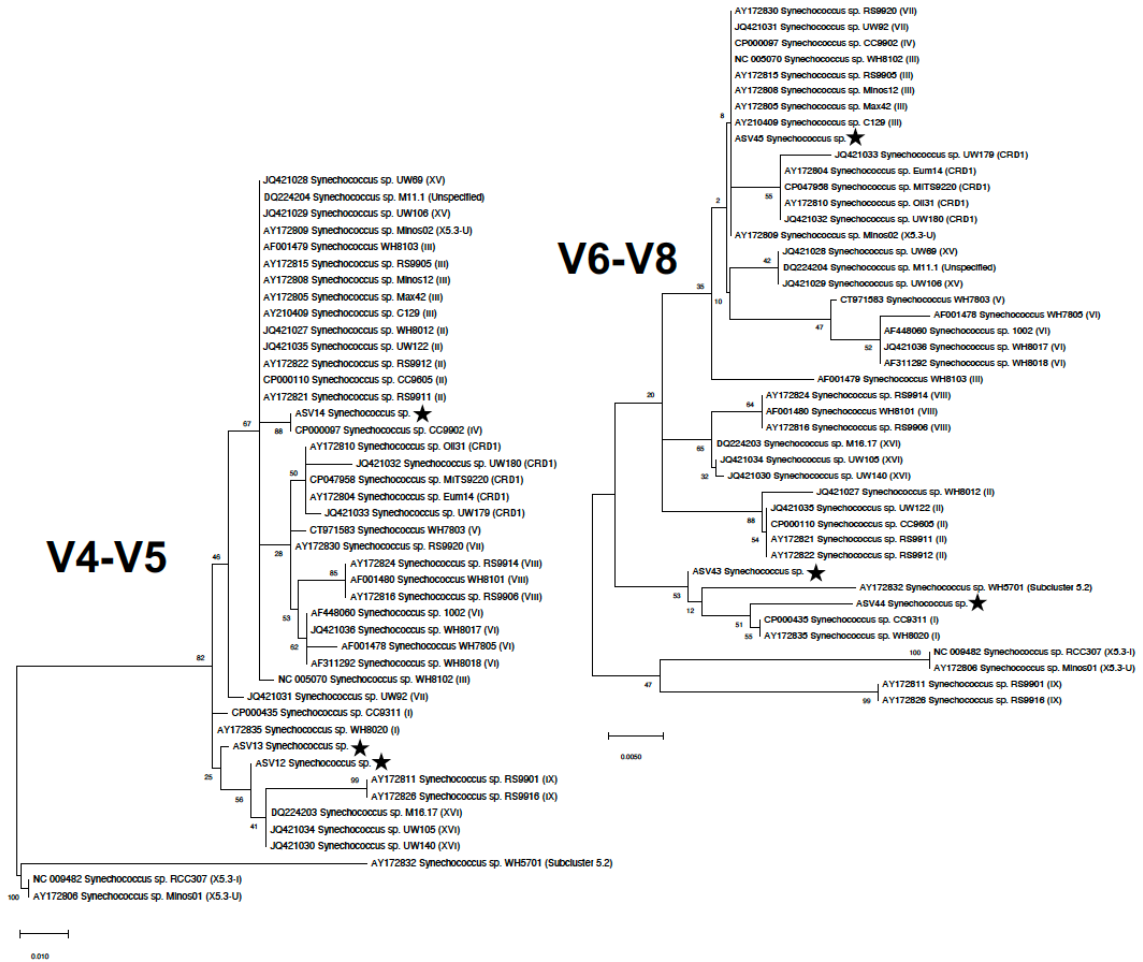


Figure 2.18. Phylogenetic assessment of *Synechococcus* ecotypes. For these trees we: aligned nucleotides via MUSCLE (Edgar, 2004), trimmed sequences to either V4-V5 or V6-V8 ASV lengths, conducted a DNA substitution model test, assumed partial deletion at 95% to account for any missing sequence that remained after trimming, and set bootstrap replicates to 500 (all conducted within MEGA (Kumar et al., 2016). Figure also makes use of literature regarding known *Synechococcus* ecotypes (Ahlgren & Rocap, 2012; Sohm et al., 2016).

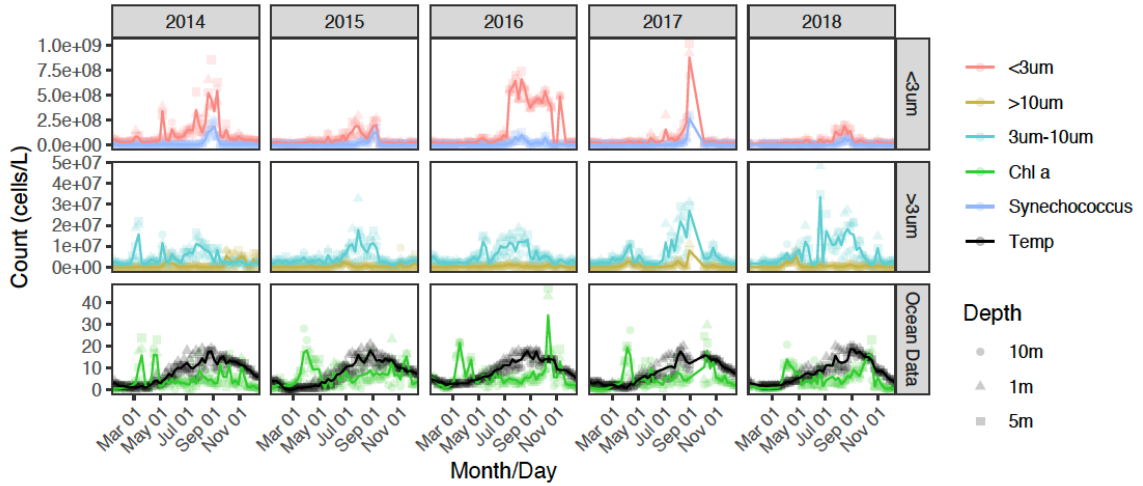


Figure 2.19. Additional time-series image of Bedford Basin flow cytometry. Figure shows all size-specific fractions and *Synechococcus*, as well as Ocean Data for chlorophyll *a* (Chl *a*) and Temperature. Lines represent averages of 1–10m data.

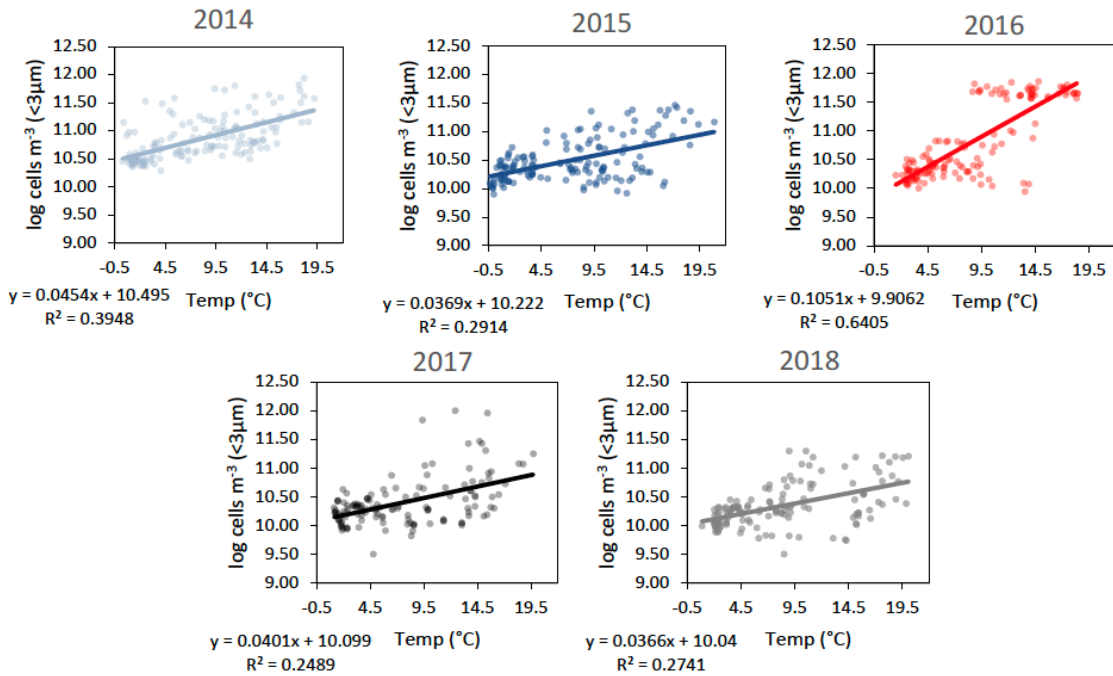


Figure 2.20. Relationship between $<3\mu\text{m}$ cell densities and temperature separated into yearly plots. Trendlines plotted as linear relationships.

CHAPTER 3

Microevolutionary Patterns in Ecotypes of the Symbiotic Cyanobacterium UCYN-A Revealed from a Coastal Time-series in the Northwest Atlantic

3.1 Abstract

Candidatus Atelocyanobacterium thalassa, also known as UCYN-A, is a globally important nitrogen-fixing microbe often found in colder temperate regions and in coastal areas where nitrogen fixation has classically been overlooked. In this study we present a high-resolution 3-year time series of UCYN-A at a coastal site in the Northwest Atlantic. Our time series integrates oceanographic measurements with weekly environmental DNA sampling for *nifH* and *16S* rRNA genes. To study UCYN-A ecotypes we used *nifH* amplicon sequencing and quantitative-PCR (qPCR) assays. Community-based *nifH* sequencing, when paired with a new quantitative assay for the often-observed UCYN-A2 ecotype, showed that high concentrations of UCYN-A dominated by A1–A4 ecotypes reoccurred annually at our study site. Although UCYN-A was detected every summer/fall, our results illustrate that the ability to observe each ecotype may be highly dependent upon sampling time given the intense interannual and weekly variability of ecotype-specific blooms; specifically, qPCR assays and *nifH* amplicon sequencing methods provided independent estimates of the relative abundance of UCYN-A across the time series, with qPCR assays indicating only a few peaks in the temporal occurrence of UCYN-A1 and A2 ecotypes, while the *nifH* ASVs indicated a more persistent presence in the diazotrophic microbial community during late summer and early fall. According to qPCR assays, in 2015 the A1 and A2 ecotypes co-occurred, while in 2016 and 2017 their appearance in the environment was temporally separated. Time series patterns also revealed that much of the rarer diversity within UCYN-A was populated by short-lived neutral mutational variants, therefore providing new insights into the microevolutionary patterns exhibited by UCYN-

A. Our data provide novel perspectives from which to interpret the delineation of UCYN-A ecotypes and their intraspecific microdiversity, while also underscoring the need to consider high-resolution datasets when attempting to generalize spatiotemporal ecologies within this well-known microbial group.

3.2 Introduction

Biological nitrogen fixation is a major source of new biologically-available nitrogen in the oceans and is a process carried out by diazotrophic microbes (Hutchins & Capone, 2022; Zehr & Capone, 2020). Prominent diazotrophs in the ocean include filamentous *Trichodesmium* (Cerdan-Garcia et al., 2021), the unicellular cyanobacterial diazotrophs of UCYN-A, *Crocospaera*, and *Cyanothece* (Zehr et al., 2001; Zehr & Capone, 2020), as well as heterotrophic bacterial diazotrophs (Cheung et al., 2021) and multiple diatom-diazotroph associations (Schvarcz et al., 2022). Herein, we focus on the Northwest Atlantic representatives of the geographically widespread UCYN-A or *Candidatus Atelocyanobacterium thalassa sensu stricto* (Farnelid et al., 2016; Krupke et al., 2015; Martínez-Pérez et al., 2016; Zehr et al., 2016).

UCYN-A lives symbiotically with haptophyte algae where it exchanges fixed nitrogen for fixed carbon produced by its algal host (Zehr et al., 2016). Unlike other cyanobacteria, and due to genomic streamlining, UCYN-A has lost genes for photosystem II and therefore cannot evolve oxygen via photosynthesis (Tripp et al., 2010; Zehr et al., 2016; Zehr et al., 2008). Recently, others have shown that UCYN-A fixes nitrogen even in the presence of nitrate (Mills et al., 2020); this pattern is likely the result of its algal host's general preference for ammonia fixed by UCYN-A over other dissolved inorganic nitrogen sources, such as nitrate (Suzuki et al., 2021). Such a tight coupling of the exchange of new nitrogen from diazotrophy exchanged for fixed carbon from the host could explain why UCYN-A has been found in areas classically thought to have less diazotrophy, such as coastal regions (Mills et al., 2020). Even so, the precise mechanisms regulating this symbiosis are still unknown given that stable UCYN-A/host associations remain generally uncultivated (Gradoville et al., 2021); for instance, others have recently demonstrated the loss of UCYN-A and maintenance of the host during cultivation attempts (Suzuki et al., 2021). UCYN-A has also been shown to fix nitrogen during the day (Landa et al., 2021;

Muñoz-Marín et al., 2019), representing yet another departure from the classical example of nighttime nitrogen-fixation and daytime carbon-fixation used by other unicellular cyanobacterial diazotrophs to overcome the adverse effects of oxygen on the iron-rich nitrogenase (Gallon, 1992; Zehr & Capone, 2020).

Overall, UCYN-A is important within the marine environment despite the sparsity of data surrounding its absolute contribution to global nitrogen fixation, given that it may be subjected to significant loss from the euphotic zone (Tang, Li, et al., 2019; Zehr, 2011). For instance, the UCYN-A/host complex is grazed by copepods (Conroy et al., 2017; Scavotto et al., 2015), bleached corals (Meunier et al., 2019), and possibly dinoflagellates (Fletcher-Hoppe et al., 2022), suggesting trophic level energy transfer via consumption by larger organisms. Furthermore, UCYN-A has been previously identified on sinking particles (Farnelid et al., 2018) and at >500m depths indicating a contribution to sinking carbon export (Karlusich et al., 2021).

Global surveys (Cabello et al., 2016; Farnelid et al., 2016; Karlusich et al., 2021) have demonstrated that UCYN-A is of widespread significance in the ocean and can occur in colder temperate regions where tropical/subtropical diazotrophs are less likely to dominate (Agawin et al., 2014; Fonseca-Batista et al., 2019). Recently, Turk-Kubo et al. (2017) and Henke et al. (2018) described upwards of eight sublineages or phylotypes within UCYN-A through the oligotyping of *nifH* environmental sequences, and these are currently named UCYN-A1 to -A8. Given that the more frequently observed A1 and A2 sublineages show differences in host specificity (Cornejo-Castillo et al., 2019), symbiont protein-coding gene similarity (Bombar et al., 2014), cell-specific nitrogen fixation rates (Turk-Kubo et al., 2021) and symbiont/host consortium sizes (Farnelid et al., 2016; Thompson et al., 2014), the UCYN-A sublineages have been referred to as ecotypes. Use of this term reflects that genetically distinct sublineages within this species exist with individual ecotypes being adapted to unique environmental conditions (Cohan & Perry, 2007; Gevers et al., 2005; Hoarfrost et al., 2019; Koeppel et al., 2008). Specifically, UCYN-A2 is found in association with larger hosts (~7.3µm in size versus ~2.3–3.6µm for A1 & A3), and reaches a higher cell density per host (~3–10 versus ~1–2 symbionts for A1), leading also to a higher nitrogen fixation rate (Cornejo-Castillo et al., 2019; Farnelid et al., 2016; Martínez-Pérez et al., 2016; Turk-Kubo et al., 2021). Earlier observations have

suggested that UCYN-A1 and -A2 may be found in the oligotrophic/open ocean and coastal waters, respectively, thus generally agreeing with the view that UCYN-A diversity can be divided into environmentally-specific ecotypes (Henke et al., 2018; Moreira-Coello et al., 2019; Turk-Kubo et al., 2017). This said, the A1 and A2 ecotypes have been known to overlap in some coastal areas including the Atlantic continental shelf of North America (Turk-Kubo et al., 2017); hence further study and *in situ* data are needed to achieve a more complete understanding of ecotype-specific environmental preferences.

Although prior time series analyses have helped to resolve the general seasonal patterns of UCYN-A, data presented by such studies have focused on time scales that are different from what we have presented herein (generally lower resolution for a longer period, or higher resolution for a shorter period). For instance, UCYN-A has been tracked via long-term monthly sampling for upwards of 8-years (Cabello et al., 2020; Fletcher-Hoppe et al., 2022; Henke et al., 2018; Moreira-Coello et al., 2019; Saulia et al., 2020) and via short-term hourly and daily sampling from several days to upwards of three months (Muñoz-Marín et al., 2019; Needham et al., 2018). For longer time series, in the Pacific both Henke et al. (2018) and Cabello et al. (2020) recently observed peaks in UCYN-A towards late summer/fall in Noumea Lagoon (New Caledonia; tropical SW Pacific) and Monterey Bay (California, USA; NE Pacific), respectively. Additionally, Moreira-Coello et al. (2019) showed highest UCYN-A (A1+A2+A4) values in surface waters of July 2015 during their periodic monthly sampling of shelf-waters near Ría de A Coruña (Spain). For shorter time series, others have initially suggested a high temporal variability for UCYN-A, with Robidart et al. (2014) showing a three orders of magnitude change in UCYN-A *nifH* copies in less than two days near Station ALOHA in the North Pacific Subtropical Gyre.

Given the demonstrated importance of UCYN-A both globally and in the coastal Northwest Atlantic (Tang et al., 2020; Tang, Wang, et al., 2019), we investigated UCYN-A occurrence patterns in the microbial community at a temperate multiyear time series site located on the eastern coast of North America (Bedford Basin, NS, Canada). Using 3-years of weekly sampling at four depths, universal *nifH* amplicon sequencing, and A1 and A2 q-PCR assays, we uncovered the seasonal and evolutionary dynamics of the numerous UCYN-A ecotypes observed at this site. Specifically, we have investigated: (i) the

microevolutionary patterns exhibited by UCYN-A *nifH* amplicon sequence variants (ASVs) through the lens of a high-resolution time series, (ii) the weekly and multiyear seasonal patterns of UCYN-A ecotypes within the coastal embayment, and (iii) the role of temperature and nutrients on defining the temporal dominance of the major A1 and A2 ecotypes within the coastal Northwest Atlantic by using qPCR assays specific for each major ecotype (A1 and A2). As a final component to our study, we also (iv) compared UCYN-A *nifH* sequences against a phytoplankton *16S* chloroplastic rRNA gene dataset collected from the same set of DNA samples to identify putative haptophyte host signatures, corresponding *16S* rRNA gene sequences for the UCYN-A1 and -A2 ecotypes, as well as other co-occurring phytoplankton.

3.3 Materials and Methods

3.3.1 Sampling and Oceanographic Data

Using Niskin bottles seawater was collected at 1, 5, 10 and 60m weekly during 2015–2017 from the Compass Buoy Station in Bedford Basin (BB; 44.6936 LAT, -63.6403 LON [or 44° 41' 37" N, 63° 38' 25" W]) and transported in the dark to Dalhousie University (Halifax, Nova Scotia, Canada) for processing. Bottles were kept in a cooler to maintain ambient temperature and reduce exposure to high light. An additional ten samples collected in 2014, were used in our phylogenetic and diversity-based analyses.

For environmental DNA (eDNA) 500mL of seawater was prefiltered using synthetic nylon mesh (2014–2015 = 160µm, 2016–2017 = 330µm) then filtered onto 47 mm diameter, 0.2µm Isopore polycarbonate filters (Millipore, Ireland) using an acid-cleaned peristaltic pump system (3% HCl rinsed, followed with extensive washes with Milli-Q water). Filters for DNA extractions were frozen in liquid nitrogen and kept at –80°C until processing.

Oceanographic data for temperature and salinity (both CTD sensor-derived), as well as bottle-derived chlorophyll *a* (chl *a*) and nutrients (for nitrate, NO₃⁻; nitrite, NO₂⁻; ammonium, NH₄⁺; and phosphate, PO₄³⁻) were collected as part of the Bedford Institute of Oceanography (BIO) 20-year Bedford Basin Monitoring Program time series (Li & Dickie, 2001) and are available directly from BIO by request and via their website (www.bio.gc.ca/science/monitoring-monitorage/bbmp-pobb/bbmp-pobb-en.php).

3.3.2 DNA extraction and amplicon sequencing

A DNeasy Plant Mini kit was used to extract DNA from eDNA filters using an enhanced lysis procedure as previously described (Zorz et al., 2019) and eluted at 50 μ L final volume. DNA concentrations were assessed using a NanoDrop 2000c (Thermo Scientific, USA). To target the *nifH* marker gene for metabarcoding, we used the nested PCR strategy of Zehr & McReynolds (1989) and Zani et al. (2000) via the primers nifH3R ‘ATRTTRTTNGCNGCRTA’, nifH4F ‘TTYTAYGGNAARGGNGG’, nifH1F ‘TGYGAYCCNAARGCNGA’, and nifH2R ‘ADNGCCATCATYTCNCC’. *NifH*, a gene encoding the Fe subunit of the nitrogenase enzyme involved in nitrogen fixation (Church et al., 2005), has been extensively used as a functional gene marker for diazotrophs (Gaby & Buckley, 2012; for recent examples see Ding et al. (2021), Gradoville et al. (2020), Raes et al. (2020), and Ratten et al. (2015)). Although some have indicated that these primers may have a natural bias against planctomycetes (Delmont et al., 2018) they continue to be used to study diazotroph communities with UCYN-A (Cabello et al., 2020). First round PCRs (25 μ L total) included: 2.5 μ L of 10 \times Buffer (Qiagen), 4 μ L of 25mM MgCl₂ (Qiagen), 2 μ L each of dNTPs (at 10mM; Invitrogen) and nifH3 & nifH4 primers (10 μ M; IDT), as well as 0.3 μ L of 20 mg/mL BSA (NEB), 9.725 μ L molecular biology grade H₂O (Invitrogen), 0.125 μ L HotStar *Taq* (Qiagen), and 2.5 μ L of DNA template. First round cycling included: 95 $^{\circ}$ C–15 min, followed by 35 cycles of 95 $^{\circ}$ C–1 min, 45 $^{\circ}$ C–1 min, and 72 $^{\circ}$ C–1 min, then a final 75 $^{\circ}$ C–10 min. Second round PCRs (10 μ L total; concentrations and suppliers as above) included: 1 μ L Buffer, 1.2 μ L MgCl₂, 0.8 μ L each of dNTPs and nifH1 & nifH2 primers, along with 0.12 μ L of BSA, 4.29 μ L of molecular biology grade H₂O, 0.05 μ L of HotStar *Taq*, and 1 μ L of nifH3/4 PCR product as template. Second round cycling included: 95 $^{\circ}$ C–15 min, followed by 28 cycles of 95 $^{\circ}$ C–1 min, 54 $^{\circ}$ C–1 min, and 72 $^{\circ}$ C–1 min, then a final 75 $^{\circ}$ C–10 min. Samples with amplified *nifH*, as assessed via agarose gel electrophoresis, were further processed for Illumina sequencing on a MiSeq instrument. To accomplish this, the nifH3/4 PCR reactions were repeated with 1/10 diluted DNA template. The product of these reactions was combined equally with the product of original undiluted nifH3/4 PCR reactions. This template was then used to generate an Illumina library using custom fusion primers merging the nifH1/2 sequences with Illumina

adapters and barcodes (Ratten, 2017). The library was purified and normalized using Just-A-Plate 96 (Charm Biotech) prior to sequencing at the Integrated Microbiome Resource center (Dalhousie University, NS, Canada).

NifH amplicon sequence variants (ASVs) were generated using *QIIME 2* version 2019.7 (Bolyen et al., 2019) along with a modified *Microbiome Helper* pipeline as described by Comeau et al. (2017) and Robicheau et al. (2022) for checking the quality of raw reads, primer removal, read pairing, ASV denoising, and the assignment of taxonomy. Modified pipeline steps included using *nifH* primers during the *cutadapt* (Martin, 2011) primer removal stage, using a *nifH* reference database (from Gaby & Buckley (2014)) and a trim-length of 325bp during *deblur* (Amir et al., 2017) ASV denoising, and using a full-length *nifH*-trained classifier during the *scikit-learn* taxonomy assigning stage via the *classify-sklearn* command (Pedregosa et al., 2011). ASVs with less than 0.1% mean sample depth were removed to account for sample bleed-through between sequencing runs. The final output was an ASV table describing the number of reads for each *nifH* ASV per sample along with putative taxonomy. Data were kept unrarefied with an average of $2,268 \pm 1,073$ SD reads per sample (see Supplemental Fig. 3.6 for distribution of sequencing depths and rarefaction curves).

We used a reciprocal alignment method to identify UCYN-A ASVs amongst the total collection of *nifH* ASVs—an approach similar in principle to the Reciprocal Best Hits method for detecting orthologs (Hernández-Salmerón & Moreno-Hagelsieb, 2020; Moreno-Hagelsieb & Latimer, 2008; Wolf & Koonin, 2012). Our strategy posits that one can retrieve all closely related UCYN-A ASVs from a pool of total *nifH* ASVs by using a local alignment search against canonical UCYN-A1 and A2 *nifH* from their reference genomes (Bombar et al., 2014; Tripp et al., 2010), in combination with a calculated nucleotide similarity threshold that accounts for *nifH* sequence dissimilarity amongst known UCYN-A ecotypes. Under this framework we prepared a workflow that could identify multiple UCYN-A ecotype ASVs (see Supplemental Methods 3.S1 and Supplemental Fig. 3.7 for more details). This workflow was completed in *Geneious Prime* version 2022.0.1 (www.geneious.com/) using *BLAST* (Johnson et al., 2008) and *MUSCLE* (Edgar, 2004) as alignment strategies. The final UCYN-A *nifH* dataset used herein contains UCYN-A ASVs subsetted from a comprehensive *nifH* ASV table along with

accompanying representative nucleotide sequences. *NifH* sequencing reads have been deposited in GenBank under the Bioproject PRJNA930772 (NCBI Resource Coordinators, 2018).

Cyanobacterial and chloroplast *16S* rRNA gene data targeting *V6-V8* variable regions for the phytoplankton present in the Bedford Basin between 2014–2017 have been previously described (in Chapter 2 and its publication (Robicheau et al., 2022)), including sample collections, DNA extraction, and a relevant phytoplankton ASV pipeline.

3.3.3 Quantitative-PCR Assays

UCYN-A1 was quantified using the qPCR assay of Langlois et al. (2008) with the following per reaction volumes (for 16 μ L total): 8 μ L of 2 \times TaqMan Universal PCR master mix (Applied Biosystems), 0.16 μ L of 10 μ M TaqMan MGB 6-FAM Probe (Applied Biosystems), 0.144 μ L each of forward and reverse 100 μ M primers (IDT), 0.32 μ L of 20mg/mL BSA (NEB), 2.232 μ L of molecular biology grade H₂O (Invitrogen), and 5 μ L of 0.5 \times DNA template. Cycling conditions were: 95 $^{\circ}$ C–10 min and then 45 cycles of 95 $^{\circ}$ C–15 sec and 60 $^{\circ}$ C–1 min (Langlois et al., 2008).

UCYN-A2 was quantified using a new assay modified from Thompson et al. (2014). Primers and probe used were: 5'-GGTTACAACAACGTTTTATGTGTTGAA-3' (forward primer), 5'-TCTGGTGGTCCTGAGCCCGGA-3' (probe), and 5'-ACCACGACCAGCACATCCA-3' (reverse primer). When this assay was tested on template DNA containing 10¹ A2 *nifH* gene copies + 10⁷ A1 *nifH* gene copies no cross-reactivity was detected (Supplemental Fig. 3.8a). Although our assay provides improved distinction from the A1 ecotype, it is assumed that it may still be somewhat cross-reactive to A3 and A4 ecotypes (Farnelid et al., 2016). Supplemental Methods 3.S2 provides additional context for our UCYN-A2 qPCR assay design. The 16 μ L total reaction volume was composed of (with suppliers as above): 8 μ L of 2 \times TaqMan Universal PCR master mix, 0.32 μ L of 10 μ M probe, 0.4 μ L each of 100 μ M forward and reverse primers, 0.32 μ L of 20mg/mL BSA, 1.56 μ L of molecular biology grade H₂O, and 5 μ L of 0.5 \times DNA template. Cycling conditions were 95 $^{\circ}$ C–10 min, followed by 45 cycles of 95 $^{\circ}$ C–15 sec and 64 $^{\circ}$ C–1 min.

Both assays were initially tested and optimized on a StepOnePlus real-time PCR thermocycler (Applied Biosystems). Full-scale analyses were run on a ViiA7 real-time PCR thermocycler (Applied Biosystems) for higher sample throughput, with the accompanying *QuantStudio* software and standard settings used to determine gene copies per reaction. Average individual reaction efficiencies at the end of the UCYN-A2 assay optimization were equal to $92\% \pm 3\%$ SD, and the final average individual efficiencies (for DNA standards + unknowns) were $A1 = 97.4\% \pm 23\%$ SD and $A2 = 87.6\% \pm 9\%$ SD (Supplemental Fig. 3.9). Efficiencies were calculated using *LinRegPCR* (Ramakers et al., 2003; Ruijter et al., 2009). DNA used for standard curves (serially diluted and ran in triplicate) and cross-reactivity tests (ran in duplicate) were ordered as individual gBlocks gene fragments (IDT; Supplemental Fig. 3.10). These DNA fragments were quantified shortly before each use via a Qubit 4 fluorometer (Invitrogen) with a Qubit 1× dsDNA High-Sensitivity assay (Invitrogen). The limit of quantification for qPCR assays was 40 copies L⁻¹.

3.3.4 Data Analysis

Unless otherwise specified, DNA sequence analyses were conducted using *Geneious Prime* version 2022.0.1 (www.geneious.com) via alignments generated using MUSCLE (Edgar, 2004). Maximum Likelihood and Neighbor-Joining phylogenies, calculations for synonymous and non-synonymous mutations relative to outgroups, and codon-based Z-tests of selection were completed in *MEGA11* (Tamura et al., 2021). For comparative purposes, we used previously published representative sequences of UCYN-A oligotypes (Turk-Kubo et al., 2017), labelled ‘oligos’ throughout. Minimum spanning networks (Bandelt et al., 1999) that reconstruct the minimum nucleotide differences connecting ASVs were produced using *POPART* (Leigh & Bryant, 2015) with the setting Epsilon = 0. *RStudio* version 2021.9.1.372 (RStudio Team, 2021), *R* version 4.1.2 (R Core Team, 2021), and the package *ggplot2* (Wickham, 2016) were used for graphical data visualizations in addition to other *R* packages listed in Supplemental Methods 3.S3. For spatiotemporal and multivariate analyses, we converted amplicon data to centered log-ratio (clr) transformed values (Gloor et al., 2017) via the *R* packages *phyloseq* (McMurdie & Holmes, 2013) and *microbiome* (Lahti et al., 2019). Our multivariate RDA analysis was computed in *RStudio*

(RStudio Team, 2021) using the package *vegan* (Oksanen et al., 2022) along with the *decostand* and *envfit* functions therein; these functions standardized our data and fit environmental vectors onto the RDA ordination, respectively [for *envfit* permutations = 999]. The network comparing copresences and mutual exclusions between *nifH* and cyanobacterial + chloroplast *16S* rRNA gene ASVs was generated in *Cytoscape* using the *CoNet* application (Faust & Raes, 2016; Shannon et al., 2003). For this network, input unrarefied and untransformed data were standardized via column normalization as part of the network construction (see Supplemental Methods 3.S3 for full *CoNet* settings; Faust & Raes, 2016).

3.4 Results

3.4.1 Diversity and Evolutionary Dynamics

In total 34,375 *nifH* ASVs (for $n = 520$ samples) were recovered from our eDNA dataset. Of these, 65 ASVs were classified as UCYN-A. An additional three variants were initially classified as UCYN-A but contained mutations suggesting that they originated from pseudogenes and were therefore excluded from the analysis [two had frameshift mutations and one had a premature stop codon]. Although UCYN-A ASVs (hereafter referred to as ‘ASVs’) represented a small fraction of the diazotroph community richness, they at times represented a large percentage of *nifH* reads per sample and therefore were considered dominant diazotrophs in the Bedford Basin. For instance, when observed at $\geq 1\%$ of the diazotroph community (Pedrós & Pedrós-Alí, 2012) the average UCYN-A relative abundance per sample (for total UCYN-A *nifH* ASVs) was $33.6\% \pm 31.4$ SD with a minimum of 1.1% and a maximum of 94.4%. Boxplots showing the unrarefied relative abundances detected for each UCYN-A variant in our entire dataset are provided as supplementary material (Supplemental Fig. 3.11).

Using the distribution of total reads per ASV (Fig. 3.1a), we divided variants into ‘major ASVs’ and ‘rare ASVs,’ where the former group had ≥ 3 reads in the entire *nifH* dataset. A Maximum Likelihood phylogeny of ASVs compared to previously known UCYN-A oligotypes (Turk-Kubo et al., 2017) and other reference sequences for UCYN-A and cyanobacterial diazotrophs indicated that ASVs recovered herein were related to known UCYN-A ecotypes of groups A1 through A4 (Fig. 3.1b). Overall, strong bootstrap

support (at 100%) confirmed that all major ASVs belonged to UCYN-A (Fig. 3.1b). We excluded the A7 oligotype from our phylogeny and main analyses as it was generally placed outside a monophyletic UCYN-A clade when we attempted to include it (Supplemental Fig. 3.12), hence its status as a UCYN-A variant is not entirely clear.

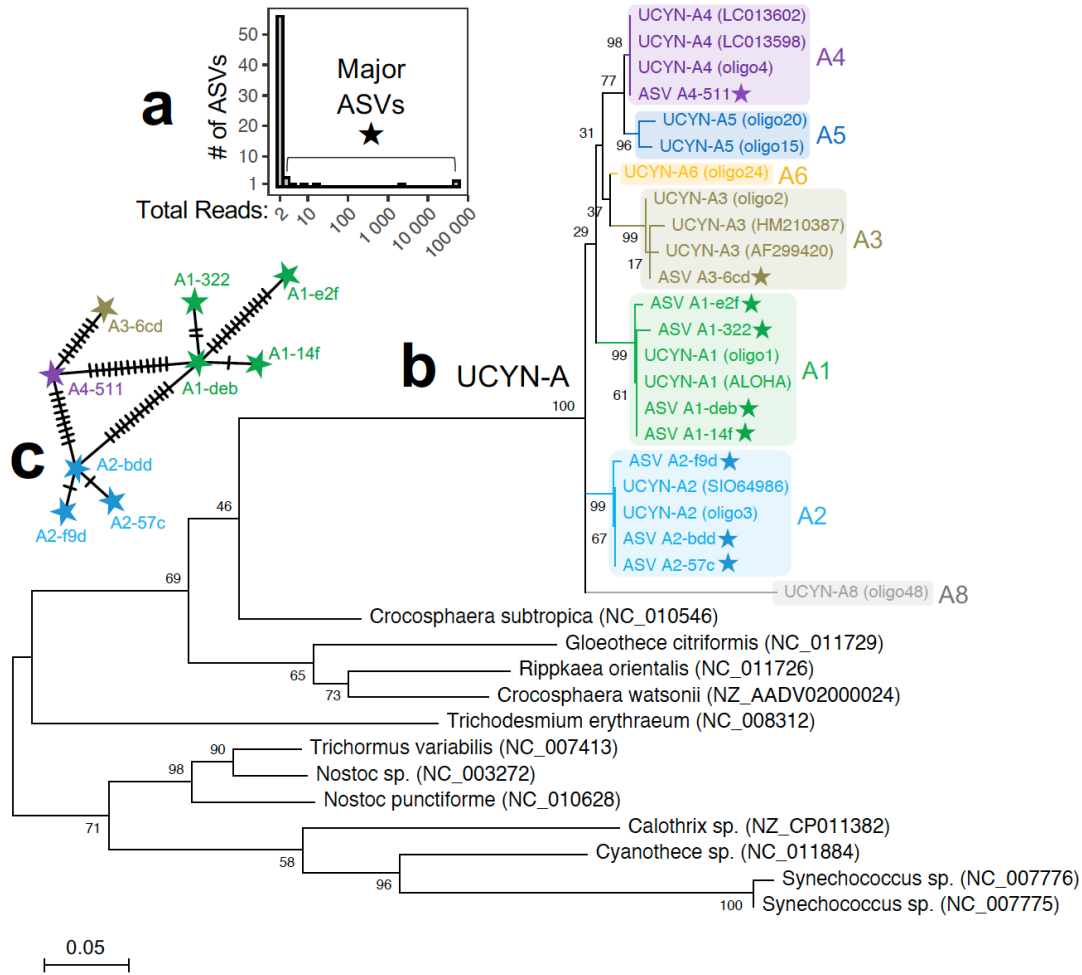


Figure 3.1. UCYN-A Ecotypes in the Bedford Basin. **(a)** Distribution of total reads observed per ASVs; most ASVs had 2 reads while only nine more dominant ASVs had >2 reads in the total dataset (starred throughout figure). Total reads for each major ASV were: A2-bdd = 50,195, A1-deb = 49,650, A4-511 = 2,586, A1-14f = 17, A3-6cd = 8, A2-f9a = 4, A1-322 = 3, A2-57c = 3, A1-e2f = 3]. **(b)** Maximum Likelihood tree of cyanobacterial diazotrophs showing UCYN-A ecotype clades and dominant UCYN-A ASVs in the Bedford Basin. **(c)** Minimum spanning network showing the fewest number of nucleotide substitutions (dashes) that connect ASVs to each other. Parameters for ML tree included: a T92+G+I substitution model, complete deletion to account for sequence gaps (248 bp in final dataset), and 1000 bootstrap replicates (final support shown at nodes). Sequences called ‘oligo’ are from Turk-Kubo et al. (2017). Minimum spanning network uses entire ASV sequence length = 325bp. Codes next to outgroup taxa are GenBank accessions.

It was possible to assign the remaining rare ASVs to known ecotypes by using a computationally simpler Neighbor-joining tree and identifying clades with relatively higher bootstrap support values (Supplemental Fig. 3.13). Many rare ASVs differed by only a few nucleotides, therefore complicating a more detailed phylogenetic reconstruction for these ASVs (see low bootstrap values in Supplemental Fig. 3.13). Although some of the major ASVs were closely related based on the Maximum Likelihood phylogeny, a minimum spanning network indicated the number of nucleotide differences amongst major ASVs could greatly vary (some had >10 differences, while others had 1–2 differences; Fig. 3.1c). This trend suggested that additional evolutionarily informative data may be obfuscated by rather short UCYN-A branches within the Maximum Likelihood tree (Fig. 3.1b).

We constructed a larger minimum spanning network to identify intraspecific relationships between all UCYN-A ASVs found at our study site (Leigh & Bryant, 2015). Although this network strategy is relatively simple (Bandelt et al., 1999), it produces an informative output. The network shows a starburst-like pattern within each ecotype, while further placing a major ASV at the center of each cluster (Fig. 3.2a). Indeed, it is quite remarkable that without any *a priori* information on ASV dominance patterns, the minimum spanning network identified from a simple alignment the correct major ASVs within each ecotype. This pattern suggests that the starbursts (or radiations) within ecotypes are an emergent feature of the UCYN-A ASVs within the Bedford Basin. Although many of the ASVs that form this network had only two sequence reads in our own dataset, this starburst pattern was also present when we applied our analysis approach to the original UCYN-A reference oligotype sequences in Turk-Kubo et al. (2017; Supplemental Fig. 3.14). This additional analysis (Supplemental Fig. 3.14) provides a robust control for our own findings given that it contains rarer variants with generally more (≥ 100) reads across samples, is produced using reference sequences stemming from a different variant calling method (i.e., oligotyping), and contains sequences generated from more geographically widespread eDNA with a different primer set (Turk-Kubo et al., 2017). Some of the additional sequence dissimilarities found in our study came from the larger *nifH* amplicon that encompasses but exceeds the smaller amplicon region targeted by the UCYN-A

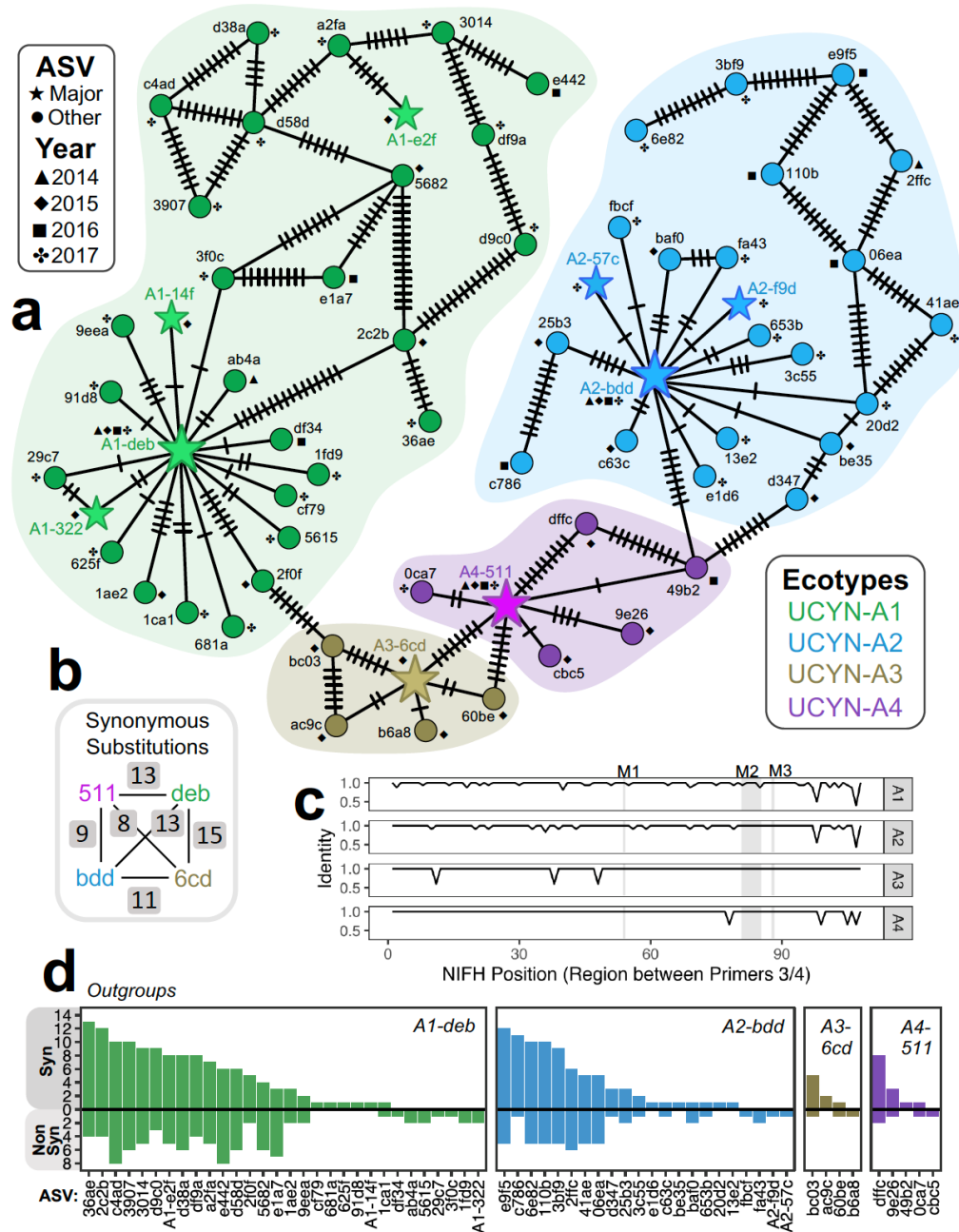


Figure 3.2. Analysis of mutations observed in UCYN-A ASVs within the Bedford Basin (BB). Main ASVs have >2 reads in dataset, other ASVs have 2 reads. **(a)** Minimum spanning network showing the minimum number of mutations that relate ASVs to each other and the years that each ASV were detected in the Bedford Basin [nucleotide differences are shown as cross-hatches], **(b)** the number of synonymous substitutions between central ASVs within each ecotype, **(c)** the position of substitutions along the translated *nifH* gene, and **(d)** whether mutations would result in synonymous or nonsynonymous mutations relative to the central ASVs used as outgroups. M1 & M3 are Fe₄S₄ motifs and M2 is the Switch II region of a binding site motif as annotated on the *nifH* gene in accession ACJ53724 (UCYN-A *nifH* conceptual translation; Tripp et al., 2010). Note only 10 samples were collected in 2014.

specific primer set used in Turk-Kubo et al. (2017), suggesting that there may be additional intra-ecotype diversity that can only be resolved by sequencing longer *nifH* gene regions (Supplemental Fig. 3.15). Note that the *nifH* ASVs used herein are 325 bp in length and the UCYN-A specific region used in previous oligotyping is ~248 bp (Turk-Kubo et al., 2017). Regardless of whether sequences originated from oligotyping or an ASV procedure, the center of ecotype-specific starbursts were sequences with 100% pair-wise similarity (Supplemental Fig. 3.15a), hence the major ASVs (or oligotypes) appear to play an important role in providing the main DNA signatures from which other rarer ASVs are related. While major ASVs, with one exception, reoccurred annually from 2014–2017, we did not observe any rare ASVs that occurred year-after-year (Fig. 3.2a). The major ASV of the A3 ecotype [ASV A3-6cd], and indeed the A3 ecotype as a whole, was only observed in 2015 (Fig. 3.2a). Rare ASVs typically appeared together with major ASVs (Supplemental Fig. 3.16), and similarly rare ASVs with more numerous nucleotide differences also occurred alongside major ASVs throughout the time series (Supplemental Fig. 3.17).

We further observed a difference in the number of nonsynonymous and synonymous mutations (and hence levels of nucleotide selection) between rare ASVs and major ASVs. Major UCYN-A ASVs had many synonymous substitutions and no nonsynonymous substitutions relative to each other, and therefore in theory code for identical *nifH* amino acid sequences within the region amplified (Fig. 3.2b). In comparison, although rare ASVs generally lacked mutations within gene regions that code for important NifH protein motifs (Fig. 3.2c), they did exhibit nonsynonymous substitutions compared to the major ASVs as outgroups (Fig. 3.2d). Z-tests of selection (Supplemental Table 3.1) also demonstrated that rare ASVs were primarily evolving in a strictly neutral fashion [44/61 ASVs or ~72%] or otherwise via purifying selection [17/61 or ~28%]. Note that mutational hotspots in A1, A2, and A4 ecotypes towards the end of the NifH region amplified (when translated into amino acids) were consistent with mutational hotspots in other diazotrophs (Supplemental Fig. 3.18).

3.4.2 High-resolution Seasonal Patterns within the Coastal NW Atlantic

In general, the Bedford Basin displayed patterns typical of a seasonally stratified temperate coastal basin (Li & Dickie, 2001). Increases in phytoplankton (detected as peaks in chl *a*) were evident at the beginning of spring when surface water begins to warm, and again at the end of summer when smaller phytoplankton are known to bloom (Li & Dickie, 2001; Li et al., 2006; Robicheau et al., 2022). Typically, macronutrients, particularly nitrate, were higher during periods of winter mixing and encountered surface-depletion in the summer months (Figs. 3.3a and 3.3b). As previously reported (Haas et al., 2021; Robicheau et al., 2022), 2016 was an anomalously warmer year than 2015 and 2017 (Fig. 3.3a).

The *nifH* genes of UCYN-A1 and A2 ecotypes were enumerated by qPCR within the Bedford Basin at weekly frequencies over three complete years at both surface (1, 5, 10 m) and 60 m depths (Fig. 3.3c), given their dominance in diazotroph communities at our study site. During this time, maximum *nifH* copies for A1 and A2 ecotypes were 2.55×10^6 *nifH* copies L⁻¹ on 17-Sept-2015 at 5m and 2.13×10^6 *nifH* copies L⁻¹ on 05-Jul-2017 at 1m, respectively (Fig. 3.3c). Across all three years, UCYN-A reliably grew in surface depths (1–10m) each year with observed maxima during the summer/early fall when ammonia and nitrate were low in the surface layer (Fig. 3.3)—a pattern consistent with UCYN-A’s known metabolic status as a diazotroph (Martínez-Pérez et al., 2016). The time series also revealed obvious differences in the annual patterns of UCYN-A growth, pointing to complex temporal occurrences for the different ecotypes within a single coastal basin (Fig. 3.3c-d). In 2015 all four ecotypes were initially present, decreased together, and then all but A3 reappeared. In 2016 and 2017, A2 was the initial ecotype present; however, eventual increases in A1 (at events E1 & E2 in Fig. 3.3c) were both preceded and succeeded by increases in A2. We suspect that the differing UCYN-A growth patterns through 2015 relative to 2016 and 2017 may be connected to environmental conditions being more generally favorable to UCYN-A during 2015 given that this year exhibited all A1–A4 ecotypes and UCYN-A at both surface and deeper 60 m depths (Fig. 3.3).

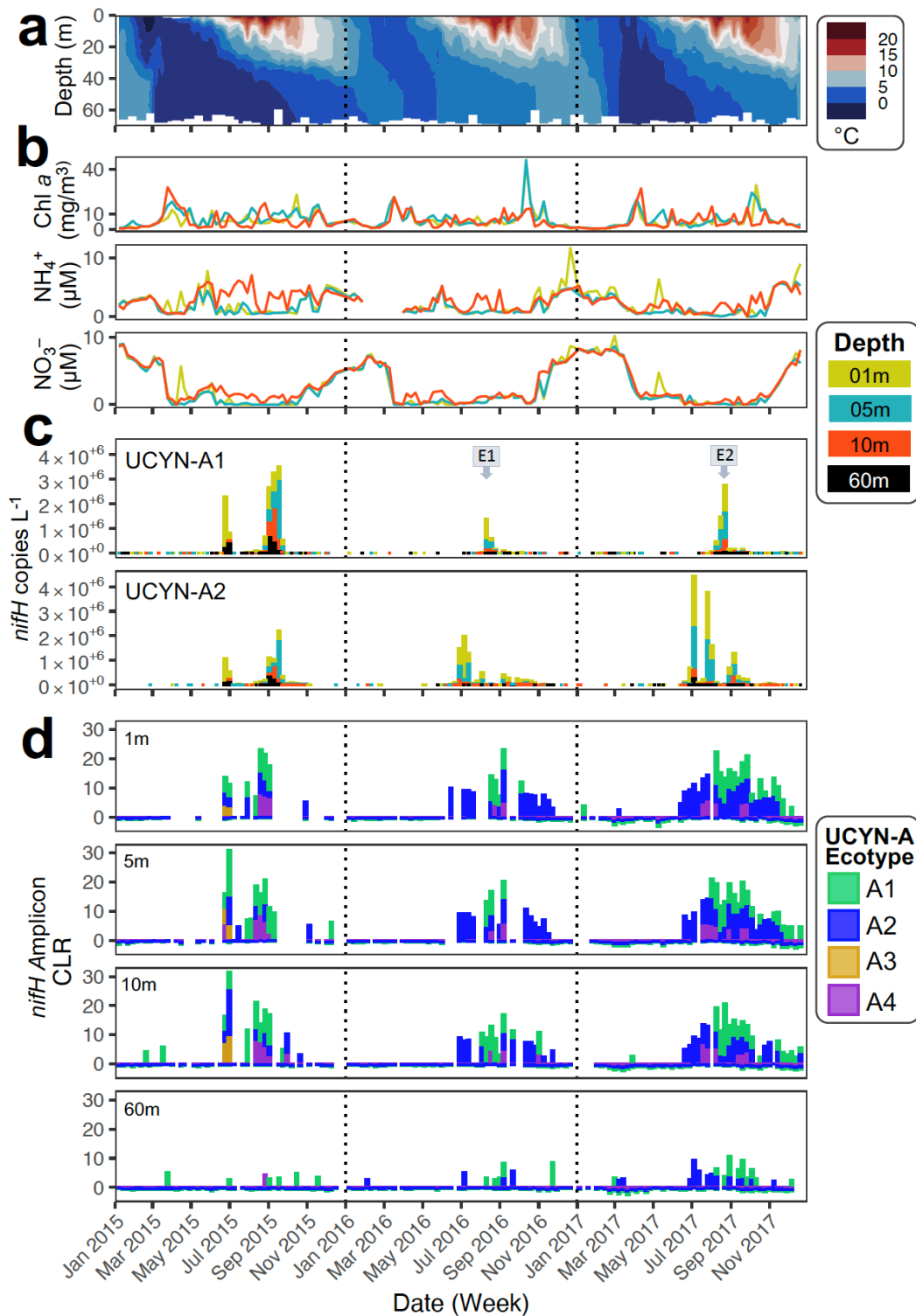


Figure 3.3. Weekly distribution of UCYN-A ecotypes in the Bedford Basin over three years. **(a)** Temperature profiles, **(b)** surface nutrient trends, **(c)** quantitative PCR profiles for UCYN-A1 and UCYN-A2, **(d)** weekly *nifH* metabarcoding profiles for UCYN-A ecotype clusters [values are center log ratios determined per sample from unrarefied sequence reads]. Events 1 and 2 (E1 & E2) are years where UCYN-A1 peaks did not overlap with UCYN-A2 peaks.

Although tidal action and occasional intrusion events of Scotian Shelf seawater are known to affect the Bedford Basin (Kerrigan et al., 2017; Shan & Sheng, 2012; Shi & Wallace, 2018). Haas et al. (2021) recently showed that 2015 had no detectable 60m intrusions and the main 60m intrusion events of 2016–2017 [in late May 2016, early November 2016, and late Dec 2017] did not coincide with the seasonal time-period that supports the growth of UCYN-A. In contrast, a mid-depth intrusion at 30m in 2017 (Haas et al., 2021; Shi & Wallace, 2018) did coincide with an initial peak in A2 presence followed by an immediate decrease in A2 before its eventual reappearance a few weeks later (Supplemental Fig. 3.19). Although the basin and shelf are physically connected, the net flow of surface water is outwards from the basin onto the shelf (Kerrigan et al., 2017). Hence, taken collectively, our observations suggest that the growth of UCYN-A in the Bedford Basin likely resulted from environmental conditions acting upon a resident or ‘overwintering’ UCYN-A population, as opposed to intrusion events being a primary source for UCYN-A in the basin. This view is also strongly supported by the very low-level detection of UCYN-A during timepoints outside its main bloom period (Fig. 3.3c).

3.4.3 *In situ* Temperature affected the Dominance of A1 versus A2 Ecotypes

Redundancy analysis (RDA) and regression fitting of environmental parameters indicated that temperature explained the highest amount of variance for major UCYN-A ASVs within samples ($r^2 = 0.56$, $p < 0.001$; Supplemental Table 3.2). Nutrients, salinity, and Chl *a* were also significantly correlated to UCYN-A composition, however, the latter two variables explained less variance ($r^2 = 0.04$ to 0.05 for Chl *a* and salinity, versus $r^2 = 0.16$ to 0.26 for nutrients; Supplemental Table 3.2 and Fig. 3.4a). The axis of RDA1 explained 39.43% of variance and mainly separated A2-bdd and A1-deb ASVs, each with >49,000 total reads in the dataset, from other major ASVs with fewer reads (3–2,586 total reads in the dataset; Fig. 3.4a). *NifH* CLR values plotted exclusively by day of the year further confirmed the role of temperature in explaining ecotype patterns in the Bedford Basin; A1 predominantly appeared in the Bedford Basin during the hottest weeks of the year (when temperatures approached $\sim 15^\circ\text{C}$ and DIN was consistently low; Fig. 3.4b). Although DIN was low when maximal CLR values for A1 were detected, comparably low DIN values were also observed prior to the appearance of A1 when A2 was initially

predominant in the basin, hence there were some periods when low nutrients were also associated with high CLR values for the A2 ecotype (Fig. 3.4b). When present, the major ASV of ecotype A4 [A4-511] also appeared around the same time as the A1 ecotype. Comparisons of *nifH* copies L⁻¹ versus temperature were likewise consistent with temperature influencing *in situ* growth of A1 versus A2, with A1 reaching higher *nifH* copies per litre of seawater at temperatures primarily above ~9°C (Fig. 3.4c). This was not the same for A2 which reached similarly high *nifH* copy numbers even below this approximate temperature threshold (Fig. 3.4c). Although salinity was relatively stable during UCYN-A's occurrence within the basin (Fig. 3.4b), we did observe that the A2 ecotype, more so than the A1 ecotype, could overlap with the fall phytoplankton bloom (Figs. 3.4a-b). Moreover, *nifH* CLR values plotted over time relative to temperature, DIN, and phosphate, demonstrated that the gradual reduction in UCYN-A towards the end of its annual appearance in the basin paralleled the seasonal decrease in temperature and progressive increase in surface macronutrients that reflect fall to winter transitions when destratification and winter-mixing come into effect (Fig. 3.4b; Li & Dickie, 2001).

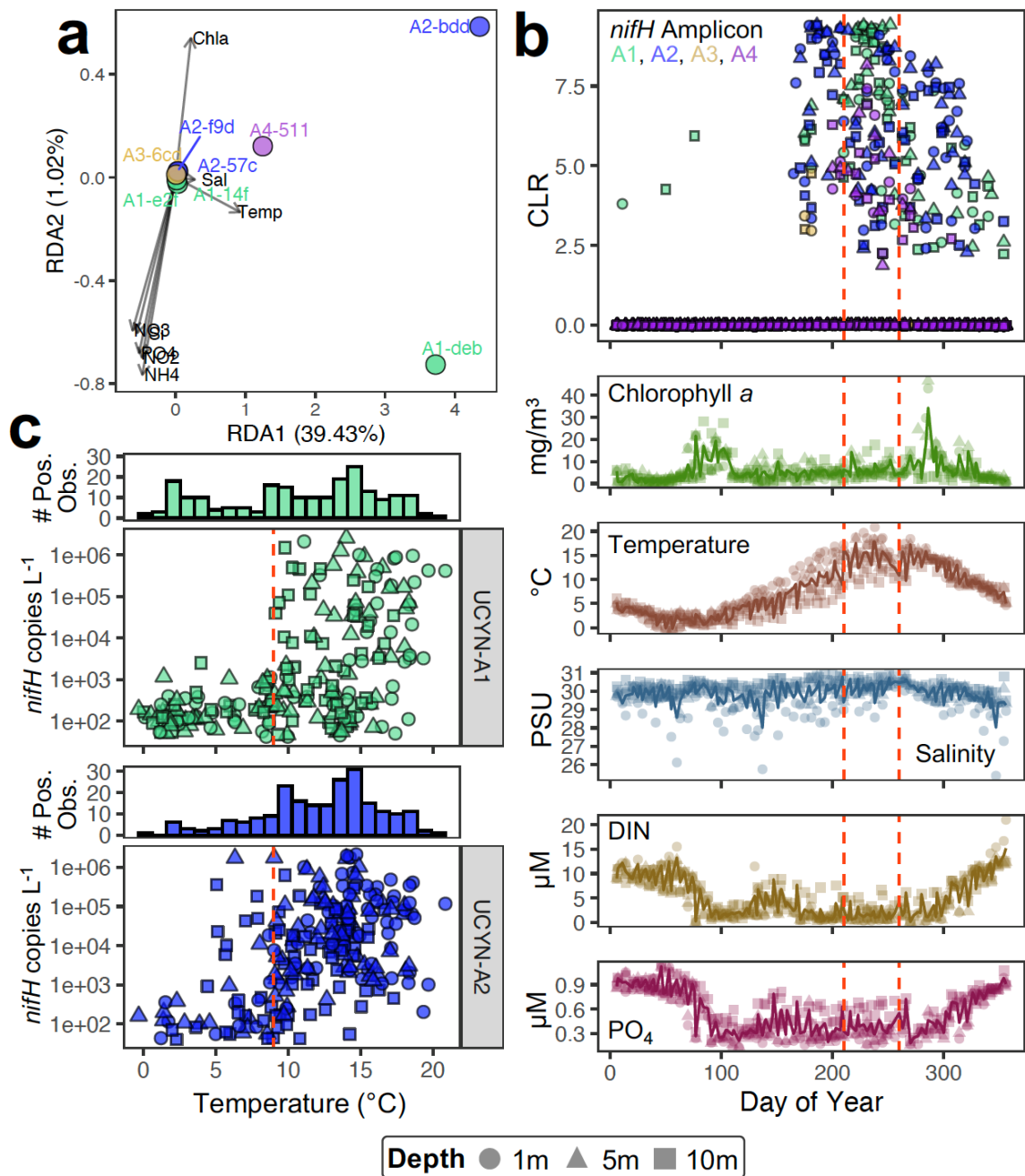


Figure 3.4. The occurrence of UCYN-A ecotypes within Bedford Basin surface waters (1–10m) in relation to environmental parameters. **(a)** Redundancy analysis (RDA) of major UCYN-A ecotypes present in the Bedford Basin along with **(b)** 3-yr of collated weekly data for CLR of individual *nifH* ASVs, chlorophyll *a*, temperature, salinity, dissolved inorganic nitrogen (DIN = nitrate + nitrite + ammonium), and phosphate, as well as **(c)** quantitative PCR counts for A1 and A2 ecotypes versus temperature. Dashed red lines (in panels b and c) demark general thresholds for major UCYN-A1 appearance/growth. The shape legend applies to both panels (b) and (c). For environmental data weekly means (as lines) are overlaid atop of individual data points. # Pos. Obs. = sum of positive observations

(i.e., total number of samples with UCYN-A detected at a given temperature). Abbreviations in (a) are chlorophyll a (Chla), salinity (Sal), temperature (Temp), nitrate (NO₃), nitrite (NO₂), silicate (Si), phosphate (PO₄), ammonium (NH₄).

3.4.4 UCYN-A Algal Hosts are Amongst Co-occurring Phytoplankton Community

A network analysis of Bedford Basin *nifH* versus chloroplast and cyanobacterial ASVs (see Supplemental Fig. 3.20 for initial network) recovered *16S* rRNA gene signatures for both A1 and A2 ecotypes (see ASV 1982* & ASV cbd7*, respectively, in Fig. 3.5a). We detected the known host of UCYN-A2 within the Bedford Basin (see ASV 5f2d** *Braarudosphaera bigelowii* in Fig. 3.5a; Cornejo-Castillo et al., 2019; Hagino et al., 2013; Thompson et al., 2014). Given the ability to accurately resolve both A1 and A2 *16S* rRNA signatures, as well as an A2 host signature via the constructed network, we consider the only Haptophyta ASV significantly co-occurring with UCYN-A1 [ASVe5c7 *Chrysochromulina* sp.] to be a putative A1 haptophyte host signature in the Bedford Basin. We also noted that ASV 45c7 *Chrysochromulina* sp. was more closely related to *B. bigelowii* cp*16S* rRNA than to any other haptophyte ASVs co-occurring with UCYN-A in the network (Supplemental Fig. 3.21 and Fig. 3.5a; Hagino et al., 2013; Thompson et al., 2014). Spatiotemporal CLR data for all ASVs that had a significant relationship to UCYN-A are further shown (see Fig. 3.5b for strong copresences, and Supplemental Fig. 3.22 for weak copresences and mutual exclusions). Hierarchical clustering allowed us to distinguish between ASVs that had copresences simply because they occurred for extended periods, rather than occurring more exclusively alongside UCYN-A (Fig. 3.5b). Collectively, our *16S* rRNA data confirm that the UCYN-A/host consortium occurs within the Bedford Basin, while also providing substantial insight into the coastal phytoplankton community that co-exists with UCYN-A within this region of the Atlantic.

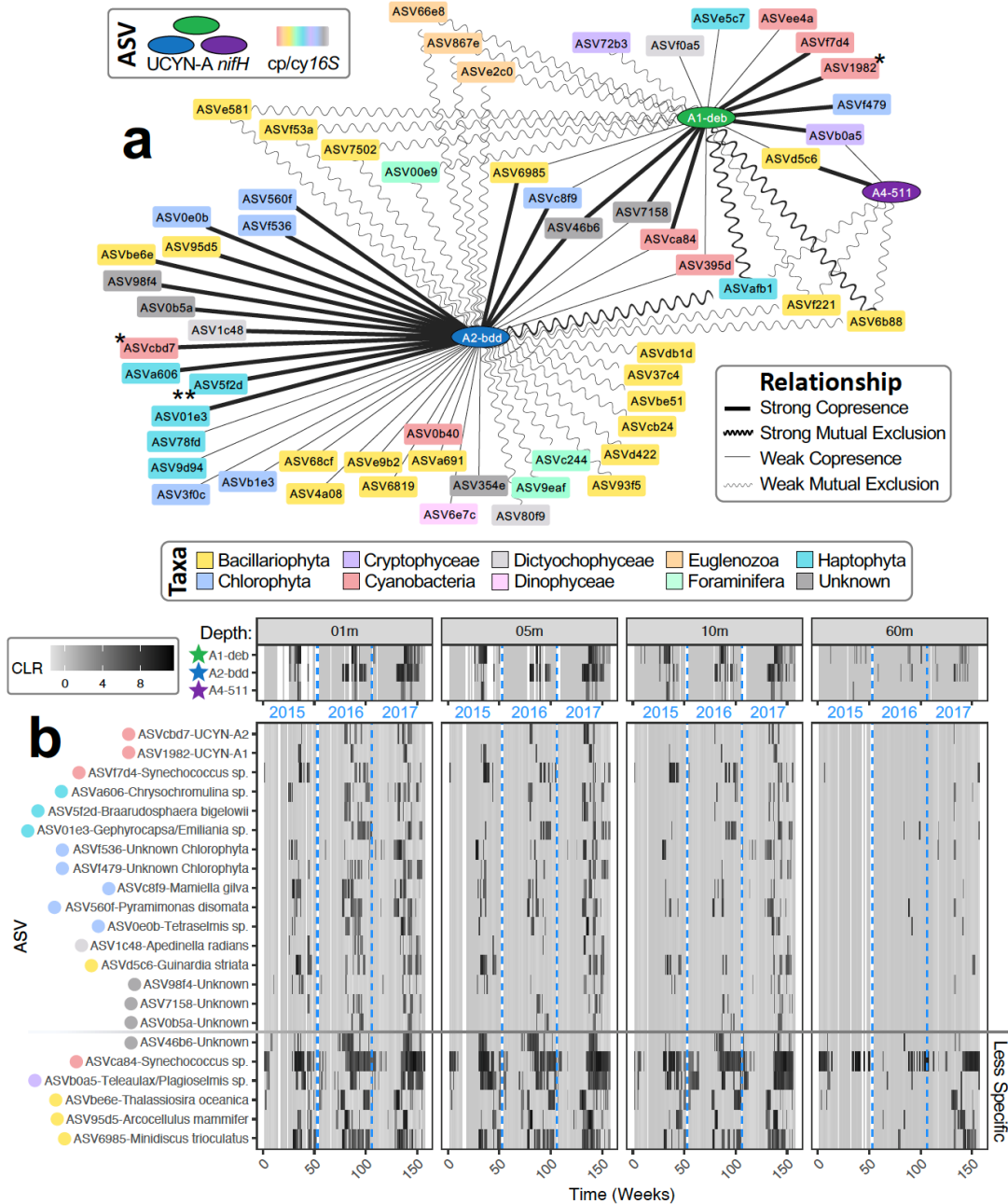


Figure 3.5. Network analysis of UCYN-A *nifH* ASVs versus phytoplankton *16S* rRNA ASVs (based on chloroplast and cyanobacterial V4-V5 *16S* rRNA gene signatures). **(a)** Subnetwork showing relationships exclusively between UCYN-A *nifH* ASVs and other phytoplankton ASVs [see Supplemental Fig. 3.20 for complete network]. **(b)** CLR spatiotemporal profiles for ASVs with strong copresence relationships to UCYN-A [lower panel] along with weekly UCYN-A *nifH* CLR values [upper panel]. Other relationships are shown in Supplemental Fig. 3.22. Hierarchical clustering was used to separate out ASVs with a less specific copresence to UCYN-A. “Strong” network scores had final p -values $< 1.11 \times 10^{-16}$ and q -values $< 5.86 \times 10^{-15}$, while “Weak” network scores were also significant with final p -values < 0.01 and q -values < 0.04 [see Section 3.9.1.3 Supplemental Methods 3.5.3 for more details on the methodologies used in the network analysis shown].

3.5 Discussion

3.5.1 Multiple UCYN-A Ecotypes Inhabit a Coastal Basin in the NWA

Both the maximum *nifH* copy numbers we detected for UCYN-A [~ 2.55 to 2.13×10^6 copies L^{-1} ; Fig. 3.3c], and our observations of the A1 to A4 ecotypes in the Bedford Basin, reaffirm that the coastal waters of the NWA contain high concentrations of UCYN-A and are home to multiple ecotypes within this species. These values align with similarly high UCYN-A *nifH* copies (1×10^7 *nifH* copies L^{-1}) found near the Mid-Atlantic Shelf Waters by Mulholland et al. (2012) and the coast of New Jersey (USA) by Tang et al. (2020), and from a broader perspective are consistent with some of the highest recorded values within this species (Cabello et al., 2020; Tang et al., 2020).

There is a strong suggestion in the literature that the A2 & A4 ecotypes often co-occur in coastal regions, while A1 & A3 often co-occur in offshore waters (Henke et al., 2018; Moreira-Coello et al., 2019; Turk-Kubo et al., 2017, 2021). However, as others have pointed out, this categorization method is not perfect given that A1 and A2 have been shown to occasionally co-exist (Messer et al., 2015; Moreira-Coello et al., 2019; Turk-Kubo et al., 2017). Based on the *nifH* community sequencing conducted herein, we observed that ecotypes A1, A2, A3 and A4 could co-occur within the inshore coastal waters of the Bedford Basin (Fig. 3.3). Also, A4 had an annual distribution pattern somewhat comparable to A1, in that this ecotype appeared during the warmest weeks of the year (Fig. 3.4). Although it is still possible that A1 and A2 have different niches on a larger geographic scale (open ocean versus coastal, respectively (Henke et al., 2018; Moreira-Coello et al., 2019; Turk-Kubo et al., 2017, 2021)), our results do argue that far more specialized niche definitions will likely be needed to explain the variable temporal dynamics exhibited by individual UCYN-A ecotypes especially within the coastal realm.

3.5.2 Phytoplankton Co-occurrences Suggest Presence of UCYN-A/Host

Consortium

Our network describing co-presences and mutual exclusions between *nifH* ASVs and chloroplast/cyanobacterial ASVs provides a comprehensive picture of the phytoplankton community associated with UCYN-A at our study site, and by extension the coastal NWA (Fig. 3.5 and Supplemental Fig. 3.22). We have demonstrated elsewhere that many of the

phytoplankton within our study site overlap with the phytoplankton found directly outside the Bedford Basin, suggesting that results from our study can be extended to the nearby Scotian Shelf (Robicheau et al., 2022). Phytoplankton with strong co-occurrences specific to UCYN-A in the Bedford Basin included: three ASVs representing different Haptophyta species [a *Chrysochromulina* sp., an ASV equally similar to *Gephyrocapsa* and *Emiliana* sp., and an ASV identified as *Braarudosphaera bigelowii*—the UCYN-A2 host (Cornejo-Castillo et al., 2019; Hagino et al., 2013; Thompson et al., 2014)], five Chlorophyta variants, *Apendinella radians* [Dictyochophyceae], *Synechococcus*, and the diatom *Guinardia striata* (Fig. 3.5). Comparison to the literature suggests parallels exist between these groups and other phytoplankton reportedly associated with UCYN-A in either the eastern North Pacific Ocean or western South Atlantic Ocean. Consistent with our study, Needham et al. (2018) observed connections between UCYN-A, *Synechococcus*, *B. bigelowii*, and an unknown Dictyochophyceae near Catalina Island, California [we considered this taxon to be different from *A. radians* found herein given that reanalysis of this unknown Dictyochophyceae yielded ~98% similarity to *Florenciella parvula* via *BLAST* (Johnson et al., 2008)]. Meanwhile, *Braarudosphaera*, *Emiliana* (now *Gephyrocapsa*; Bendif et al., 2019), *Thalassiosira*, *Teleaulax*, and *Chrysochromulina* were recovered by G erikas Ribeiro et al. (2018) from *nifH* and autotrophic *18S* rRNA network assemblages near Brazil. Of interest, these authors also directly correlated a *Chrysochromulina 18S* rRNA signature (a member of the Prymnesiales (Hagino et al., 2013)) with an *18S* rRNA signature from the host of UCYN-A1 (G erikas Ribeiro et al., 2018). These trends—in addition to our own findings that the only Haptophyta co-occurring with UCYN-A1 was a *Chrysochromulina* sp. (ASVe5c7 in Fig. 3.5 and Supplemental Fig. 3.22)—are in general agreement with current views that a small prymnesiophyte is the host of UCYN-A1 (Cabello et al., 2016; Thompson et al., 2014).

Not only does our network analysis provide evidence of the symbiotic UCYN-A/host consortium within the Bedford Basin, but presumably other nonsymbiotic microbial interactions involving UCYN-A are embedded within our *nifH* versus *16S* rRNA datasets. Such interactions may include phytoplankton growth connected to the input of new nitrogen from UCYN-A (Wang et al., 2021) or potential predator-prey interactions (Farnelid et al., 2021; Frias-Lopez et al., 2009; Hartmann et al., 2013). Although a full

investigation into these topics was outside the scope of our study, we anticipate that the fine-scale *in situ* measurements offered by the Bedford Basin time series will provide a rigorous baseline from which to formulate and test new hypotheses regarding the interactions between the UCYN-A/host consortium and other phytoplankton.

3.5.3 UCYN-A Temporal Variability and Microevolutionary Patterns

Our time series of UCYN-A (Fig. 3.3) largely agrees with prior research focusing on temporal and seasonal patterns, where peaks in UCYN-A occur towards summer/fall (Cabello et al., 2020; Henke et al., 2018; Moreira-Coello et al., 2019). However, unlike prior time series of UCYN-A we observed: (i) a strong temporal difference between the occurrence of the A1 and A2 ecotype that was revealed using a more specific qPCR assay for the UCYN-A2 ecotype and (ii) a reset of all minor UCYN-A *nifH* variants each year and the retention of the dominant UCYN-A variants. These two findings will be discussed below.

3.5.3.1 The Dominance of UCYN-A1 and UCYN-A2 in the Coastal NW Atlantic Shows Strong Temporal Variability that is Modulated by Temperature and Nutrients

Our findings suggest that prior to our modified qPCR assay that was presented herein, the UCYN-A2 assay was cross-reactive to the A1 ecotype especially at higher A1 *nifH* copy numbers; hence, our analyses likely provide a more accurate assessment of the temporal differences between the A1 and A2 ecotypes. For instance, an important trend in our time series was that A1–A4 ecotypes appeared together in 2015, while in 2016 and 2017 high UCYN-A1 *nifH* gene abundances displaced high UCYN-A2 *nifH* gene abundances (see events E1 and E2 in Fig. 3.3). In light of the high UCYN-A copy numbers expected for our study region (Mulholland et al., 2012; Tang et al., 2020) and the relatively higher growth rate for UCYN-A compared to more tropical/subtropical *Trichodesmium* (Martínez-Pérez et al., 2016), had we not improved upon the specificity of the UCYN-A2 qPCR assay and implemented less frequent sampling, these events may not have been as clearly defined. Overall, from the temporal data presented one can conclude that UCYN-A ecotypes (and hence their niches) may overlap depending on both yearly and weekly

dynamics. Our analyses therefore point to high-resolution sampling as being critical for understanding the complex *in situ* temporal niches of UCYN-A ecotypes. E1 and E2 were very short-lived events that occurred over a matter of 2–3 weeks (Fig. 3.3c). Hence, it is conceivable that similar fine scale patterns, if occurring elsewhere in the ocean, could explain why repeated sampling may uncover certain ecotypes in areas where they were initially considered absent. For example, A1 was recently found to dominate in the Southern California Current system (during May and October 2017) where prior sampling had initially detected only A2 in this coastal region (between 2010–2011; Thompson et al., 2014; Turk-Kubo et al., 2021).

In the Bedford Basin, temperature was the most predictive parameter explaining the occurrence of major UCYN-A ASVs (Fig. 3.4; Supplemental Table 3.2). Henke et al. (2018) also found that temperature was the most predictive variable for UCYN-A abundance; however, in their study UCYN-A1 was associated with lower temperatures (using qPCRs and oligotyping). Consequently, our results are more consistent with others who have instead proposed a link between UCYN-A2 and cooler waters (Henke et al., 2018; Turk-Kubo et al., 2017). It is important to note that, as Henke et al. (2018) indicate, a lack of overlap between temperatures across disparate studies presents a challenge for determining how the environment controls UCYN-A patterns within the ocean. For instance, during 2015–2017, surface temperatures at our study site never rose above 21°C (Fig. 3.4b). Thus, it is possible that the observations in the Bedford Basin captured dynamics within UCYN-A's lower temperature range instead of its upper temperature range; UCYN-A's dynamics at >22°C being more clearly captured by others such as Moisander et al. (2010) and Moore et al. (2018). Consistent with this view was our finding that UCYN-A (and especially UCYN-A1) in the Bedford Basin had a strong preference for temperatures closer to 10°C or above (Fig. 3.4) during the late summer period when DIN was also low. This temperature pattern is in agreement with recent work in the western Arctic Ocean, where highest UCYN-A *nifH* copies in the region were reported at ~10°C (Harding et al., 2018). Arctic UCYN-A1 nitrogen fixation rates at this temperature were also found to parallel rates in the subtropical North Atlantic (Harding et al., 2018).

3.5.3.2 Annual Cycles Reshape the Entire Community Composition of Rare UCYN-A ASVs

Our analysis of nucleotide differences between ASVs revealed that multiple rarer variants within each UCYN-A ecotype shared a close relationship to a central variant (Fig. 3.2). These central (or major) ASVs were found at higher relative abundances in our own samples and had 100% identity to the main A1–A4 *nifH* signatures recovered elsewhere (Turk-Kubo et al., 2017). We consider this feature (i.e., central ASVs with radiations of rarer ASVs; Fig. 3.2a) to be a key microevolutionary pattern within UCYN-A. Although rare ASVs had a total of two sequencing reads in our dataset, several lines of evidence lead us to conclude that the observed nucleotide differences represent true biological mutations in the natural community rather than experimental error (for e.g., errors from sequencing and/or PCR amplification). The first line of evidence is that the starburst patterns observed within our minimum spanning network was also observed with a larger independent dataset from the study of Turk-Kubo et al. (2017; Supplemental Fig. 3.14). The reproducibility of our analysis with a different dataset from the literature suggests trends were likely not the result of a lower sequencing depth or our analytical approach. The second line of evidence is that all ASVs of the A3 ecotype were effectively rare in the Bedford Basin despite the known biological relevance of the central A3 ASV; A3-6cd was identical in *nifH* sequence to both the A3 *nifH* described by Cornejo-Castillo et al. (2019) and to the reference Oligo2 (A3) *nifH* defined by Turk-Kubo et al. (2017). Although others have suggested that rare sequences are likely to contain spurious ASVs (Reitmeier et al., 2021), the case of A3 in the Bedford Basin implies that rare *nifH* ASVs are likely biologically relevant rather than derived from sequencing errors. Finally, the third line of evidence for a biological explanation hinges on the fact that the distribution of mutations we observed across the UCYN-A *nifH* ASVs (Fig. 3.2c) was consistent with mutations occurring in the *nifH* of other diazotrophs (Supplemental Fig. 3.18)—a pattern that is suggestive of mutational hotspots likely representing true interspecific biological variation within the *nifH* gene. Supplemental Discussion 3.S1 provides a more detailed explanation of the above three lines of evidence.

Given the prior context, we concluded that our *nifH* time series provides a unique perspective from which to further explore the microevolutionary patterns of UCYN-A. In

accordance to previous studies (Henke et al., 2018; Moreira-Coello et al., 2019; Turk-Kubo et al., 2017), we observed major and minor (i.e., rare) variants for multiple UCYN-A ecotypes that vary temporally. However, the high temporal resolution achieved in our study extends our understanding of co-occurring rare UCYN-A variants and provides insight into the type of selective pressure acting upon these variants compared to dominant UCYN-A variants. We found that the majority of rare ASVs were under neutral selection and did not reoccur annually at our study site (Supplemental Table 3.1 and Fig. 3.2). Although sampling location may play a role in determining the dominance of a variant within a given dataset (Turk-Kubo et al., 2017), the temporal patterns within the Bedford Basin nevertheless imply that at least some of the rare UCYN-A *nifH* intraspecific diversity is likely the result of neutral mutational variation within each ecotype. This feature is in direct contrast with mutations separating the major ASVs, which reoccurred annually and had numerous synonymous substitutions and zero nonsynonymous substitutions, implying that these ASVs are all under purifying selection (Yang & Bielawski, 2000). Previous Ka/Ks ratio calculated for >200 genes between A2 versus A1, and A2 versus A3, also revealed purifying selection between published genomes for aforesaid ecotypes (Cornejo-Castillo et al., 2016, 2019). The lack of annually reoccurring rare ASVs ultimately implies that seasonal cycling restructured the UCYN-A community composition especially for rare ASVs (Fig. 3.2). Given that a large fraction of rare UCYN-A ASVs exhibited neutral selection, and that rare diversity was reshaped annually, suggests that these two processes can be prominent microevolutionary features within naturally occurring populations of UCYN-A. In future, it will be important to also consider these features when attempting to further delineate UCYN-A diversity into any additional ecotypes. These results also emphasize the continued need for ecotype specific case studies (e.g., Cornejo-Castillo et al. (2019)) for confirming ecological and genomic differences across UCYN-A ecotypes.

3.6 Conclusions

By following UCYN-A ecotypes across a high-resolution 3-year weekly time series, we have taken significant steps towards filling current knowledge gaps regarding the coastal dynamics of UCYN-A. More specifically, our results and findings: (i) uncovered A1, A2, A3, and A4 ecotypes within the diazotrophic community of the Bedford Basin (a

coastal basin in the NWA) by delineating *nifH* sequences into ASVs (Callahan et al., 2017; Nearing et al., 2018) and showed that UCYN-A1 and A2 hosts were likely prymnesiophyte algae, consistent with other locations in the ocean (Cornejo-Castillo et al., 2019; Thompson et al., 2014); (ii) contributed to emerging observations of other phytoplankton that can also co-occur with UCYN-A within the marine environment (for e.g., *Synechococcus*); (iii) used a new UCYN-A2 qPCR assay with increased specificity to reveal that at a single coastal site the timing of UCYN-A blooms for A1 and A2 ecotypes could strongly vary both interannually and on weekly timescales; and finally, (iv) demonstrated that seasonal cycling could reshape the rarer fraction of UCYN-A's intraspecific diversity. Furthermore, this rare intraspecific diversity itself likely also contained many neutral mutational variants of the more dominant UCYN-A sequences within the environment. Overall, these results illustrate new complexities regarding UCYN-A's intraspecific diversity and the temporal patterns exhibited by its ecotypes, thus highlighting an ongoing need for high-resolution time-series datasets that measure this marine microbe. Without such data, it will inevitably remain challenging to interpret rarer *nifH* diversity within UCYN-A, to formulate more accurate distribution patterns for each UCYN-A ecotype, and to predict the *in situ* response of individual ecotypes to future environmental shifts originating from either anthropogenic perturbation at the coastal land-sea interface or from global climate change (Doney, 2010; Hutchins & Capone, 2022; Li et al., 2018; Nogales et al., 2011; Sohm et al., 2011; Tang, Wang, et al., 2019; Zehr, 2011).

3.7 Acknowledgements

The *nifH* molecular time series presented herein involved the dedicated participation of many individuals who helped with weekly sampling over the years (including I. Luddington, J-M. Ratten, C. Willis, J. Zorz, C. Mackie, and S. Rose). The Bedford Institute of Oceanography is especially thanked for the collection of the non-molecular ocean data that was presented herein and for welcoming our weekly participation in the Bedford Basin Monitoring Program. Likewise, the Ocean Frontier Institute technical pool is also thanked for assisting with weekly sampling. We also thank K. Turk-Kubo, who through personal communication provided more direct access to published oligotype sequences. This project was funded by the National Sciences and Engineering Research Council of Canada through

a Discovery grant awarded to J. LaRoche and a CGS-Doctoral Award received by B. Robicheau. The Ocean Frontier Institute and a Canada Foundation for Innovation Grant also provided funding for this project (awarded to J. LaRoche). B. Robicheau was further supported by a NS Graduate Scholarship and a Killam Predoctoral Award from Dalhousie University.

3.8 Author Contributions

B. Robicheau and J. LaRoche conceived the study. B. Robicheau conducted lab work, performed all data analyses, generated all initial data interpretations, and wrote the entire manuscript. J. Tolman assisted B. Robicheau with data processing and molecular lab work. D. Desai provided bioinformatics support. J. LaRoche additionally provided input that helped further refine the data interpretations and text.

The authorship order for a future journal submission of this Chapter will be: Brent M. Robicheau, Jennifer Tolman, Dhvani Desai, and Julie LaRoche.

3.9 Supplemental Information for Chapter 3

3.9.1 Supplemental Text for Chapter 3

3.9.1.1 Supplemental Methods 3.S1: Workflow for Retrieving UCYN-A ASVs from Universal *nifH* ASV dataset

Our workflow is based on the following key steps: (a) the generation of initial *nifH* ASVs, (b) the determination/calculation *in silico* of a minimum percent-identity threshold expected for all known UCYN-A ecotypes for a given *nifH* region, and then (c) the use of this minimum threshold to filter through a collection of globally-aligned BLAST hits (Johnson et al., 2008) that have initially been retrieved by locally aligning *nifH* ASVs to the published *nifH* of UCYN-A1 and UCYN-A2 genomes using a comparatively lower threshold. Reference *nifH* sequences are taken from ALOHA (Tripp et al., 2010) and SIO64986 (Bombar et al., 2014) UCYN-A strains. The final ASV lists generated (one from each procedure of aligning *nifH* ASVs to A1 and to A2) are then subsequently merged followed by manual inspection to remove duplicates, as well as nonsense and frameshift

mutations that are suggestive of pseudogenes. In this way, we effectively cast a wide net searching for ASVs that we suspect will include both UCYN-A and additional non-UCYN-A sequences (for e.g., other cyanobacteria), and then further narrow down this list by identifying those ASVs that indeed match to already identified ecotypes using a predetermined threshold (Moreno-Hagelsieb & Latimer, 2008). As mentioned in the main text, the general assumption is that known ecotypes of A1 & A2 should match each other and other ecotypes (A3, A4, etc.) in a reciprocal fashion. The minimum threshold used was 94.35% *nifH* similarity, with full step-by-step details provided in Supplemental Fig. 3.7. We therefore assume that ASVs falling below this threshold are unlikely to be UCYN-A. For context, this value is above the ~82% *nifH* similarity between UCYN-A and closely-related *Cyanothece* sp. ATCC 51142 [calculated using a MUSCLE alignment (Edgar, 2004)] and below the UCYN-A1 versus UCYN-A2 *16S* rRNA similarity of 99% (Bombar et al., 2014). Although we do recognize that the *nifH* similarity for UCYN-A8 versus UCYN-A1 and A2 presumably falls below our calculated threshold of 94.35%, a separate search for any UCYN-A8 *nifH* ASVs in our dataset found only BLAST hits with 79–91.6% similarity [at >90% sequence coverage] to the previously reported UCYN-A8 (oligo8) from (Henke et al., 2018; Turk-Kubo et al., 2017). Therefore, we do not suspect any UCYN-A8 ASVs were present within the Bedford Basin time series during the years of our study.

3.9.1.2 Supplemental Methods 3.S2: Additional Overview and Context for New UCYN-A2 qPCR Assay

As others have indicated (for e.g., Stenegren et al. (2018)), the cross-reactivity of the UCYN-A2 assay of Thompson et al. (2014) can depend on cycling conditions. Even with a higher annealing temperature of 64°C and a standard thermocycler speed (Stenegren et al., 2018), it has been our experience that the A2 assay remained cross-reactive towards 10⁷ A1 *nifH* gene copies (Supplemental Fig. 3.8a). We therefore designed a new modified forward primer extending the primer of Thompson et al. (2014) by one nucleotide, so that the 3' end of the primer differs from the A1 target (Supplemental Fig. 3.8b). One should note that our new assay does not address the known issue that the widely-used A2 qPCR assay is likely somewhat non-specific towards A3 and A4 ecotypes due to only having one

mismatch within the primers/probe set (Farnelid et al., 2016; Supplemental Fig. 3.8b). Due to this high sequence similarity, we did not test the cross-reactivity of the new A2 assay to A3/A4.

3.9.1.3 Supplemental Methods 3.S3: Additional R Packages and Settings used during Data Analyses

The following additional *R* packages (R Core Team, 2021) were used during data analyses/visualizations: *BiocManager* (Morgan, 2022), *reshape2* (Wickham, 2007), *metR* (Campitelli, 2021), *cmocean* (Thyng et al., 2016), *scales* (Wickham & Seidel, 2022), *cowplot* (Wilke, 2020), *tidyverse* (Wickham et al., 2019), *ggrepel* (Slowikowski, 2021), *stats* (R Core Team, 2021), and *ggdendro* (de Vries & Ripley, 2022).

For the *nifH* versus cyanobacterial + chloroplast *16S* rRNA network *CoNet* parameters (Faust & Raes, 2016) were set at: Matrix Type = count selected, Input Filtering = row_minocc selected and set at 20, standardization = col_norm selected with sum of filtered rows included, Methods = Pearson Correlation & Spearman Correlation & Mutual Information & Bray Curtis Dissimilarity & Kullback-Leibler Dissimilarity selected, Threshold Setting = 1000 top and bottom edgeNumber selected, Permutation = shuffle_rows with 100 iterations and renormalize setting selected (Faust et al., 2012), Bootstrapping = 100 iterations with p-values merged via “brown” (Brown, 1975) and unstable edges filtered, Multiple Test Correction = “banjaminihochberg” at P-value threshold of 0.05. All of these settings are the same as those recommended in the current *CoNet* 2017 tutorial for microbial communities (Website: https://psbweb05.psb.ugent.be/conet/microbialnetworks/conet_new.php; Last Accessed: 13-Jun-2022; (Faust & Raes, 2016)).

3.9.1.4 Supplemental Discussion 3.S1: Are Microevolutionary Trends Observed in Our Study Biologically Relevant (i.e., Originating due to Mutations Occurring in Nature) or Erroneous?

During our analyses, rare ASVs had at most two cumulative sequencing reads across all samples (Fig. 3.1a), thus calling into question their status as true biologically relevant mutations occurring in nature rather than potentially spurious or erroneous nucleotide

differences. Upon deeper examination, our datasets support the former inference (i.e., true biology) rather than the latter (i.e., error) as the best explanation. As a control for low sequencing depth, the ASV procedure, and our choice of molecular marker, we reran our minimum spanning analysis using the reference sequences described by Turk-Kubo et al. (2017); in this study, ASVs had ~100 reads or more, as opposed to the two reads per rare ASV that was typical in our dataset. This reanalysis showed that major trends (i.e., starburst patterns/radiations within each ecotype) were consistent across two different datasets (compare Fig. 3.2 versus Supplemental Fig. 3.14). Consequently, one cannot assume that the trends observed in said figures were simply the result of low sequencing depth or using an ASV versus oligotyping approach. Closer consideration of the UCYN-A3 ecotype provides another line of evidence supporting the biological origin of rare ASVs. Although the central ASV of the A3 ecotype in our study [i.e., ASV A3-6cd] was deemed ‘major’—due to our usage of the total reads per ASV across all samples to define rare versus major ASVs (Fig. 3.1a)—this ASV could actually be considered rare within the Bedford Basin as all five UCYN-A3 ASVs occupied less than 1% of the *nifH* community when present [equating to 1–3 reads when observed], hence in a practical sense all A3 ASVs were rare within the *nifH* community sequenced. Given this context, one can then turn to the question of biological relevance for these rare A3 ASVs. The relative abundance of ASV A3-6cd in the Bedford Basin was never above 0.898% during our study (Supplemental Fig. 3.11), yet it shares 100% similarity to the *nifH* gene reported in the 2019 study of the A3 ecotype [Accession Code: MH815013] (Cornejo-Castillo et al. 2019). Furthermore, ASV A3-6cd shares 100% similarity to Oligo2 (A3) from the study of Turk-Kubo et al. (2017), where the authors report that Oligo2 occupied slightly more reads within their total dataset than the dominant Oligo3 (A2) oligotype. Hence, although this ASV was rare at our study site, this is not always the case at other locations in the ocean (Turk-Kubo et al., 2017). Although erroneous sequences may be expected to appear in the rare fraction of community sequencing efforts (Reitmeier et al., 2021), the case of the A3 ecotype demonstrates that rare *nifH* ASVs might represent true mutational differences in the natural community. Others posit that rare ASVs may result from polymerase errors (Nearing et al., 2018). Our most informative analysis addressing this point is the distribution of mutations across translated NifH sequences; within A1, A2, and A4 ecotypes two mutational hotspots were

present at the end of the *nifH* region amplified by our selected primer set (Fig. 3.2c). By comparing mutational differences across 136 non-UCYN-A diazotrophs we observed that these two hotspots were shared amongst diazotrophs (Supplemental Fig. 3.18). Given the variety of methodologies likely used to acquire these 136 reference genomes, it seems improbable that the distribution of mutations within our own UCYN-A ASVs would correspond to *nifH* mutational hotspots in other diazotrophs had they mainly originated from polymerase amplification errors. Although some of the rare ASVs may still be generated via sequencing errors, the above points on interstudy comparisons, A3 similarity, and mutational distributions taken together suggest that (based on inference to the best explanation; Prosser et al., 2020) biology rather than errors shaped the evolutionary patterns and interspecific diversity detected for UCYN-A in the Bedford Basin.

3.9.2 Supplemental Tables for Chapter 3

Table 3.1. Z-test of selection for rarer UCYN-A ASVs. Probabilities for rejecting neutrality are shown, as well as the probabilities of rejecting strict neutrality in favour of either positive or purifying selection. Tests are completed in a pair-wise fashion for each non-dominant UCYN-A ASV versus dominant UCYN-A ASV within each ecotype. Analysis uses Nei-Gojobori (Proportion) model/method and assumes complete deletion for sites with gaps or missing data. Abbreviations: H_0 = null hypothesis, H_A = alternative hypothesis, d_S = No. of synonymous substitutions per synonymous sites, d_N = No. of nonsynonymous substitutions per nonsynonymous sites. Bold values are significant at $\alpha = 0.05$. Conducted in *MEGAI1* (Tamura et al., 2021)

Outgroup	ASV	Probabilities for each Z-test of Selection					
		Strict-Neutrality		Positive Selection		Purifying Selection	
		$H_0 = d_N = d_S$	$H_A = d_N \neq d_S$	$H_0 = d_N = d_S$	$H_A = d_N > d_S$	$H_0 = d_N = d_S$	$H_A = d_N < d_S$
	ASV-e442	0.1397		1.0000		0.0699	
	ASV-e1a7	0.5517		1.0000		0.2759	
	ASV-df34	0.3183		0.1592		1.0000	
	ASV-df9a	0.0169		1.0000		0.0084	
	ASV-d58d	0.0781		1.0000		0.0390	
	ASV-d38a	0.0239		1.0000		0.0120	
	ASV-d9c0	0.0063		1.0000		0.0031	
	ASV-cf79	0.3163		1.0000		0.1581	
	ASV-c4ad	0.0134		1.0000		0.0067	
	ASV-ab4a	0.1582		0.0791		1.0000	
	ASV-a2fa	0.0318		1.0000		0.0159	
A1-deb	ASV-5682	0.2543		1.0000		0.1272	
	ASV-5615	0.1582		0.0791		1.0000	
	ASV-3907	0.0071		1.0000		0.0035	
	ASV-3014	0.0094		1.0000		0.0047	
	ASV-681a	0.3163		1.0000		0.1581	
	ASV-625f	0.3163		1.0000		0.1581	
	ASV-91d8	0.3163		1.0000		0.1581	
	ASV-36ae	0.0007		1.0000		0.0004	
	ASV-29c7	0.3183		0.1592		1.0000	
	ASV-9eea	0.3618		1.0000		0.1809	
	ASV-3f0c	0.3183		0.1592		1.0000	
	ASV-2f0f	0.0520		1.0000		0.0260	
	ASV-2c2b	0.0014		1.0000		0.0007	
	ASV-1fd9	0.1582		0.0791		1.0000	
	ASV-1ca1	0.5239		1.0000		0.2619	
	ASV-1ae2	0.1845		1.0000		0.0923	

Outgroup	ASV	Probabilities for each Z-test of Selection					
		Strict-Neutrality		Positive Selection		Purifying Selection	
		$H_0 = d_N = d_S$	$H_A = d_N \neq d_S$	$H_0 = d_N = d_S$	$H_A = d_N > d_S$	$H_0 = d_N = d_S$	$H_A = d_N < d_S$
A1-deb	A1-e2f		0.0173		1.0000		0.0086
	A1-322		0.1582		0.0791		1.0000
	A1-14f		0.3163		1.0000		0.1581
A2-bdd	ASV-fbcf		0.3183		0.1592		1.0000
	ASV-fa43		0.1582		0.0791		1.0000
	ASV-e9f5		0.0014		1.0000		0.0007
	ASV-e1d6		0.3163		1.0000		0.1581
	ASV-d347		0.1250		1.0000		0.0625
	ASV-c786		0.0008		1.0000		0.0004
	ASV-c63c		0.5198		1.0000		0.2599
	ASV-be35		0.3163		1.0000		0.1581
	ASV-baf0		0.7450		1.0000		0.3725
	ASV-653b		0.5191		1.0000		0.2596
	ASV-110b		0.0078		1.0000		0.0039
	ASV-41ae		0.1034		1.0000		0.0517
	ASV-25b3		0.1843		1.0000		0.0922
	ASV-20d2		0.3163		1.0000		0.1581
	ASV-13e2		0.3163		1.0000		0.1581
	ASV-06ea		0.1034		1.0000		0.0517
	ASV-6e82		0.0049		1.0000		0.0024
ASV-3c55		0.2453		1.0000		0.1227	
ASV-3bf9		0.0091		1.0000		0.0046	
ASV-2ffc		0.0787		1.0000		0.0394	
A2-f9d		0.3183		0.1592		1.0000	
A2-57c		0.3183		0.1592		1.0000	
A3-6cd	ASV-bc03		0.0350		1.0000		0.0175
	ASV-b6a8		0.3183		0.1592		1.0000
	ASV-ac9c		0.1546		1.0000		0.0773
	ASV-60be		0.5198		1.0000		0.2599
A4-511	ASV-dffc		0.0080		1.0000		0.0040
	ASV-cbc5		0.3183		0.1592		1.0000
	ASV-49b2		0.3163		1.0000		0.1581
	ASV-9e26		0.1249		1.0000		0.0625
	ASV-0ca7		0.5191		1.0000		0.2596

Table 3.2. Significance of environmental predictors determined using the envfit function in *vegan* (Oksanen et al., 2022). Number of permutations used = 999.

	PC1	PC2	r^2	Pr(>r)
Temperature	0.99613	0.08794	0.5639	0.001 ***
NO ₃	-0.98852	0.15112	0.2594	0.001 ***
Si	-0.98411	0.17756	0.212	0.001 ***
PO ₄	-0.97762	0.21036	0.1912	0.001 ***
NO ₂	-0.97205	0.23479	0.1653	0.001 ***
NH ₄	-0.96515	0.26169	0.1639	0.001 ***
Salinity	0.99842	0.05612	0.0499	0.002 **
Chl <i>a</i>	0.8979	-0.4402	0.0407	0.002 **

Significance levels: *** 0; ** 0.001; * 0.010

3.9.3 Supplemental Figures for Chapter 3

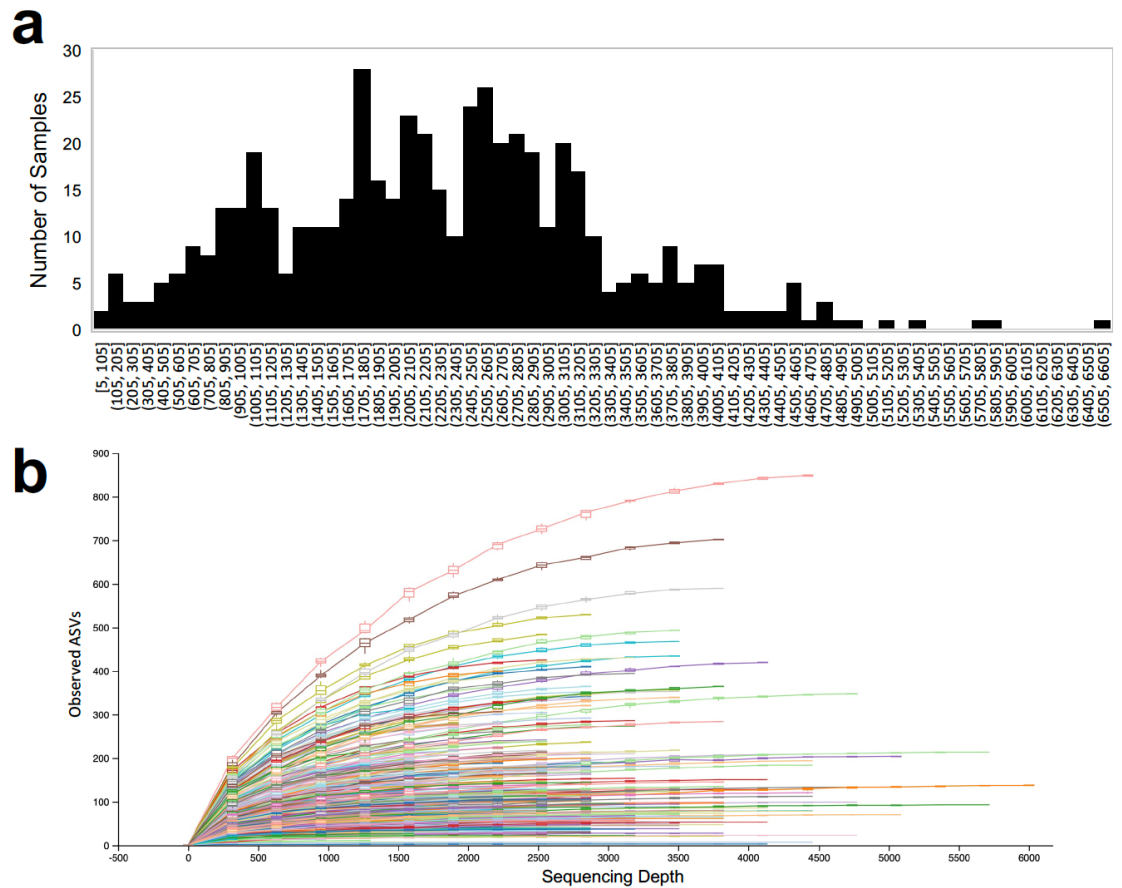


Figure 3.6 Summaries of raw sequencing data. Shown are **(a)** frequency distribution for final number of *nifH* reads per sample, and **(b)** rarefaction curves [individual samples colored] produced as part of the *QIIME2 View* (Bolyen et al., 2019) output.

(1) Complete QIIME2 Pipeline (Bolyen et al. 2019). **(2) Generate ASV table of all *nifH* ASVs per sample and a list of reference sequences for each of these ASVs**

Parameters of note:
 Trim Length = 325bp
 ASV denoising strategy = *deblur* (Amir et al. 2017)
 Read stitching strategy = *PEAR* (Zhang et al. 2014)

Generate *nifH* ASVs

(3) Determine Percent-Identity Threshold for Known Ecotypes:

3.1 Collect known reference ecotype sequences
3.2 Align reference ecotype sequences via MUSCLE (Edgar 2004)
Note: Sequences come from Turk-Kubo et al. (2017), Tripp et al. (2010), Bombar et al. (2014), Halm et al. (2012), Zehr et al. (2001), and Shiozaki et al. (2015)
3.3 Trim to same 325bp region as used in (1).
*Note: Some reference sequences do not have full length *nifH*, so are <325bp.*
3.4 Export matrix of Percent-Identity [Inset A]; colors are inter-ecotype comparisons.

Calibrate Search Strategy

Inset A. RAW Pairwise-identity DATA (%)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) A1 ALOHA Accession		100	95.97	96	94.44	95.06	95.16	96	96	95.97	95.16	94.76	97.18
(2) A1 oligo1	100		95.97	95.97	94.35	94.76	95.16	95.97	95.97	95.97	95.16	94.76	97.18
(3) A2 oligo3	95.97	95.97		100	95.16	95.56	95.97	96.77	96.77	96.77	95.16	94.76	97.18
(4) A2 SIO64986 Accession	96	95.97	100		95.68	96.3	95.97	97.23	97.23	96.77	95.16	94.76	97.18
(5) A3 HM210387	94.44	94.35	95.16	95.68		98.77	99.19	96.6	96.6	95.97	95.56	95.56	97.18
(6) A3 AF299420	95.06	94.76	95.56	96.3	98.77		99.6	97.22	97.22	96.37	95.97	95.97	97.58
(7) A3 oligo2	95.16	95.16	95.97	95.97	99.19	99.6		96.77	96.77	96.77	96.37	96.37	97.98
(8) A4 LC013598	96	95.97	96.77	97.23	96.6	97.22	96.77		100	100	98.39	97.98	97.98
(9) A4 LC013602	96	95.97	96.77	97.23	96.6	97.22	96.77	100		100	98.39	97.98	97.98
(10) A4 oligo4	95.97	95.97	96.77	96.77	95.97	96.37	96.77	100	100		98.39	97.98	97.98
(11) A5 oligo15	95.16	95.16	95.16	95.16	95.56	95.97	96.37	98.39	98.39	98.39		98.79	97.18
(12) A5 oligo20	94.76	94.76	94.76	94.76	95.56	95.97	96.37	97.98	97.98	97.98	98.79		96.77
(13) A6 oligo24	97.18	97.18	97.18	97.18	97.18	97.58	97.98	97.98	97.98	97.98	97.18	96.77	

3.5 Determine minimum and maximum pairwise-identity between ecotypes [Inset B & C]

Inset B. MINIMUM THRESHOLD							Inset C. MAXIMUM THRESHOLD						
	A1	A2	A3	A4	A5	A6		A1	A2	A3	A4	A5	A6
A1							A1						
A2	95.97						A2	96					
A3	94.35	95.16					A3	95.16	96.3				
A4	95.97	96.77	95.97				A4	96	97.23	97.22			
A5	94.76	94.76	95.56	97.98			A5	95.16	95.16	96.37	98.39		
A6	97.18	97.18	97.18	97.98	96.77		A6	97.18	97.18	97.98	97.98	97.18	

Minimum PI % = 94.35 Maximum PI % = 98.39

(4) Identify ASVs Similar to UCYN-A:

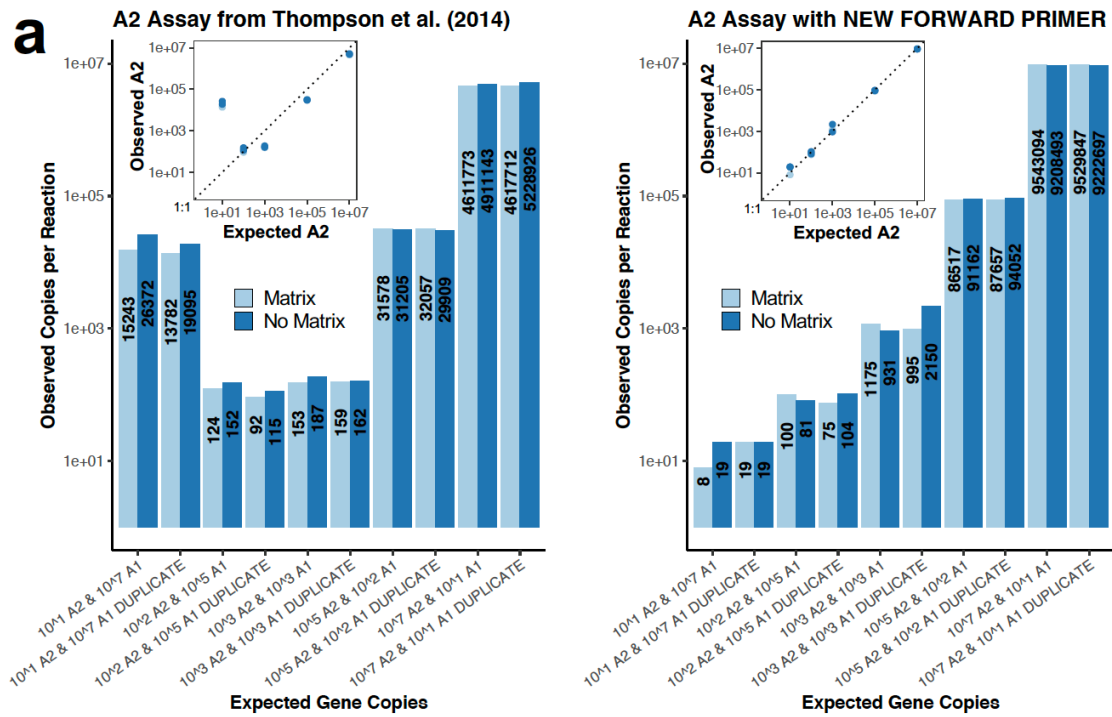
4.1 Locally align via BLAST (Johnson et al. 2008) the total ASV sequence collection of (2) to UCYN-A *nifH* from UCYN-A1 ALOHA and A2 SIO64986 genomes (Tripp et al. 2010; Bombar et al. 2014). Limit alignment search to 325bp target region above. MEGABLAST parameter selected and maximum hits = 300
Note: these settings were lax enough to return both similar and dissimilar matches.
4.2 Globally align via MUSCLE (Edgar 2004) all ASV sequences identified as hits in (4.1)
4.3 Export a pairwise-identity matrix for each search (one for ALOHA & one for SIO64986).
4.4 Use 94.35% threshold calculated in (3.5) to create list of UCYN-A like ASVs.
 There can be overlap between the two searches of (4.2), therefore remove duplicates.

(5) Filter UCYN-A ASVs into final list of ASVs:

5.1 Translate nucleotides and removing ASVs with nonsense mutations
5.2 Globally align ASVs to ALOHA and SIO64986 *nifH* references via MUSCLE (Edgar 2004) and remove ASVs with indels suggestive of frameshift mutations.
5.1 Use final list of UCYN-A ASV names to filter total *nifH* ASV table of step (2).

Retrieve UCYN-A ASVs

Figure 3.7. Workflow for retrieving UCYN-A ASVs from total *nifH* ASVs list.



b

A2 Assay	Forward Primer A2	Probe A2	Reverse Primer A2
Thompson	GGTTACAACAACGTTTTATGTTGAA...TCTGGTGGTCCTGAGCCCGGA..TGGATGTGCTGGTCGTGGT		
Robicheau	GGTTACAACAACGTTTTATGTTGAA..TCTGGTGGTCCTGAGCCCGGA..TGGATGTGCTGGTCGTGGT		
A1-deb	GGCTA T AACAACGTTTTATG C GTTGAG..TC C GGTGGTCCTGAGCC T GGA..TGGATGTGCTGGTCGTGGT		
A2-bdd	GGTTACAACAACGTTTTATGTTGAA..TCTGGTGGTCCTGAGCCCGGA..TGGATGTGCTGGTCGTGGT		
A3-6cd	GGTTACAACAACGTTTTATGTTGAA..TCTGGTGGTCCTGAGCC T GGA..TGGATGTGCTGGTCGTGGT		
A4-511	GGTTACAACAACGTTTTATGTTGAA..TCTGGTGGTCCTGAGCC T GGA..TGGATGTGCTGGTCGTGGT		

Figure 3.8. Modified UCYN-A2 quantitative PCR (qPCR) assay. **(a)** Cross-reactivity tests to determine if UCYN-A2 assay detects UCYN-A1. A gradient of A2:A1 standard DNA template is shown on the x-axis (ran in duplicate). Original Thompson et al. (2014) A2 qPCR assay shown on the left, A2 assay with modified forward primer (described herein) shown on right. **(b)** diagram showing nucleotide differences between the two assays (Thompson et al. (2014) = “Thompson”; Assay described herein = “Robicheau”) and major UCYN-A ASVs belonging to A1, A2, A3, and A4 ecotypes in our study. A box encases the A2 target sequence.

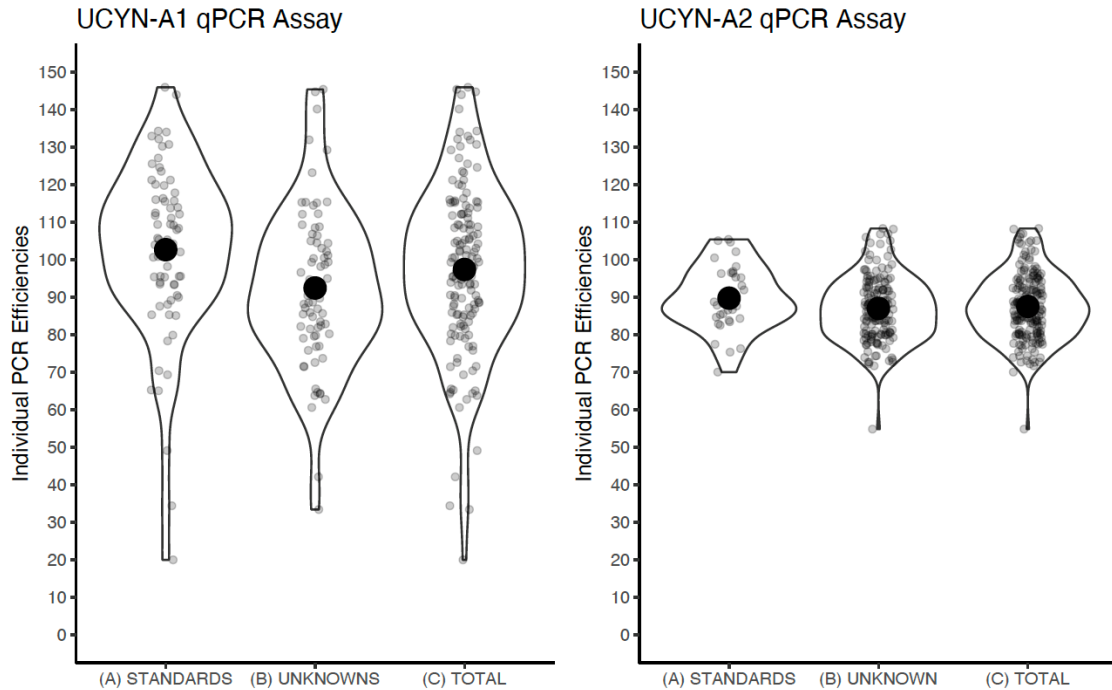


Figure 3.9. Distributions for final individual quantitative-PCR efficiencies for UCYN-A1 and UCYN-A2 assays (expressed as percentages). Mean values (black dots) are as follows: A1 Standards = 102.7%, A1 Unknowns = 92.5%, A1 Total = 97.4%, A2 Standards = 89.8%, A2 Unknowns = 87.0%, A2 Total = 87.6%.

>UCYN-A1-gBlock-standard

```
TTGGTTGTGACCCTAAAGCAGATTCCACACGTCTAATGTTGCATTGTAAAGCACAAACCACTGTT
TTACATTTAGCTGCAGAAAGAGGAAGTGTAGAAGATATTGAACTTGATGAAGTAGTAATTCCTGG
CTATAACAACGTTTTATGCGTTGAGTCCGGTGGTCCTGAGCCTGGAGTTGGATGTGCTGGTTCGTG
GTATTATTACTGCTATCAACTTCCCTTGAAGAAGAAGGTGCTTACGAAAACCTAGATTTTCGTATCT
TATGATGTATTAGGAGACGTTGTTTGTGGTGGTTTTCGCTATGCCTATCCGTGAAGGAAAAGCACA
AGAAATCTACATCGTTACTTCTGGTGAAATGATGGCAATGTACG
```

>UCYN-A2-gBlock-standard

```
TTGGTTGTGACCCTAAAGCAGATTCCACCCGTTTAAATGTTGCACTGTAAAGCACAAACCACTGTT
TTACATTTAGCTGCAGAAAGAGGAAGTGTAGAAGATATTGAACTTGACGAAGTAGTAATTCCTGG
TTACAACAACGTTTTATGTGTTGAATCTGGTGGTCCTGAGCCCGGAGTTGGATGTGCTGGTTCGTG
GTATTATTACTGCTATCAACTTTCTTGAAGAAGAAGGTGCTTACGAAAATCTAGATTTTCGTATCT
TACGATGTATTAGGAGACGTTGTTTGTGGTGGTTTTCGCTATGCCTATCCGTGAAGGAAAAGCACA
AGAAATCTACATCGTTACTTCTGGTGAAATGATGGCAATGTACG
```

Figure 3.10. References for UCYN-A1 and UCYN-A2 gBlock gene fragments (IDT) in FASTA format. Note that A1-gBlock-standard is identical to partial *nifH* of UCYN-A ALOHA (Accession: CP001842; Tripp et al., 2010) and A2-gBlock-standard is identical to partial *nifH* of A2 strain CPSB-1 (Accession: AP024987; Suzuki et al., 2021).

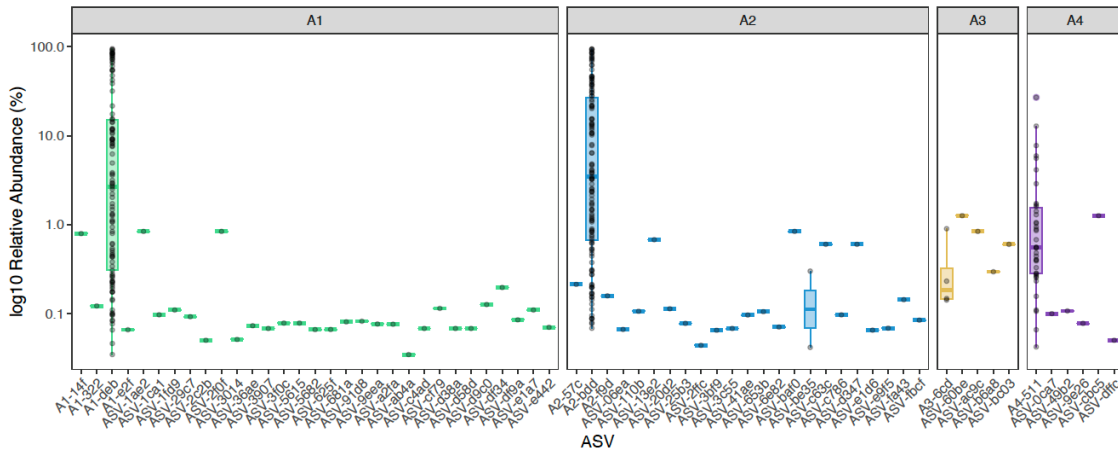


Figure 3.11. Relative abundance distributions for each UCYN-A ASV. Data are unrarefied. Ecotypes are individually colored (see ecotype names at top of panels).

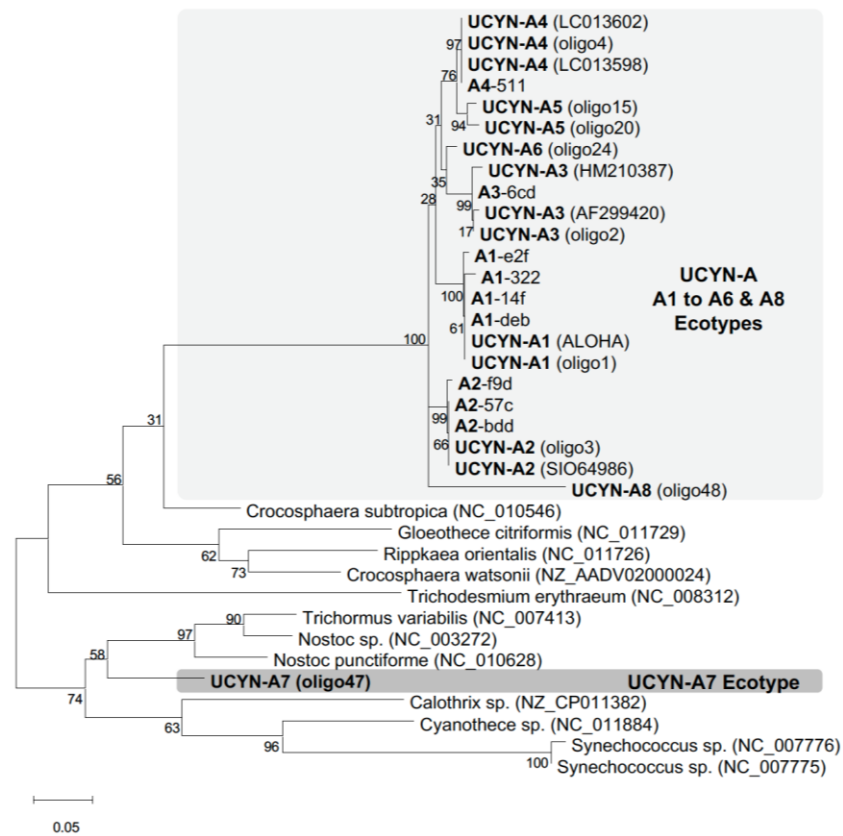


Figure 3.12. Maximum Likelihood tree of UCYN-A ASVs, UCYN-A oligotypes, and other cyanobacterial diazotrophs. Demonstrates placement of UCYN-A7 ecotype (lower highlighted box) outside of the main UCYN-A clade containing other ecotypes (upper highlighted box). Tree was constructed assuming a T92+G+I model and used complete deletion for sites with gaps. Bootstrap support (for 1000 replicates) shown at nodes. Codes in parentheses are either oligotypes from Turk-Kubo et al., (2017) or GenBank accession codes.

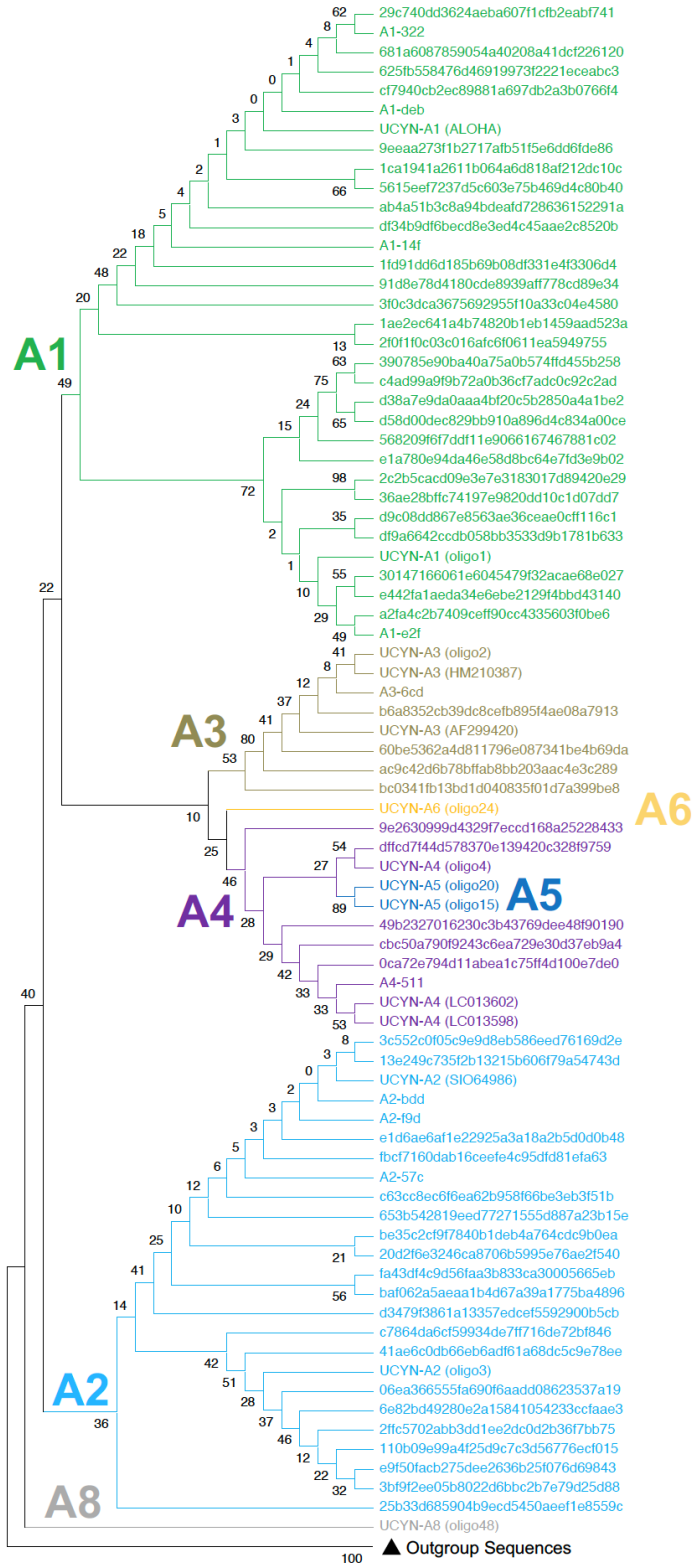


Figure 3.13. Neighbor-joining tree of UCYN-A phylotypes. Tree constructed using p-distances, pairwise deletion for gaps/ambiguous nucleotides, and bootstrap values (at nodes) using 1000 replicates. Alignment trimmed to 325bp *nifH* 3/4 amplicon. Outgroups are the same as in Fig. 3.1.

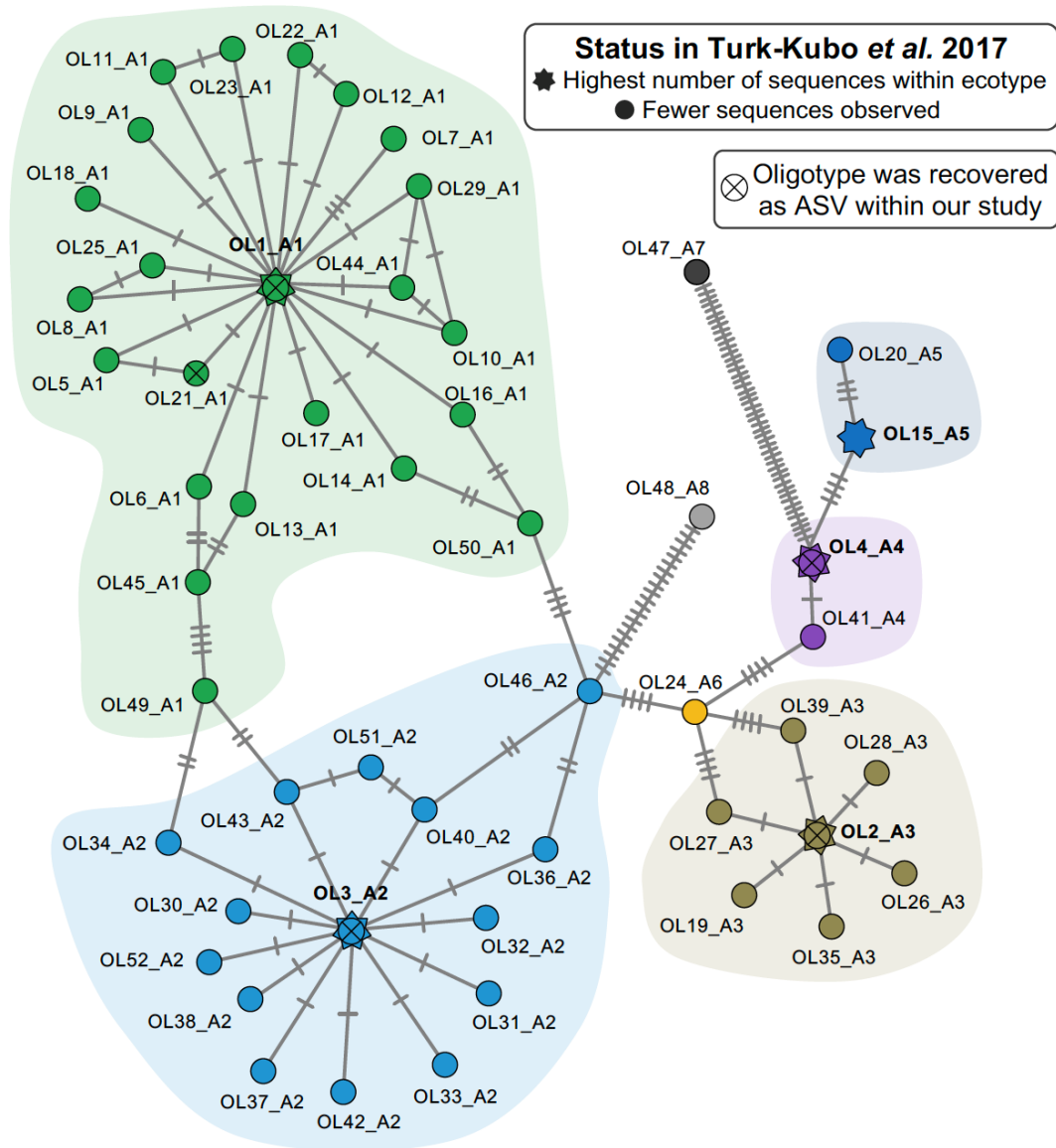


Figure 3.14. Minimum spanning nucleotide substitution network for fifty-two UCYN-A oligotypes identified by Turk-Kubo et al. (2017). Their study was based on using metabarcoding from a UCYN-A specific primer set and therefore the counts for rare reads are a lot higher (≥ 100 reads for A1-A6 ecotypes; Turk-Kubo et al., (2017)). Analysis uses 251 *nifH* sites in MUSCLE alignment. Note that for legibility, the original oligotype names used by Turk-Kubo et al. (2017) have been simplified to OL1–OL52, whereby OL = “Oligotype”. The oligotypes derived by Turk-Kubo et al. (2017) come from samples located across the globe (North Pacific, Southwest Pacific, and North and South Atlantic).

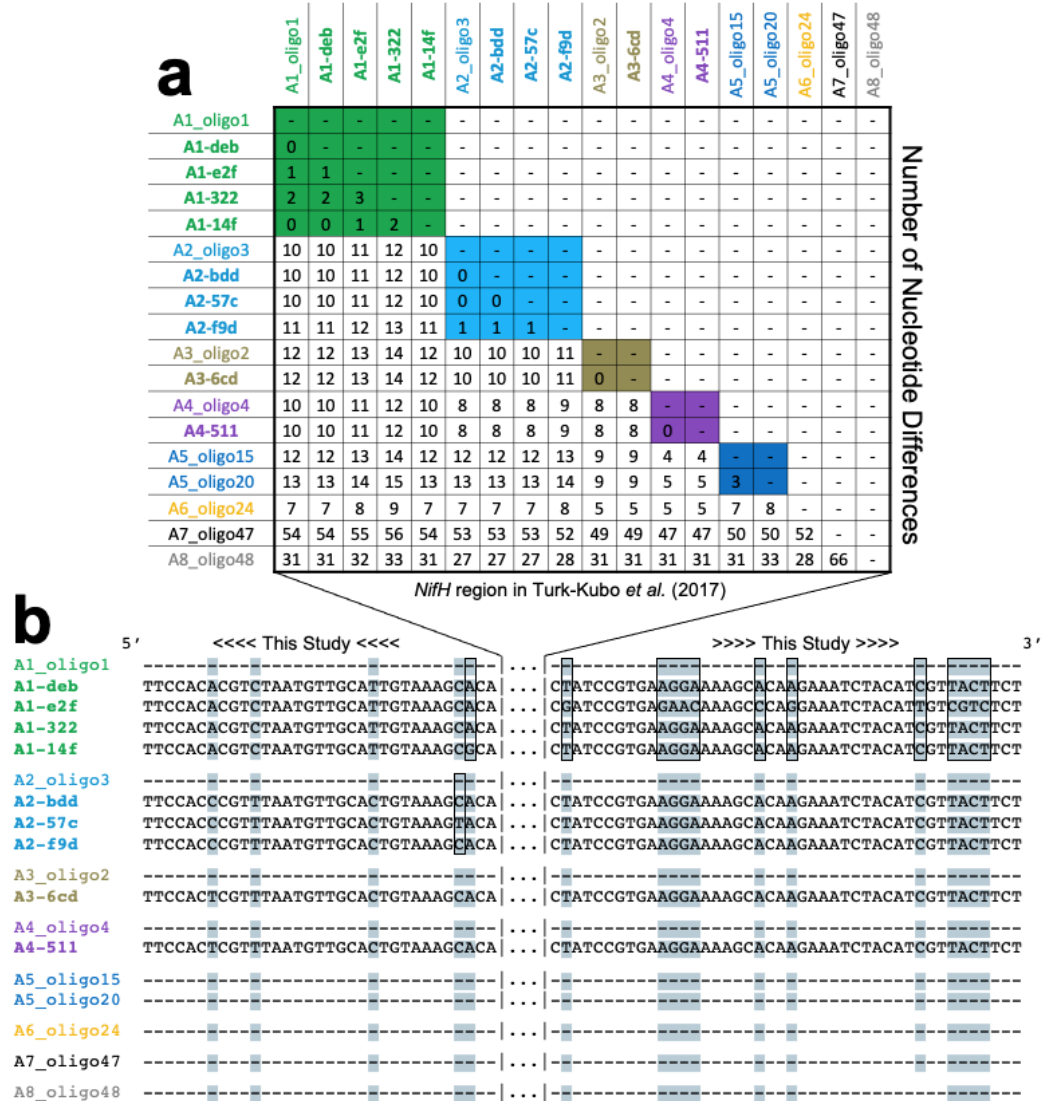


Figure 3.15. Nucleotide differences between UCYN-A ASVs reported herein (bolded) and UCYN-A oligotypes reported in Turk-Kubo et al. (2017). Alignment generated via MUSCLE. **(a)** Shows matrix of the number of nucleotide differences [intra-ecotype comparisons are colored] for the *nifH* region reported by Turk-Kubo et al. (2017), **(b)** shows nucleotide differences through visualization of the alignment itself where our ASV sequences extend beyond the 5' and 3' ends of the *nifH* region in Turk-Kubo et al. (2017) [non-conserved sites highlighted grey and intra-ecotype differences are encased by boxes].

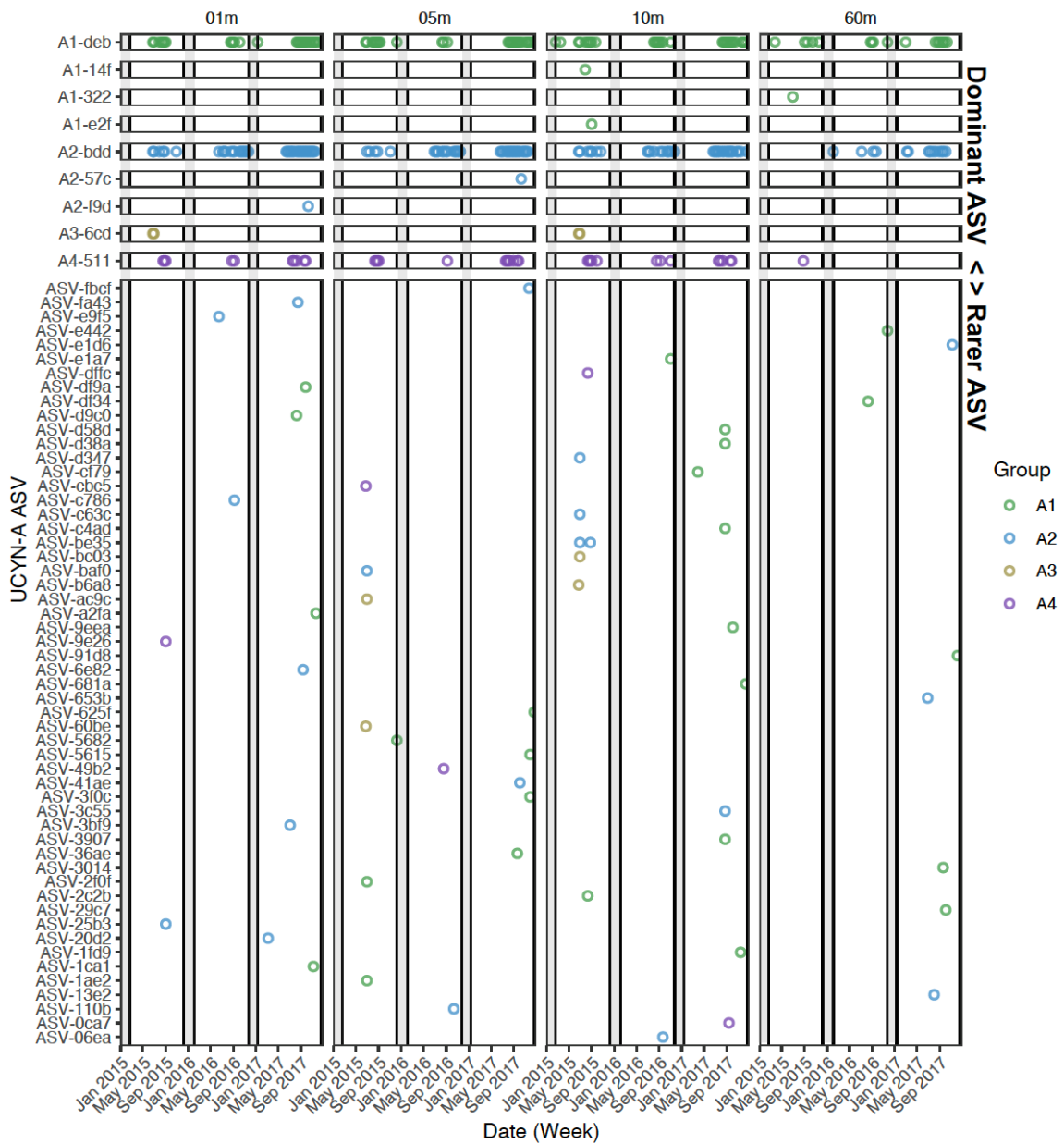


Figure 3.16. UCYN-A *nifH* ASVs present in Bedford Basin between 2015–2017. Raw reads were converted to presence/absence data. Note that absent values likely reflect the compositional nature of the data, however, these data can still be used to generally assess if rarer ASVs were present before, during, and/or after the appearance of dominant ASVs. Shaded vertical lines indicate times when dominant UCYN-A ASVs were not present in the dataset.

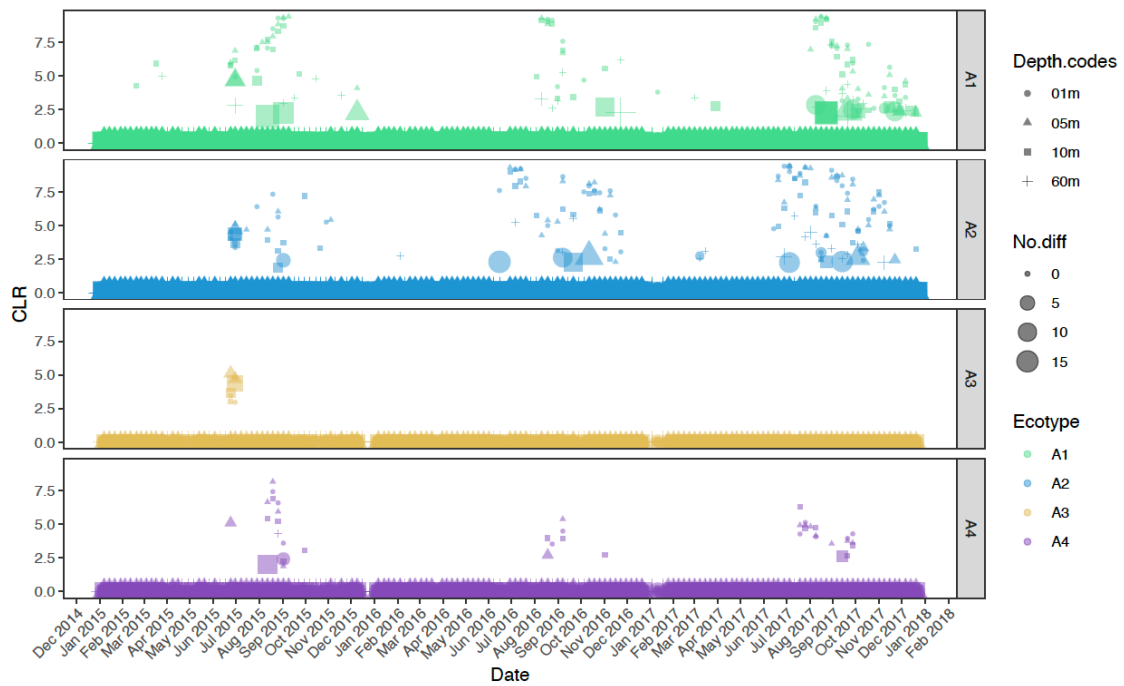


Figure 3.17. CLR values for each of the 65 UCYN-A ASVs over time along with the number of nucleotide differences (No.diff) each observation has relative to the major ASV within each group. Values where No.diff = 0 are observations of the major ASV within each group. For the purposes of this analysis major ASVs for each group were set to A1-deb for UCYN-A1, A2-bdd for UCYN-A2, A3-6cd for UCYN-A3, and A4-511 for UCYN-A4. CLR values of zero indicate that a given UCYN-A ASV was not detected in a sample.

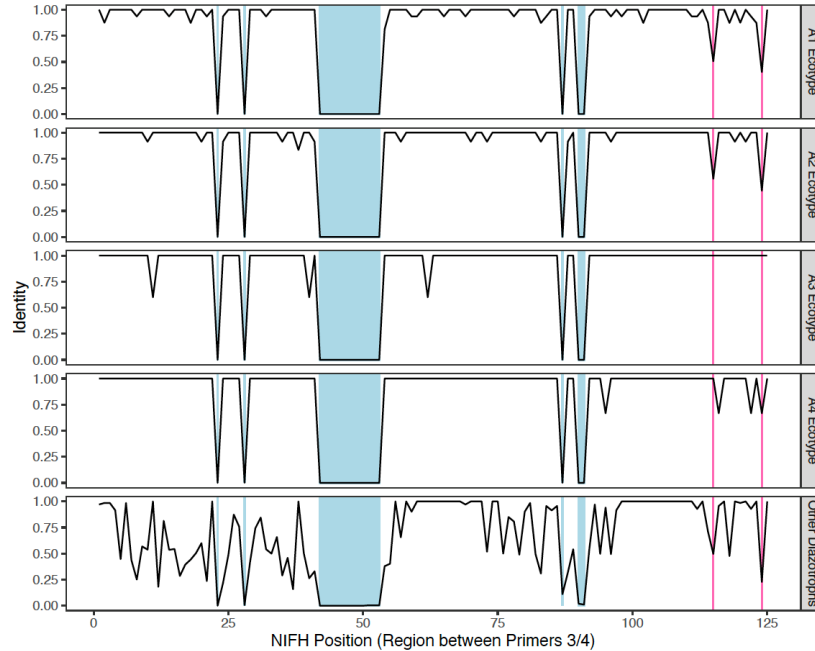


Figure 3.18. *NifH* amino acid site conservation within UCYN-A ecotypes relative to amino acid sites conserved within other non-UCYN-A diazotrophs. “Other Diazotrophs” are a collection of translated *nifH* amino acid sequences for 136 reference diazotroph genomes. Blue = regions with low identity scores due to large insertions/deletions between sequences in the alignment; pink = UCYN-A amino acid sites with <0.6 identity. GenBank accession codes for the 136 reference genomes used in analysis are: CP003046, NZ_FNOO01000067.1, NZ_JTHE01000274.1, NZ_JXXD01000121, NZ_MRDM01000012.1, NC_000916.1, NC_002678.2, NC_002932.3, NC_002936.3, NC_002939.5, NC_002977.6, NC_003030.1, NC_003272.1, NC_003552.1, NC_003901.1, NC_004463.1, NC_004547.2, NC_005090.1, NC_005296.1, NC_005791.1, NC_005863.1, NC_007298.1, NC_007355.1, NC_007413.1, NC_007493.2, NC_007498.2, NC_007512.1, NC_007514.1, NC_007517.1, NC_007775.1, NC_007776.1, NC_007778.1, NC_007952.1, NC_007958.1, NC_008312.1, NC_008576.1, NC_008639.1, NC_008702.1, NC_008741.1, NC_008781.1, NC_008789.1, NC_009012.1, NC_009049.1, NC_009253.1, NC_009337.1, NC_009428.1, NC_009434.1, NC_009437.1, NC_009445.1, NC_009464.1, NC_009483.1, NC_009485.1, NC_009523.1, NC_009617.1, NC_009633.1, NC_009635.1, NC_009637.1, NC_009675.1, NC_009706.1, NC_009712.1, NC_009767.1, NC_009921.1, NC_009937.1, NC_010337.2, NC_010424.1, NC_010524.1, NC_010546.1, NC_010628.1, NC_010794.1, NC_010814.1, NC_010831.1, NC_011026.1, NC_011027.1, NC_011060.1, NC_011145.1, NC_011146.1, NC_011206.1, NC_011420.2, NC_011666.1, NC_011726.1, NC_011729.1, NC_011761.1, NC_011768.1, NC_011769.1, NC_011830.1, NC_011832.1, NC_011884.1, NC_012796.1, NC_012881.1, NC_012997.1, NC_013173.1, NC_013194.1, NC_013216.1, NC_013851.1, NC_013943.1, NC_014410.1, NC_014500.1, NC_014664.1, NC_014762.1, NC_014973.1, NC_015216.1, NC_015416.1, NC_015709.1, NC_016616.1, NC_016629.1, NC_019757.1, NC_021149.1, NZ_AADV02000024.1, NZ_AA EW02000007.1, NZ_AA XW01000004.1, NZ_AP012549.1, NZ_ARWE01000001.1, NZ_CP007053.1, NZ_CP007215.2, NZ_CP009632.1, NZ_CP010803.1, NZ_CP010869.1, NZ_CP011382, NZ_CP011412.1, NZ_CP013021.1, NZ_CP023715.1, NZ_DF850488.1, NZ_JAFO01000001.1, NZ_JAGC01000009, NZ_JH109153.1, NZ_JH600070.1, NZ_JH993797.1, NZ_JOMG01000002.1, NZ_JPEO01000004.1, NZ_JRKM01000001.1, NZ_JTDI01000004.1, NZ_KB889967.1, NZ_KB905847.1, NZ_KE386569.1, NZ_KK214997.1, NZ_LYXA01000001.1

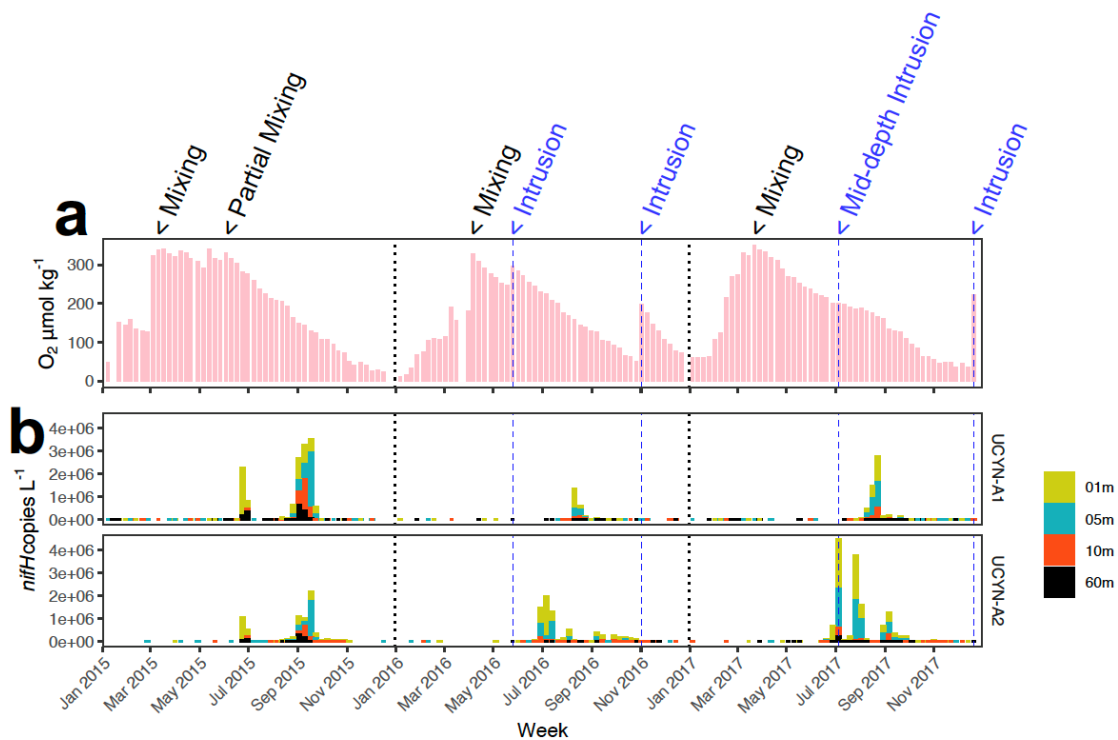


Figure 3.19. Shelf water 60m intrusion events during 2015–2017 Bedford Basin time series. Mixing and intrusion events (listed along the top of figure) have been identified and defined elsewhere by Haas et al. (2021) using dissolved oxygen measured at 60m. For comparative purposes, (a) the dissolved oxygen (O_2) data of Haas et al. (2021) and available at <https://doi.org/10.1594/PANGAEA.914705> has been replotted here relative to (b) our UCYN-A quantitative-PCR datasets for the A1 and A2 ecotypes.

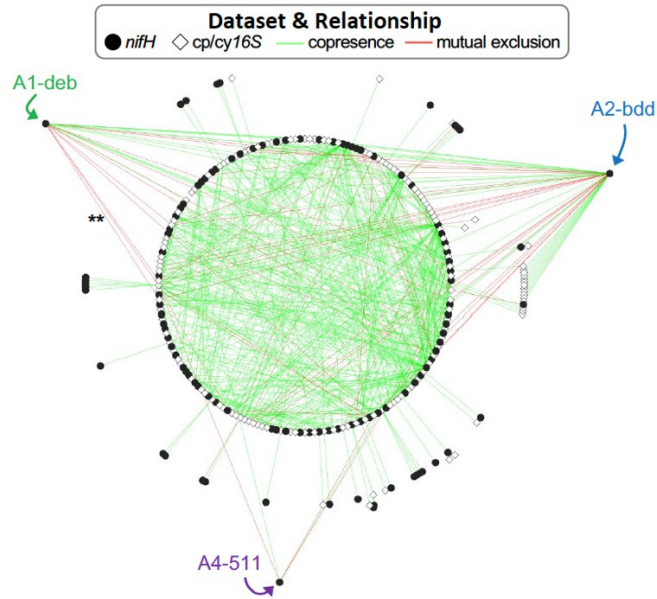


Figure 3.20. Complete network analysis of *nifH* versus phytoplankton *16S* rRNA (*cp/cy16S* rRNA gene). Figure in main text (Fig. 3.5) shows only relationships between UCYN-A and other phytoplankton but was based on this original network that includes all diazotrophs and phytoplankton with a minimum occurrence >2 across the entire dataset.

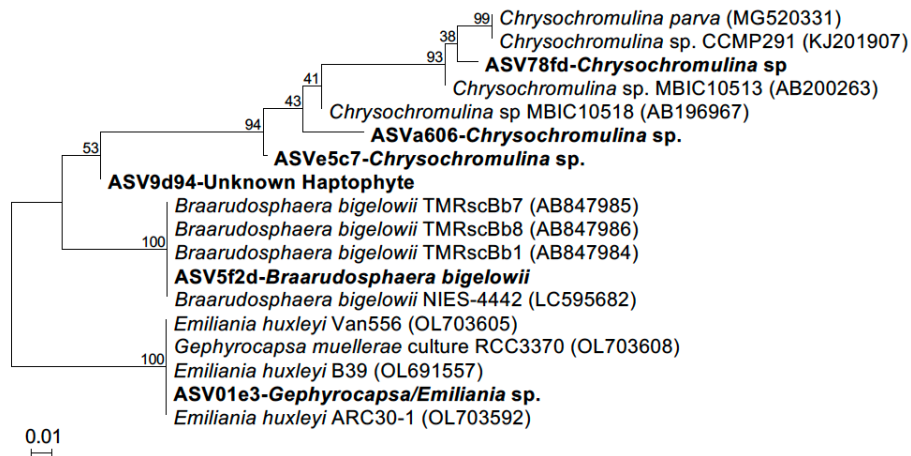


Figure 3.21. Maximum likelihood phylogeny of Haptophyta *cp16S* rRNA ASVs with strong and weak copresences to UCYN-A *nifH* ASVs. Accession codes for NCBI sequences with highest pair-wise similarity to ASVs are shown in parentheses. Complete deletion used to account for sites with missing data, tree assumes a K2+G model, and bootstrap values (1000 replicates) shown at nodes.

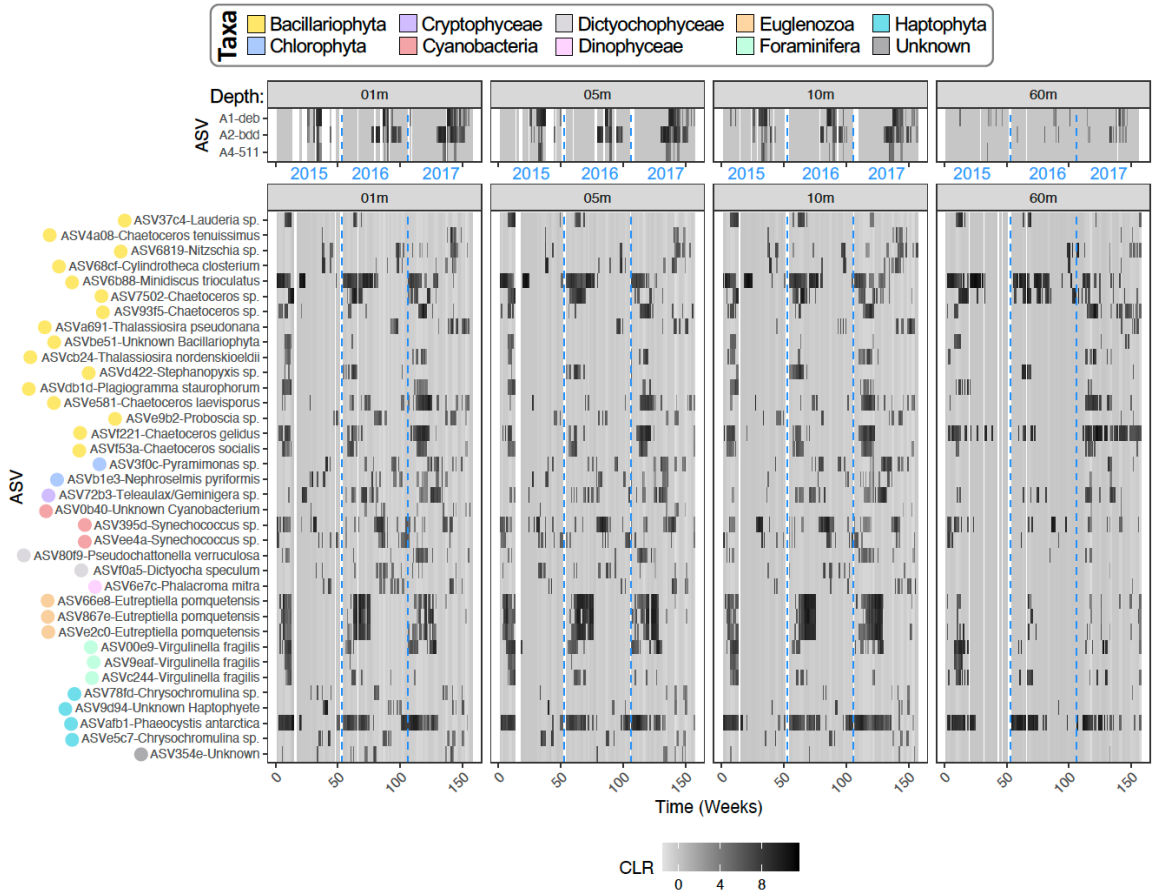


Figure 3.22. Spatiotemporal patterns for phytoplankton *16S* rRNA ASVs with weak copresence and mutual exclusions in relation to UCYN-A *nifH* ASVs. Sequencing data were CLR transformed.

CHAPTER 4

Marine Nitrogen-fixers in the Canadian Arctic Gateway are Dominated by Biogeographically Distinct Non-Cyanobacterial Communities

4.1 Abstract

Although marine diazotrophs occur globally, significantly less is known about their ecology and diversity within polar regions of the ocean. Herein we describe the diazotroph community patterns present during a GEOTRACES cruise through the Canadian Arctic Gateway (CAG) that occurred between July–August 2015. We identified differences in the dominant *nifH* signatures (i.e., variants) throughout the CAG. In the Labrador Sea, *nifH* variants corresponding to *Bradyrhizobium* sp. and *Vitreoscilla* sp. were dominant, while in Baffin Bay, a *nifH* variant from *Stutzerimonas stutzeri* was dominant. In comparison, the Canadian Arctic Archipelago (CAA) was characterized by a broader set of dominant *nifH* variants (for e.g., members of Desulfobulbaceae, Desulfuromonadales, *Arcobacter* sp., *Vibrio* spp., and *Sulfuriferula* sp. were detected). Although dominant diazotrophs fell within known non-cyanobacterial *nifH* Clusters I and III, the majority of these *nifH* variants were not identified in: (i) a 5-year weekly *nifH* time series that we established from the coastal NW Atlantic and (ii) a collection of *nifH* variants previously reported for the western Arctic Ocean (Shiozaki et al. 2018). Furthermore, the majority of dominant *nifH* variants in the CAG shared poor sequence similarity (<92% nucleotide pair-wise identities based on $\geq 95\%$ sequence coverage) to other *nifH* sequences present in a recently published global non-cyanobacterial diazotroph catalogue (compiled by Turk-Kubo et al. (2022)). We also provide a comprehensive assessment of UCYN-A abundances within the CAG, demonstrating that this species was widespread throughout the region, but major A1 and A2 ecotypes were only detected at very low copy numbers in quantitative-PCR assays (~ 30 *nifH* copies L⁻¹ or less). Finally, our analysis of all diazotrophs variants detected within the

CAG showed a stark division between the diazotrophic communities of the Labrador Sea and more northern Baffin Bay and CAA sites during sampling. Although more work will be needed to determine if this pattern reoccurs annually, our results establish that this previously unknown biogeographic community division can occur within the CAG. Hence, our study advances our understanding of non-cyanobacterial diazotrophs within the CAG by shedding light on their finer-scale community dynamics.

4.2 Introduction

The Arctic Ocean is undergoing major environmental change due to climate-related warming (Ardyna et al., 2014), yet this ocean and its nearby seas are among the most understudied oceanographic areas because of their remote location and relatively harsher climate (Boeuf et al., 2014). In recent years, considerable progress has been made in the elucidation of polar microbial communities (Deming, 2002; Dinasquet et al., 2018; Edwards et al., 2020; Zhang et al., 2020); for instance, Arctic marine bacteria experience strong seasonal changes (Kellogg et al., 2019) due to differing light and sea ice regimes between winter and summer seasons (Dinasquet et al., 2018), and bacteria in the Arctic Ocean have structured biogeography (for e.g., between surface and deep waters and from Eurasian to Canadian Arctic basins; Galand et al., 2009). Early cloning studies (e.g., Pommier et al. (2007)) have also shown that marine heterotrophic microbes in the Arctic Ocean include both cosmopolitan and polar-associated community members (summarized by Boeuf et al. (2014)). The degree of cosmopolitanism and endemism between taxonomic groups also applies to intraspecific microbial diversity, as seen for example with SAR11 phylotypes that are restricted to the Arctic Ocean (Kraemer et al., 2019).

In the spring, melting sea ice and increased solar irradiance initiate an Arctic phytoplankton bloom that contributes to polar primary productivity (Campbell et al., 2017; Comeau et al., 2011; Schuback et al., 2017). Nitrogen is generally the limiting nutrient in the Arctic Ocean, particularly for photosynthesis/primary production (Comeau et al., 2011; von Friesen & Riemann, 2020), with Fernandez-Mendez et al. (2016) recently showing that surface N:P ratios point to nitrogen limitation in the Central Arctic Ocean. Diazotrophs that can convert atmospheric dinitrogen (N_2) to ammonia via nitrogen fixation are an important microbial group that can overcome such limitation (Sohm et al., 2011; Tang,

Wang, et al., 2019); however, it is only recently that their presence within the Arctic Ocean has been fully recognized due to previously held assumptions regarding how inorganic nitrogen, temperature, and oxygen shape diazotroph biogeography in the ocean (Breitbarth et al., 2007; Fernández-Méndez et al., 2016; Gallon, 1992; Shiozaki et al., 2018; Sipler et al., 2017; Stal, 2017; Zehr & Capone, 2020, 2021a, 2021b).

Two main studies initially identified the presence of diazotrophs in the Arctic and showed that the dinitrogenase reductase gene (*nifH*) could be amplified in the Central Arctic Ocean (Damm et al., 2010) and that *nifH* community patterns could be recovered in the Arctic via next-generation sequencing (Farnelid et al., 2011). Interestingly, the latter work was a global survey of *nifH* and showed that Baffin Bay (within the Canadian Arctic Gateway) was an extreme outlier relative to other regions in the global ocean with respect to its diazotrophic community composition (Farnelid et al., 2011). No cyanobacterial diazotrophs were recovered from the Baffin Bay microbial DNA samples and no non-cyanobacterial phylotypes were shared with other sites sampled across the globe, hence suggesting likely arctic endemism for these diazotrophs (Farnelid et al., 2011). Heterotrophic diazotrophs (or non-cyanobacterial diazotrophs (NCDs), more precisely; Turk-Kubo et al., 2022) are now considered widespread (Bombar et al., 2016; Farnelid et al., 2011; Langlois et al., 2015; Riemann et al., 2010), with more recent surveys (e.g., *Tara* metatranscriptomes) continuing to confirm their dominance and *nifH* expression in the world's oceans (Farnelid et al., 2011; Salazar et al., 2019).

More recent studies support the consensus that diazotrophs from the Arctic Ocean mainly belong to NCDs of clusters I and III (Blais et al., 2012; Díez et al., 2012; Fernández-Méndez et al., 2016; Shiozaki et al., 2018), with notable exceptions including *Trichodesmium* detection in sea ice brine from Fram Strait near Greenland (Díez et al., 2012), and the observation of cluster IV diazotrophs in sea-ice and melt ponds near the Central Arctic Ocean (Fernández-Méndez et al., 2016). Although they do not represent a large proportion of the diazotrophic community, the symbiotic unicellular cyanobacterial diazotroph known as *Candidatus Atelocyanobacterium thalassa* (or UCYN-A) has been recently recovered within western Arctic waters of the Chukchi Sea (Harding et al., 2018; Shiozaki et al., 2018). UCYN-A exchanges fixed-N for fixed-C from its algal host (Martínez-Pérez et al., 2016) and consequently has lost the ability to fix carbon via

photosynthesis (Tripp et al., 2010). Single cell nitrogen fixation rates are also comparable between polar and non-polar UCYN-A (Harding et al., 2018; Zehr & Capone, 2021a). In addition to UCYN-A, *Richelia* and *Epithemia* diatom-diazotroph associations (Caputo et al., 2019; Schvarcz et al., 2022), as well as arctic-associated ultrasmall (less than 0.22 micron) *Arcobacter* (Karlusich et al., 2021) have also recently been found at higher latitudes, hence providing even further evidence that diazotrophs are part of the Arctic marine microbiome. Measurable nitrogen fixation rates range from 0.02 nmol N L⁻¹ day⁻¹ in Baffin Bay (Blais et al., 2012) to 17.2 nmol N L⁻¹ day⁻¹ in the coastal Chukchi Sea (Sipler et al., 2017; also see Shiozaki et al. (2018)). Sipler et al. (2017) estimated that Arctic shelves alone (during ice-free periods) could account for as much as 2.7% of global nitrogen fixation. Although effects may also be regionally specific (von Friesen & Riemann, 2020), the significant nitrogen fixation rates therefore indicate that although diazotrophs represent a low percentage of bacterioplankton in the Arctic (Karlusich et al., 2021; Salazar et al., 2019) the input of new nitrogen from these microbes may still be important.

Although diazotrophs are now considered broadly distributed at polar latitudes, their geographic distributions and diversity across various regions of the Arctic Ocean are underexplored (Karlusich et al., 2021; Meiler et al., 2022; Shiozaki et al., 2017; von Friesen & Riemann, 2020). Herein, we aim to help bridge this critical knowledge gap by further investigating diazotroph diversity based on *nifH* amplicon sequencing from the Labrador Sea into Baffin Bay and onwards towards the Canadian Arctic Archipelago (CAA). Specifically, we identify dominant diazotrophs present within this Arctic region, along with their correlations to ocean conditions such as depth, temperature, size fraction, oxygen levels, and selected macro and micronutrients. Given recent reports of UCYN-A in the Arctic Ocean (Harding et al., 2018; Shiozaki et al., 2018), we further conducted quantitative PCR assays of UCYN-A ecotypes (A1 and A2) to assess their presence within eastern Canadian Arctic waters. Since dominant Arctic phylotypes have previously been found to be generally less similar from those of other oceans (Turk-Kubo et al., 2022), we also sought to further establish any degree of arctic endemism for diazotrophs identified as important in the region. This was accomplished by aiming to identify *nifH* amplicon sequence variants (ASVs) that were dominant in the Canadian Arctic Gateway from within

a 5-year weekly *nifH* time series that we have established in a temperate NW Atlantic fjord (Bedford Basin, NS, Canada). Furthermore, we compared dominant *nifH* reference sequences identified herein to more global *nifH* sequence sets (from Delmont et al. (2021) and Turk-Kubo et al. (2022)), as well as to known sequences from the western Arctic Ocean (Shiozaki et al., 2018) to further identify any possible Arctic endemism. Overall, it is critical to elucidate the diversity and ecology of diazotrophs within the Arctic sector, given that this fraction of the ocean microbiome can contribute to Arctic Ocean primary productivity via the generation of new fixed-nitrogen (von Friesen & Riemann, 2020).

4.3 Methods

4.3.1 Cruise Samples and their Environmental Data

Water samples were collected via Niskin bottles aboard the *CCGS Amundsen* between July-10-2015 to August-20-2015 during the ArcticNet1502 (GN02) 2015 cruise to the Arctic. This expedition was part of an ArcticNet and Canadian Arctic GEOTRACES joint effort (Anderson et al., 2014) to study the Labrador Sea and Northwestern Passages. DNA was collected at eleven stations (Fig. 4.1) that were spread throughout the Labrador Sea (K1 and LS2), Baffin Bay (BB1, BB2, and BB3), and Canadian Arctic Archipelago (CAA; CAA1–2 and CAA4–7). At each of these stations, approximately 4 L of seawater [average = 3.93 L] was subjected to peristaltic filtration to collect biomass onto 3 μ m and 0.2 μ m Isopore polycarbonate membrane filters (Millipore, Ireland) in series such that the large fraction represents $\geq 3\mu$ m and the small fraction 0.2–3 μ m. DNA filters were immediately frozen and kept at -80°C until further laboratory processing.

Oceanographic data collected during the GN02 cruise were both CTD sensor-derived [for temperature, salinity, oxygen, and fluorescence] and bottle-derived [for nitrate, nitrite, phosphate, silicate and dissolved and particulate trace metals], and are publicly available as part of the GEOTRACES Intermediate Data Product 2021 Version 1 dataset (IDP2021; GEOTRACES Intermediate Data Product Group, 2021). Additional dissolved iron and manganese data can be accessed from the supplemental data reported in Colombo et al. (2020) and additional total particulate iron, manganese, vanadium, and phosphorus from the supplemental data of Colombo et al. (2021 and 2022). Due to the tight water budget on the GEOTRACES cruises, DNA samples and other bottle-derived oceanographic data were

occasionally collected on different CTD casts at a given station. Therefore, nutrient and trace metal data were matched to the closest depth and time at which DNA samples were collected at each station.

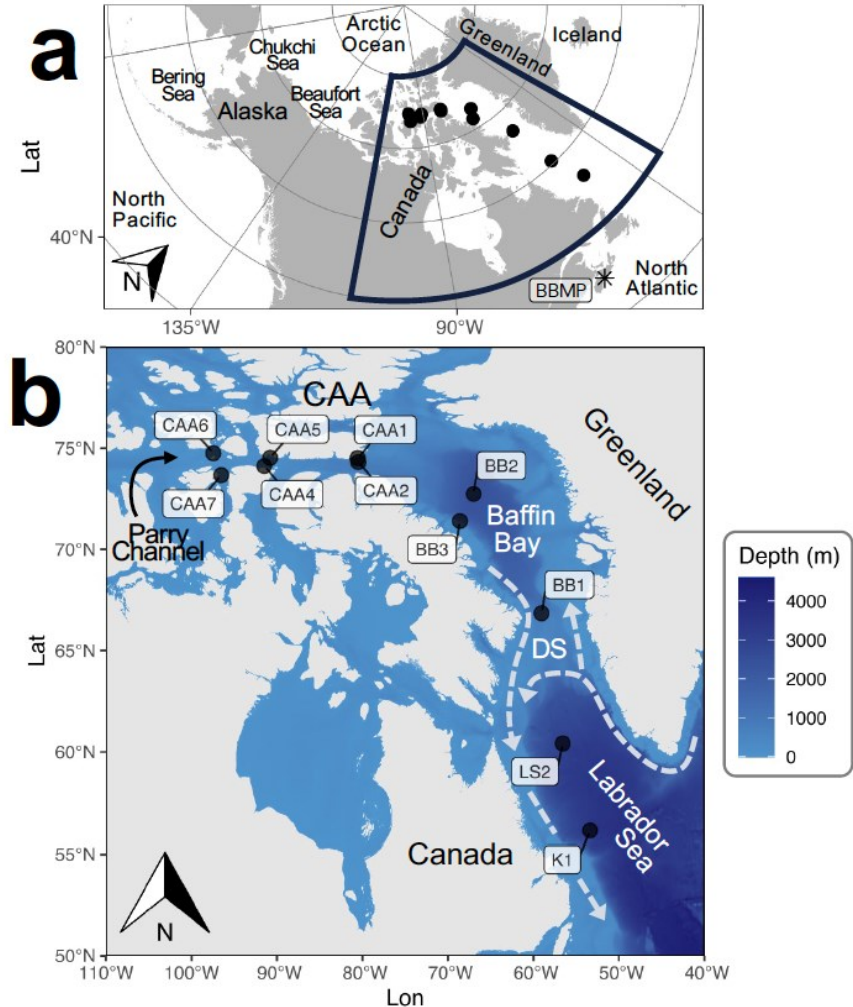


Figure 4.1. Sites sampled along the Canadian Arctic Gateway during the ArcticNet1502 (GN02) GEOTRACES expedition. Samples are from July–August 2015. **(a)** Arctic Ocean region covered by *nifH* sampling during our study and the location of the weekly Bedford Basin Monitoring Program (BBMP; asterisk), dark box indicates borders of map shown in panel b. **(b)** Detailed map of study sites with bathymetry (background color) from *ETOPO1* dataset (Amante & Eakins, 2009; NOAA National Geophysical Data Center, 2009; Simons, 2019). Light grey arrows depict the main circulation through and near the Davis Strait (DS) with Arctic outflows on the western side and inflow from the West Greenland and Irminger Currents on the eastern side, as well as the southern flowing Labrador Sea Current further south along the western side of Labrador Sea (after Colombo et al. (2020), Frago et al. (2016), and Lehmann et al. (2022)). CAA stands for Canadian Arctic Archipelago. Note that water flows mainly from west (Canadian Basin/Beaufort Sea) to east (towards Baffin Bay) along Parry Channel especially on the southern side of the channel (see Lehmann et al. (2022)).

4.3.2 Coastal NW Atlantic Time-series Samples and their Environmental Data

Intermittently during 2014 [once in each of March, June, September, and December; 10 samples total] and continuously each week for five years between Jan 2015 to December 2019, we collected DNA samples within the coastal NW Atlantic (in Bedford Basin at site BBMP in Fig. 4.1a—this location [44.6936 LAT, -63.6403 LON] is within a fjord comprising the Halifax harbour (Nova Scotia, Canada)). Seawater samples from 1, 5, and 10m (surface) and 60m (bottom) were collected at a compass buoy located at the BBMP site using Niskin bottles. Biomass from 500 mL was filtered onto Isopore polycarbonate filters (Millipore, Ireland) using a peristaltic pump such that size ranges were: 0.2–160 μ m (i.e., $\geq 0.2\mu$ m) for 2014–2015, 0.2–330 μ m (i.e., $\geq 0.2\mu$ m) for 2016–2017, and 0.2–3 μ m (small fraction) and $\geq 3\mu$ m (large fraction) for 2018–2019. Filters were immediately frozen at -80°C after each collection.

Measurements for temperature, salinity, and oxygen (all CTD sensor-derived) and nitrate (Niskin-derived) from the Bedford Basin are accessible by request via the Bedford Institute of Oceanography (BIO) website: www.bio.gc.ca/science/monitoring-monitorage/index-en.php. These oceanographic data from the Bedford Basin are collected each week as part of the BIO Bedford Basin Monitoring Program (Li & Dickie, 2001; Li et al., 2006).

4.3.3 DNA Extractions, *nifH* Amplicon Sequencing, and UCYN-A Quantitative-PCRs

All DNA filters (cruise and time series) were processed using the same procedure. DNA was extracted [50 μ L final volume] using a DNeasy Plant Mini kit (Qiagen; Zorz et al., 2019). A nested PCR method was used to amplify the *nifH* diazotroph marker gene from each sample using the *nifH3/nifH4* [ATRTTRTTNGCNGCRTA/TTYTAYGGNAARGGNGG] and *nifH1/nifH2* [TGYGAYCCNAARGCNGA/ADNGCCATCATYTCNCC] primer pairs from Zehr & McReynolds (1989) and Zani et al. (2000). PCRs [PCR 1 (25 μ L) | PCR 2 (10 μ L)] were carried out with: 10 \times Buffer (Qiagen) [2.5 μ L | 1 μ L], 25mM MgCl₂ (Qiagen) [4 μ L | 1.2 μ L], 20mg/mL BSA (NEB) [0.3 μ L | 0.12 μ L], molecular biology grade H₂O (Invitrogen) [9.725 μ L | 4.29 μ L], HotStar *Taq* (Qiagen) [0.125 μ L | 0.05 μ L], template DNA [2.5 μ L of extracted DNA | 1 μ L of

nifH3/4 PCR product], and 10nM dNTPs (Invitrogen) and 10 μ M *nifH* primers (IDT)[2 μ L each | 0.8 μ L each]. Thermocycler settings were: [PCR 1] 95°C–15 min, 35 cycles of 95°C–1 min, 45°C–1 min, and 72°C–1 min, then 75°C–10 min; PCR 2 increased the annealing temperature to 54°C and decreased the number of cycles to 28. DNA amplicon products of 359bp detected in second-round PCRs via agarose gel electrophoresis were further processed for Illumina Next-generation sequencing of the *nifH* gene. Briefly, this involved repeating the nifH3/4 PCR 1 (PCR1d) with 1/10 diluted DNA, and then combining equal amounts of the products from PCR 1 and PCR1d as template for a modified PCR2 (PCR2f) with *nifH* fusion primers (Ratten, 2017) that combined Illumina adaptors and barcodes with nifH1/nifH2. Thermocycler settings were as above with a 52°C annealing temperature and 35 cycles of amplification. Barcoded products were normalized and purified using a Just-A-Plate 96 kit (Charm Biotech). The final *nifH* library pool was sequenced on an Illumina MiSeq instrument (Integrated Microbiome Resource, Dalhousie University, Halifax, NS, Canada). *NifH* sequencing data are deposited under NCBI Bioprojects PRJNA930772 for the Bedford Basin time series and PRJNA931255 for the Arctic GEOTRACES 2015 expedition (NCBI Resource Coordinators, 2018).

To enumerate UCYN-A, we used the quantitative-PCR (qPCR) assay of Langlois et al. (2008) for the UCYN-A1 ecotype: 8 μ L TaqMan Universal PCR master mix (2 \times ; Applied Biosystems), 0.16 μ L TaqMan MGB 6-FAM probe (10 μ M; Applied Biosystems), 0.144 μ L of each forward and reverse primer (100 μ M; IDT), 0.32 μ L BSA (20 mg/mL; NEB), 2.232 μ L of molecular biology grade H₂O (Invitrogen), and 5 μ L DNA (0.56 \times), with thermocycler settings of: 95°C–10 min then 45 cycles of 95°C–15 sec and 60°C–1 min. For the UCYN-A2 ecotype we developed a new assay modified from Thompson et al. (2014): 8 μ L TaqMan (2 \times), 0.32 μ L probe 5'-FAM-TCTGGTGGTCCTGAGCCCGGANFQ-3' (10 μ M), 0.4 μ L each of forward primer 5'-GGTTACAACAACGTTTTATGTGTTGAA-3' and reverse primer 5'-ACCACGACCAGCACATCCA-3' (both stocks at 100 μ M), 0.32 μ L BSA (20 mg/mL), 1.56 μ L molecular biology grade H₂O, and 5 μ L DNA (0.56 \times), with thermocycler settings of 95°C–10 min then 45 cycles of 95°C–15 sec and 64°C–1 min. Additional details for the qPCR assay of UCYN-A2 can be found in Chapter 3 of this dissertation. Both assays were run on a ViiA7 real-time PCR system with QuantStudio software (Applied Biosystems) to

assess the final number of gene copies per reaction. Assuming 4L of filtered seawater per DNA sample, our limit of quantification (LOQ) for both assays was 5 copies L⁻¹, and average qPCR efficiencies were 106% ± 24% SD for A1 assay and 91% ± 10% SD for A2 assay (calculated using *LinRegPCR* (Ramakers et al., 2003; Ruijter et al., 2009)). A Qubit 4 fluorometer and 1× HS dsDNA kit (Invitrogen) were used to quantify gBlock gene fragments of the *nifH* gene (IDT) for generating A1 and A2 standard curves (standards were run in triplicate with PCR-grade water (Invitrogen) used as a non-template control).

4.3.4 Data Analyses

4.3.4.1 Amplicon Sequence Variant (ASV) Calling

ASVs were generated from *nifH* sequencing data for both the Arctic Cruise and Bedford Basin time series using *QIIME 2* version 2019.7 (Bolyen et al., 2019) following a workflow modified from the Microbiome Helper pipeline of Comeau et al. (2017). For primer removal, the *nifH*1/2 primers were used with *cutadapt* (Martin, 2011). Reads were denoised into ASVs using *deblur* (Amir et al., 2017) with a *nifH* reference set (Gaby & Buckley, 2014) and a consistent trim-length of 325bp. ASVs with a frequency of <0.1% of the mean sample depth were removed due to assumed bleed-through between sequencing runs. Final ASV tables were converted to percent relative abundances [calculated as: (# reads per ASV in sample ÷ total reads in sample) * 100] and to centered-log ratio (CLR) values using *phyloseq* and *microbiome R* packages (Gloor et al., 2017; Lahti et al., 2019; McMurdie & Holmes, 2013; R Core Team, 2021). The final mean sampling depth per sample was 2,504 ± 907 SD reads.

4.3.4.2 Taxonomic Identifications

Diazotroph taxonomies were assigned by placing their reference sequences into the phylogeny described by Kapili & Dekas (2021). This was accomplished using: *hmmer* version ≥ 3.1 (see hmmer.org; Eddy, 2011) for the alignment of sequences, *EPA-ng* (Barbera et al., 2019) for the placement of sequences into the Kapili & Dekas (2021) phylogeny, and *gappa* (Czech et al., 2020) to assign the final taxonomy based on the tree placements. The implementation of this pipeline required the use of *EMBOSS* software package (Barbera et al., 2019). To ensure the Kapili & Dekas (2021) sequences ($n = 8,877$)

would encompass currently known diazotroph genomes, we also subjected reference *nifH* genes ($n = 3,490$) from the genome taxonomy database (GTDB; Chaumeil et al., 2022) to the above pipeline. Out of this GTDB set, 613 sequences were not identified correctly based on their known taxonomy, while 977 sequences with known taxonomy were not placed in the tree, while the remaining 1,900 sequences were correctly assigned. To account for the possibility that our Arctic ASVs might belong to the prior two groups, we used *BLAST* (Johnson et al., 2008) to align our Arctic *nifH* sequences to the above misidentified/unclassified reference sequences; seven ASVs were renamed as a result, six of which were rare, and the seventh (Pseudomonadales) was incorrect due to a recent name change within the genus *Pseudomonas* to *Stutzerimonas* (Lalucat et al., 2022).

As a final check of the taxonomy, we locally aligned all dominant ASV reference sequences against the NCBI nucleotide collection *nr/nt* (Acland et al., 2014) via *BLAST* (Johnson et al., 2008); although this was useful for inferring the current best-matches to our ASVs, we have defaulted to the tree placement taxonomy as it was typically more conservative by placing sequences mainly at higher classification levels. Scores for alignments to the *NCBI* collection (Acland et al., 2014) and to Kapili & Dekas (2021) are provided in Supplemental Table 4.3. Note that we defined dominant ASVs as those that either: (a) had the most reads across samples and contributed to ~80% of the total reads in the *nifH* dataset or (b) were within the top four ASVs per sample based on total reads, and then from this (c) were within ≥ 3 samples and reached $\geq 1\%$ relative abundance within at least one sample. Hence, dominant ASVs are those that are non-rare both in relative abundance and presence across samples.

4.3.4.3 Graphics & Statistics

R version 4.1.2, *RStudio* version 2021.09.1.372, and *ggplot2* (Wickham, 2016) were used for graphical data analyses (R Core Team, 2021; RStudio Team, 2021) along with additional R packages listed in Supplemental Methods 4.S1. Multivariate analysis was conducted using *vegan* (Oksanen et al., 2022) with its *decostand* function to standardize oceanographic data and *envfit* function to fit environmental vectors on the RDA [permutations = 999]. Aitchison distances (dissimilarity index) between the *nifH* of the Bedford Basin time series and the Arctic cruise samples were calculated using the *vegdist*

function in *vegan* (Oksanen et al., 2022); the Bedford Basin ASV table used for distance calculations had been filtered to only include ASVs with ≥ 25 total reads across samples to reduce the computation time.

To prepare a table indicating the broader environmental categories that were significantly associated with each dominant ASV (i.e., Table 4.1), the *multipatt* function in the *indicspecies R* package was used to conduct a multi-level pattern analysis of species patterns versus site groupings [settings were function = *r.g* and number permutations = 9,999] (De Cáceres & Legendre, 2009). Note that Table 4.1 represents a summary of indicator species test results for all dominant ASVs (individual test results are provided in Supplemental Table 4.5). The individual results (Supplemental Table 4.5) were generated such that environmental data are first grouped into categories, samples are then assigned to said categories, and then finally the indicator species test is run for each environmental parameter to determine the dominant ASVs that are significantly associated to each given category (e.g., $< 100\text{m}$ or $\geq 100\text{m}$ for water column depth). To select category ranges for each environmental condition the general data distribution for each variable was used (hence, the results must be taken within the context of the categories that were defined).

The intersections of ASVs across samples was plotted using the *UpSetR* package (Conway et al., 2017). Maps were generated with *rnaturalearth* (South, 2017), *ggrepel* (Slowikowski, 2021), *ggspatial* (Dunnington, 2022), and *ggOceanMaps* (Vihtakari, 2022) packages with land and island data from Natural Earth Data (www.naturalearthdata.com) and also bathymetry data from the online *ETOPOI* topography dataset [*etopo180*] accessed via *ERDDAP* and its *griddap* protocol/access form (Amante & Eakins, 2009; NOAA National Geophysical Data Center, 2009; Simons, 2019). The institutions/creators listed for bathymetry data are the National Oceanic and Atmospheric Administration & the National Geophysical Data Center. Section plots of oceanographic data were also generated using *ETOPOI* data (Amante & Eakins, 2009) and *Ocean Data View* (Schlitzer, 2002, 2021).

Phylogenetic and sequence analyses were carried out in *Geneious Prime* version 2022.2.2 (www.geneious.com); therein, *RaxML* 8.2.11 was used to generate Maximum Likelihood (ML) trees for dominant NCD and cyanobacterial *nifH* sequences using rapid bootstrapping [1,000 replicates], the *GTR GAMMA* model, and a search for the best-scoring

ML tree [random seed = 12,345] (Stamatakis, 2014). Trees were built from codon-aligned nucleotide sequences generated by first aligning translated *nifH* sequences using *MUSCLE* (Edgar, 2004) and then converting back to nucleotides using *PAL2NAL* (Suyama et al., 2006). Trees visualizations were built using *iTOL* (Letunic & Bork, 2021). The Bedford Basin time series *nifH* reference sequences (with singleton ASVs removed) were searched for Arctic *nifH* ASVs using their alphanumeric names given that both datasets were processed with identical parameters. We also used the standalone *BLAST* (Altschul et al., 1990) function within *Geneious Prime* to align our *nifH* ASV reference sequences to sequences reported in Kapili & Dekas (2021) [to help with the taxon identifications described above], as well as Shiozaki et al. (2018) [to compare to the western Arctic Ocean] and Delmont et al. (2021) and Turk-Kubo et al. (2022) [to assess general endemism]. Note that we compare our ASVs to those in the study of Shiozaki et al. (2018) given that their sampling occurred from early September to early October 2015, and hence close to the same time period that our samples were collected.

4.4 Results

4.4.1 Diazotroph Biogeography within the Canadian Arctic Gateway

Our sequencing effort identified 2,490 *nifH* ASVs total within the Canadian Arctic Gateway (Fig. 4.2a). *NifH* genes were detected in both the upper (>100m) and lower (<100m) water column at all stations (Fig. 4.2a) except for stations BB2 and CAA4, where they were found only at lower depths. Likewise, diazotrophs were detected consistently in both the large (>3 μ m) and small (0.2–3 μ m) size fractions at every station (Fig. 4.2a). Due to the compositional nature of the *nifH* dataset, we converted reads to CLR values (Gloor et al., 2017). In our case, 0.1%, 1%, 4%, and 10% relative abundances correspond to approximate thresholds of 2.5, 4.5, 6, and 7 CLR values, respectively (Supplemental Fig. 4.8). At a broader taxonomic level, the entire set of ASVs fell within thirteen taxonomic groups that could be separated by class (Fig. 4.2a). Based on the total distribution of ASVs across stations, the readily observable trend characterizing *nifH* diversity at the time of sampling was the division of diazotroph communities into three distinct biogeographical groups: group (i) included the K1 and LS2 stations in the Labrador Sea, group (ii) included the BB stations in Baffin Bay and the CAA2 surface, and group (iii) included the remaining

CAA stations from the Canadian Arctic Archipelago (Fig. 4.2a). The Labrador Sea group was predominantly composed of Betaproteobacteria, while the CAA group was predominantly composed of Gamma-, Delta-, and Epsilonproteobacteria, as well as ASVs belonging to Bacteroidetes (mainly at CAA1 & CAA5) and other ASVs grouped into Unknown Bacteria (Fig. 4.2a). Since ASVs with a higher relative abundance can obfuscate rarer diazotrophs, we also plotted individual ASV patterns across stations based on the same major taxonomic groupings (Fig. 4.2b). These values point to distinct spatial patterns among the rarer diazotrophs, including: (i) Alphaproteobacteria predominant within the Labrador Sea group, (ii) diazotrophs belonging to Nitrospirae towards the surface at station CAA1, and (iii) rarely observed Planctomycetes and Verrucomicrobia ASVs detected mainly within the CAA group (Fig. 4.2b). Cyanobacterial diazotrophs were infrequent and detected in only five samples (see below); NCDs were therefore greatly overrepresented relative to cyanobacterial diazotrophs within the Canadian Arctic Gateway (Fig. 4.2). A comparison of the number of shared ASVs between samples further confirmed that while individual sites had a large set of unique ASVs (with BB stations having the fewest total number of ASVs overall), site pairings within the CAA group and within the Labrador Sea group shared more ASVs with each other than for site pairings across these two groups (Supplemental Fig. 4.9). Hence, stations within the Labrador Sea and CAA are more similar to nearby stations than to those further away with respect to their shared ASVs. An analysis of a lower taxonomic level also indicated that specific Orders typified each Class (Supplemental Fig. 4.10). For instance, Orders detected across the water column at some stations included: (i) Burkholderiales at station K1, (ii) Pseudomonadales at station BB1, and (iii) Vibrionales, Desulfobacterales, Desulfuromonadales, and Campylobacterales at multiple CAA stations (Supplemental Fig. 4.10).

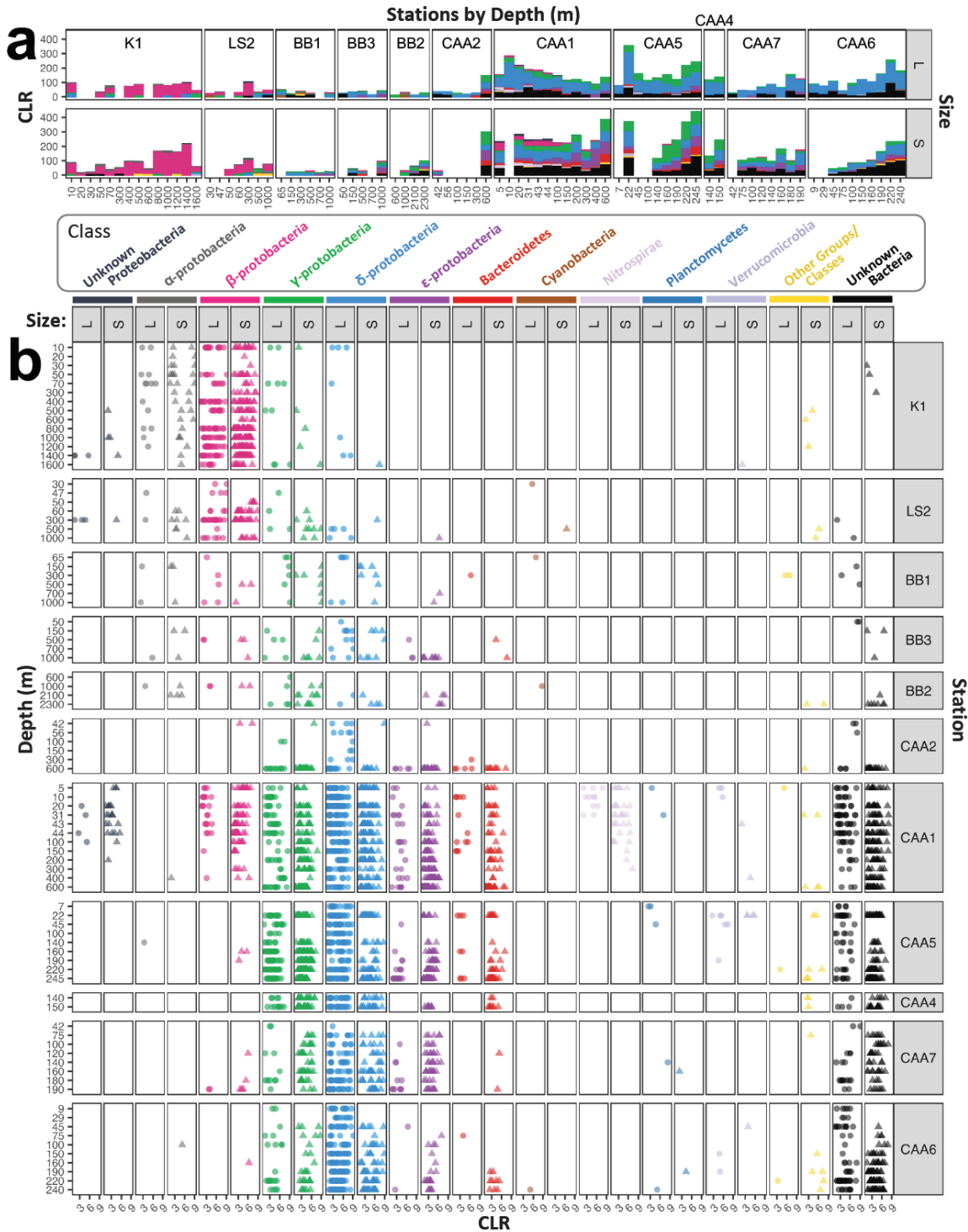


Figure 4.2. Major *nifH* Classes present in Canadian Arctic Gateway during GN02 Expedition. **(a)** Sum of centered-log ratio (CLR) values for all ASVs within each Class for each station and depth. **(b)** CLR values for individual ASVs at each station and depth by Class. DNA filters were size-fractionated where L = $\geq 3\mu\text{m}$, and S = $0.2-3\mu\text{m}$. Analyses were limited to CLR values >0 to account for data points where ASVs are absent. Rarefaction curves for datasets used herein shown in Supplemental Fig 4.22. Seventeen

ASVs had stop codons and were therefore suggestive of pseudogenes, these were all very rare occurring in only one sample each with overall averages of 3 reads, 0.17% relative abundances, and 2.3 CLR scores across all samples. Based on taxonomic placement, these pseudogenes all fell within the numerous proteobacteria and unknown bacteria found in the Labrador Sea and CAA stations, given their extreme rarity they remain in the analyses.

Total diversity for all ASVs was further analysed via redundancy analysis (RDA) and regression fitting of oceanographic measurements (Fig. 4.3). Oceanographic conditions significantly correlated to diazotrophic diversity during our study were temperature, depth, salinity, oxygen, fluorescence, nitrate, and phosphate (all p -values < 0.05; Supplemental Table 4.2). Nitrite and silicate were not significantly correlated to diazotrophic diversity within our samples (p -values > 0.05; Supplemental Table 4.2). Trace metals were not included in the RDA analysis given that their concentrations were not available for half of the DNA samples, but they are instead considered below with respect to dominant ASVs. The RDA analysis required many axes to explain the variance in the dataset (the first seven axes explained only ~8.56% of the variance). Even so, the first two axes explaining the most variance (5%) divided samples from the Labrador Sea (K1 & LS2) from those of the Baffin Bay (BB) and CAA region in the first axis (RDA1; Fig. 4.3). The Labrador Sea samples were associated with higher temperatures (Fig. 4.3). The second axis (RDA2) mainly divided surface samples from the CAA (particularly CAA1 and CAA5) from deeper CAA samples (and to some degree Labrador Sea samples) that were associated with higher salinities and lower oxygen (Fig. 4.3). These trends were evident in the RDA analysis regardless of size fraction (Supplemental Fig. 4.11).

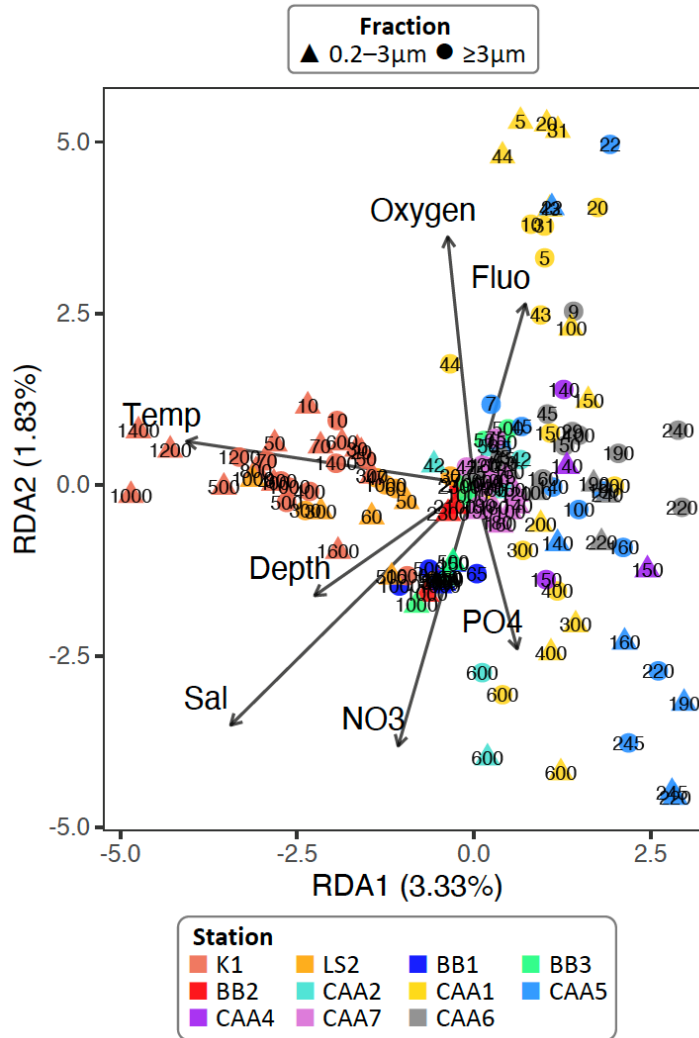


Figure 4.3. Multivariate analysis of *nifH* communities for Canadian Arctic Gateway during GN02 expedition. Shown is a redundancy analysis (RDA) for all *nifH* ASVs with environmental vectors for temperature (Temp), dissolved oxygen (oxygen), salinity (Sal), nitrate (NO₃), phosphate (PO₄), and fluorescence (Fluo).

Oceanographic summary plots for temperature, salinity, fluorescence, and oxygen are provided (Fig. 4.4a). Similarly, nutrient data for nitrite, nitrate, phosphate, and silicate (Si), as well as trace metal data for dissolved iron (D-Fe) and manganese (D-Mn) and total particulate iron (TP-Fe), vanadium (TP-V), and phosphate (TP-P) are also shown (Fig. 4.4b and Supplemental Fig. 4.12). Iron and vanadium metals were selected as they are especially relevant to the nitrogenase enzyme (Zehr et al., 2003). No data were available for Molybdenum. Given that detailed descriptions for all the above parameters have been published elsewhere (Colombo et al., 2019, 2020, 2021, 2022; Lehmann et al., 2019;

Schuback et al., 2017), they will only be summarized briefly here. Our samples were collected towards the end of seasonally expected ice coverage [end of July into August](Lehmann et al., 2019; Randelhoff et al., 2020). At the time of sampling, the mixed-layer depth was distinct and extended to a maximum of approximately 40m at LS2 (Fig. 4.4; Schuback et al., 2017). Within this upper layer and extending to the subsurface chlorophyll maximum (SCM) (located sub-mixed layer), an increased fluorescence signature is attributed to phytoplankton growth [highest SCM at 28m for BB3](Fig. 4.4; Schuback et al., 2017). Lower temperatures were seen below the mixed layer, and nitrate and phosphate nutrient concentrations were lower [$<10\mu\text{M}$ for nitrate and $<1.35\mu\text{M}$ for phosphate above 50m] in the mixed layer (Fig. 4.4; Schuback et al., 2017). According to Schuback et al., (2017), N:P ratios below 16:1 also suggested nitrate limitation in the mixed-layer throughout the Labrador Sea, Baffin Bay, and CAA. Colder saltier seawater is evident in the deep water of the Baffin Bay and warmer saltier seawater in the Labrador Sea, whereas, less saline seawater characterizes the surface of the Baffin Bay and CAA stations where surface warming and freshwater inputs from the Arctic region are apparent [e.g., from glacial and sea ice meltwater](Fig. 4.4; Colombo et al., 2019; Lehmann et al., 2019). In the deep and older Baffin Bay waters, a reduction in oxygen corresponds to higher nutrient concentrations (Fig. 4.4; Lehmann et al., 2019). Dissolved iron and manganese are higher in the CAA and lower in the Labrador Sea (Fig. 4.4b and Supplemental Fig. 4.12; Colombo et al., 2020). In the CAA, dissolved trace metals come from the benthic layer and can be advected to Baffin Bay (Colombo et al., 2020, 2021). Total particulate iron and vanadium are higher near the bottom of the water column and at boundary currents, while total particulate phosphate is reflective of primary productivity at the surface (Fig. 4.4b and Supplemental Fig. 4.12; Colombo et al., 2022).

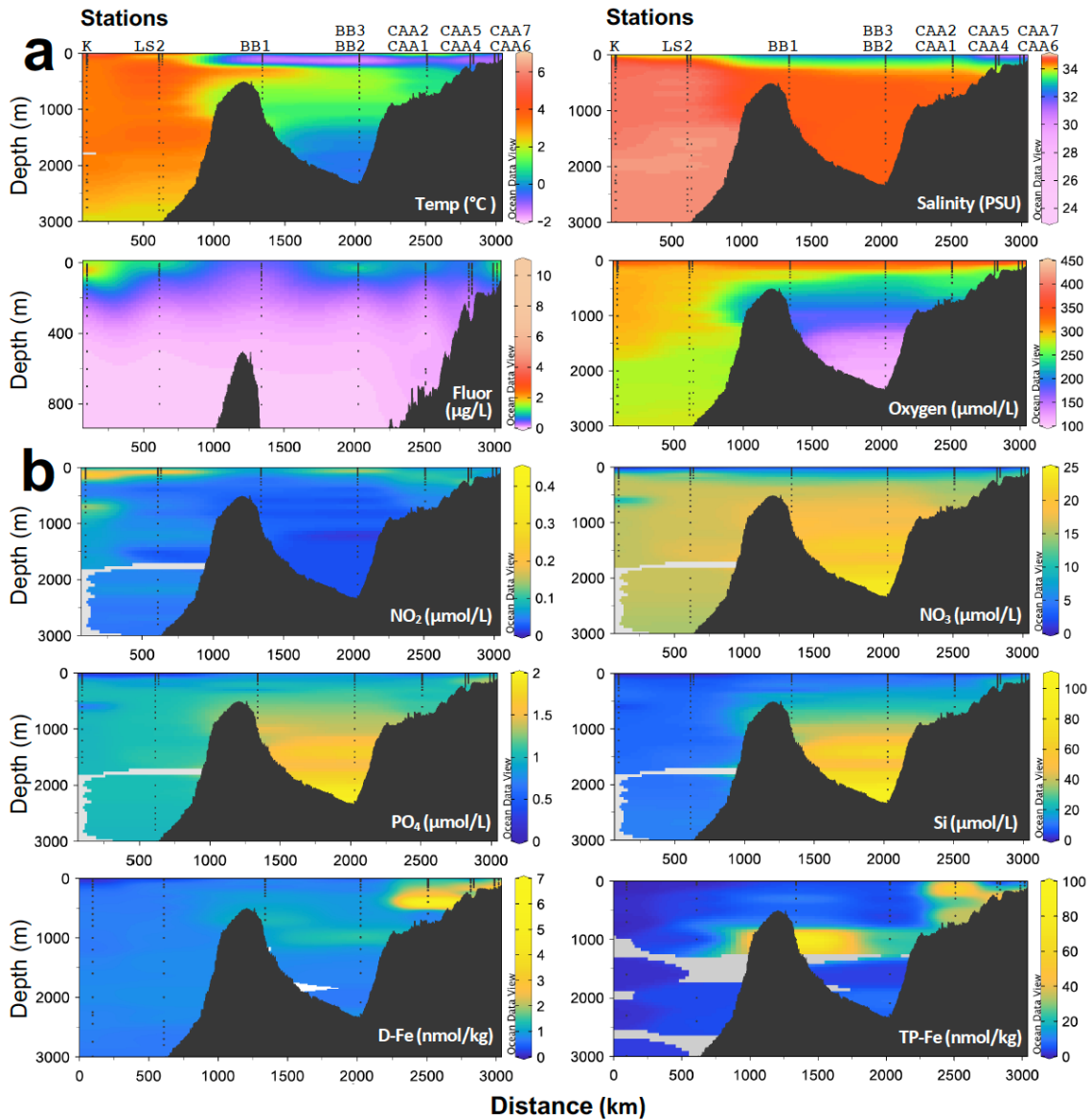


Figure 4.4. Oceanographic data for Canadian Arctic Gateway during 2015 GN02 expedition. **(a)** Sensor-derived temperature (Temp), salinity, fluorescence (Fluor), and dissolved oxygen (Oxygen). **(b)** Bottle-derived nutrients and trace metals from *IDP2021* dataset (GEOTRACES Intermediate Data Product Group, 2021), and supplemental data in Colombo et al. (2020, 2021, 2022). Shown are NO_2^- (nitrite), NO_3^- (nitrate), PO_4^{3-} (phosphate), Si (Silicate), D-Fe (dissolved iron), and TP-Fe (total particulate iron). Dark regions outline the bottom bathymetry (*ETOPO1* dataset; Amante & Eakins, 2009) along the transect based on distance (km) starting from the outer Labrador Sea (at K1). No TP-Fe data available for station CAA2. Plots were generated with *Ocean Data View* (Schlitzer, 2002, 2021).

4.4.2 Dominant CAG Diazotrophs and their Environmental Links

We identified and further characterized 106 dominant *nifH* ASVs throughout our dataset of the Canadian Arctic Gateway. Recall that dominant ASVs met the following criteria: had the most reads (which contributed to ~80% of the total *nifH* dataset) *or* were within the top four ASVs per sample *and* were in ≥ 3 samples *and* reached $\geq 1\%$ relative abundance in at least one sample. Although these 106 ASVs represent only 4.3% of the total observed ASVs, they account for 52% of the *nifH* reads. The remaining non-dominant ASVs were less widespread across our samples and belonged to the broader groups already mentioned (Fig. 4.2 and Supplemental Fig. 4.10). Therefore, we will not consider these ASVs in greater detail. A phylogenetic tree for the 53 ASVs that occurred in more than seven samples shows the diversity encapsulated by the most frequently observed ASVs ($53/106 = 50\%$ of the dominant ASVs; Fig. 4.5). A summary table outlining the results for multi-level pattern analyses identifies the environmental categories that significantly correlated with each dominant ASV (or groups of ASVs if they share very similar taxa; based on p -values ≤ 0.05 ; Table 4.1). Complete spatial profiles for all dominant ASVs at each station and depth are also provided as supplemental information (see Supplemental Figs. 4.13 and 4.14), as are alignment scores for the final taxonomy shown (Supplemental Table 4.3). The dominant diazotrophs in the Canadian Arctic Gateway during our study all belonged to NCD phylotypes within diazotrophic Clusters I and III; there were no dominant ASVs detected for clusters IB, II, and IV (Fig. 4.5).

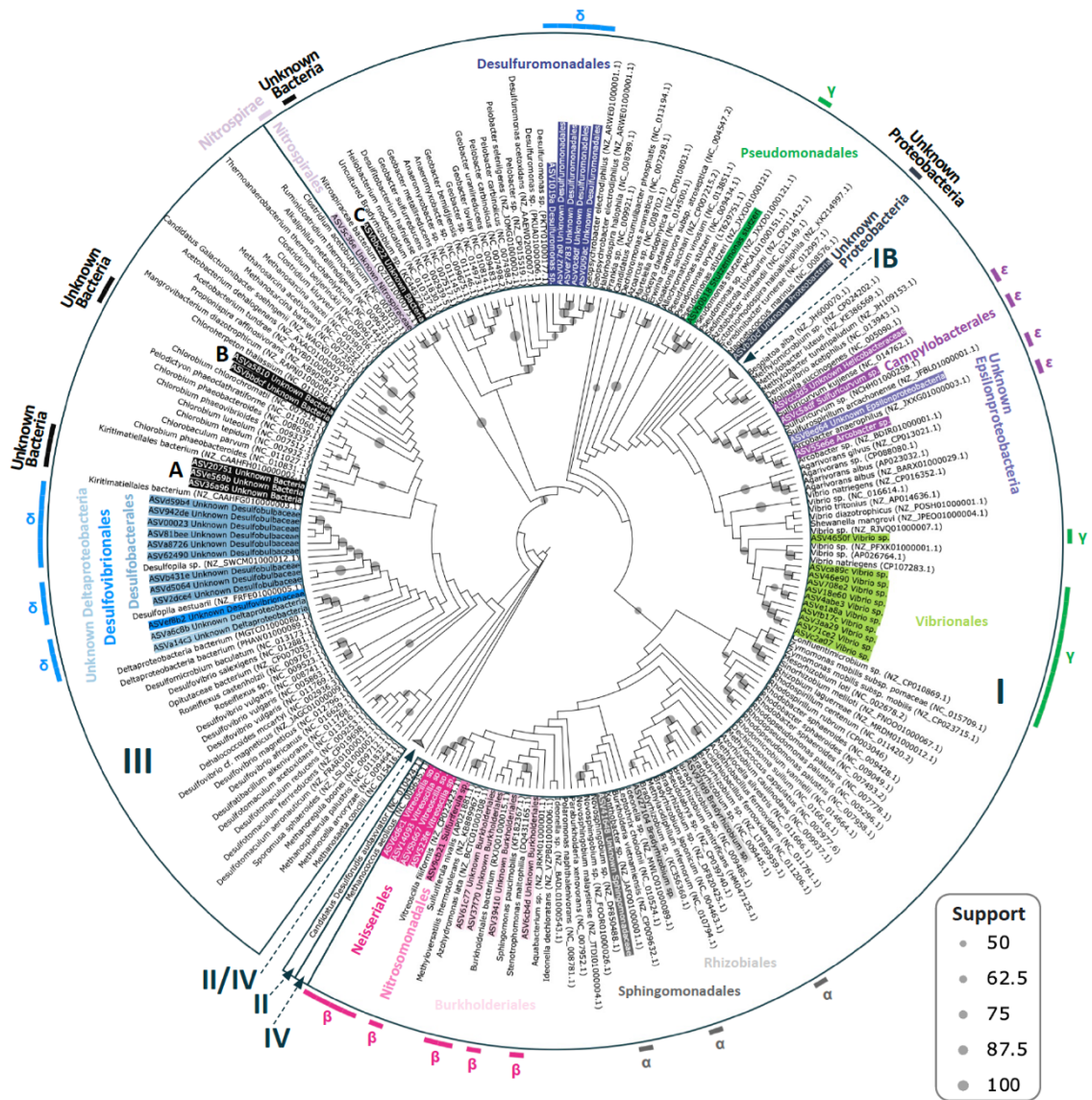


Figure 4.5. Maximum Likelihood tree of major *nifH* amplicon sequence variants (ASVs) present in the Canadian Arctic Gateway during the GN02 expedition. ASVs shown were those present in ≥ 7 samples. To increase legibility, branch lengths were not used (see Supplemental Fig. 4.21 for branch lengths). Non-colored sequences are a collection of closest reference sequences currently matching each ASV as retrieved via *BLAST* from NCBI (Acland et al., 2014; Johnson et al., 2008) and from Kapili & Dekas (2021). The tree also includes other diazotrophs for which taxonomy and genomes are known. Support values (Support) are bootstrap values based on 1,000 replicates. The tree is built using a codon-alignment. For colored labels that represent Arctic ASVs collected herein: large roman numerals (I–IV) shown in boxes = major diazotroph clades, outer labels around circle = taxonomic classes, inner labels around branches = taxonomic orders. Greek characters refer to groups of proteobacteria.

Table 4.1. Summary of multi-level pattern analyses for dominant ASVs (or ASV groups) in CAG GN02 cruise based on sample groupings. Analysis limited to ASVs in ≥ 7 samples. ASVs are denoted as belonging to a certain sample category based on having statistically significant p -values ≤ 0.05 (at $\alpha = 0.05$). KS & LS = Labrador Sea, BB = Baffin Bay, and CAA = Canadian Arctic Archipelago. Categories are: size fractions (S), water column depth (D), temperatures (T), salinities in PSU (Sa), fluorescence $\mu\text{g/L}$ (F), dissolved oxygen $\mu\text{mol/L}$ (O_2), nitrite $\mu\text{mol/L}$ (NO_2^-), nitrate $\mu\text{mol/L}$ (NO_3^-), phosphate $\mu\text{mol/L}$ (PO_4^{3-}), silicate $\mu\text{mol/L}$ (Si), dissolved Fe nmol/kg (DI-Fe), total particulate Fe nmol/kg (TP-Fe), total particulate vanadium pmol/kg (TP-V). Each category is divided into low (L) and high (H) values based on the general distribution of values and concentrations in Fig. 4.4 (color codes defined under each column heading). Dashes (-) represent groupings that were insignificant; for example, ‘Sm/-’ would mean that the ASVs were either associated with the small size fraction or insignificantly associated to either size fraction. Question marks (?) indicate that although ASVs may have been associated to the high or low category, they were also significantly associated with samples where no trace metal data were collected, hence any significance may be an artifact of reduced sampling. Test results for individual ASVs are shown in Supplemental Table 4.5, and ocean data are from *IDP2021* (GEOTRACES Intermediate Data Product Group, 2021) and (Colombo et al., 2020, 2021, 2022). ‘UNKN’ stands for Unknown.

ASV Name (n = number of ASVs)	Region	S	D	T	Sa	F	O_2	NO_2^-	NO_3^-	PO_4^{3-}	Si	DI-Fe	TP-Fe	TP-V
		0.2-3 μm $\geq 3\mu\text{m}$	<100m $\geq 100\text{m}$	<2 $^\circ\text{C}$ $\geq 2^\circ\text{C}$	<32 ≥ 32	<2 ≥ 2	<268 ≥ 268	<0.1 ≥ 0.1	<10 ≥ 10	<0.8 ≥ 0.8	<20 ≥ 20	<2 ≥ 2	<25 ≥ 25	<100 ≥ 100
■ <i>Bradyrhizobium</i> spp. ($n = 2$)	K	Sm/-	De/-	H	H/-	-	H	-	L/-	L	L	L?/-	L?/-	L?/-
■ UNKN Burkholderiales ($n = 4$)	K/LS	Sm/-	-	H	H/-	-	H/-	-	-	L/-	L/-	L/-	-	-
■ <i>Vitreoscilla</i> spp. ($n = 4$)	K/LS	Sm/-	-	H	-	-	H	-	-	-	L/-	L/-	L?/-	L?/-
■ <i>Stutzerimonas stutzeri</i> - ASVb0b18	BB	-	De	-	H	-	L	-	H	H	-	-	L	L
■ UNKN Desulfuromonadales sp. ($n = 4$)	BB/CAA/-	Lr/-	-	L/-	L/-	-	H/-	-	-	-	-	-	H/L/-	H/L/-
■ <i>Sulfuriferula</i> sp. - ASV9cb21	CAA	-	Su	L	L	-	H	-	L	L	L	-	H	H
■ <i>Sulfuricurvum</i> sp. - ASVc5adf	CAA	-	Su	L	L	-	H	-	L	L	L	-	H	H
■ UNKN Nitrospiraceae - ASV5c36e	CAA	-	Su	-	-	-	H	-	L	L	L	-	H	H
■ Unknown Bacteria C ($n = 1$)	CAA	-	Su	L	-	-	H	-	L	L	L	-	H	H
■ <i>Arcobacter</i> sp. - ASV55e6e	CAA	-	De	L	-	-	L	H	H	H	H	-	H	-
■ <i>Vibrio</i> spp. ($n = 11$)	CAA/-	Sm/-	De/-	L/-	H/-	H/-	L/-	H/-	H/-	H/-	-	H/H?/-	H?/-	H?/-
■ Unknown Bacteria A ($n = 3$)	CAA/-	Lr/-	Su/De/-	L/-	-	-	H/-	L/-	L/H/-	H/-	-	-	H?/-	H?/-
■ Unknown Bacteria B ($n = 2$)	CAA/-	Sm/-	De	L/-	-	-	-	-	H	-	-	H?/-	-	-
■ UNKN Desulfobulbaceae ($n = 9$)	CAA/-	Lr/-	-	L/-	L/-	H/-	-	H/-	-	-	H/-	H/-	-	-
■ UNKN Desulfovibrionaceae - ASVef8b2	CAA	Sm	-	L	-	-	-	-	-	-	-	H	-	-
■ UNKN Sphingomonadaceae - ASV41648	All*	Sm	-	-	-	-	-	L	-	-	-	-	-	-
■ <i>Desulfuromonas</i> sp. - ASV1019a	CAA*	-	-	-	-	-	-	-	-	H	-	-	-	-
■ UNKN Deltaproteobacteria ($n = 2$)	CAA1*	Sm	Su	-	-	-	H/-	-	L	L	-	-	H	H
■ UNKN Epsilonproteobacteria - ASV8ed64	CAA*	Sm	-	-	-	-	L	L	H	H	H	-	-	-
■ UNKN Helicobacteraceae - ASVcdd5	CAA*	-	De	-	-	-	L	-	H	H	-	H	-	-
■ UNKN Proteobacteria - ASVb20cf	CAA1*	-	Su	-	-	-	H	-	L	L	-	-	H	H

* Can only be inferred from spatial profiles but not from statistical tests (this can occur if ASV was present at only one station, was present across all stations, or was observed too infrequently for statistical test to be significant).

Consistent with earlier analyses of the full Arctic *nifH* dataset, dominant ASVs likewise showed spatially distinct distribution patterns and were correlated to a particular region within the Canadian Arctic Gateway (Table 4.1). To summarize, within the Labrador Sea dominant ASVs belonged to the genera *Bradyrhizobium* and *Vitreoscilla*, as well as to the order Burkholderiales (all within Cluster I; Fig. 4.5 and Table 4.1). Given their association to the Labrador Sea, these ASVs were all associated with higher temperatures, higher oxygen levels, and generally lower trace metal concentrations (Table 4.1). In Baffin Bay, one single dominant ASV belonged to *Stutzerimonas stutzeri* (formerly

Pseudomonas stutzeri) from Cluster I; this ASV was present primarily throughout deeper (>100m) samples and was associated with higher salinities, lower oxygen levels, and higher concentrations of nitrate and phosphate (Fig. 4.5; Table 4.1). In the CAA, various dominant ASVs were present and were related to *Desulfopila* spp. (the Desulfobulbaceae), *Desulfuromonas* sp. or *Geopsychrobacter* sp. (the Desulfuromonadales), *Sulfuricurvum* spp. (Campylobacterales), *Arcobacter* sp. (also Campylobacterales), *Vibrio* spp. (the Vibrionales), and *Sulfuriferula* sp. (Nitrosomonadales). The CAA region also contained several ASVs that were less taxonomically resolved including an Unknown Desulfovibrionaceae, two Unknown Deltaproteobacteria, an Unknown Nitrospiraceae, an Unknown Epsilonproteobacteria, and an Unknown Sphingomonadaceae (though this is most likely a *Novosphingobium* sp. (Fig. 4.5; Table 4.1). Although environmental conditions associated with dominant CAA ASVs were more variable than those observed for the Labrador Sea and Baffin Bay, these ASVs largely correlated with lower temperatures and could somewhat be divided by depth (Table 4.1). Interestingly, the CAA was home to three dominant ASV groups that could only be identified to the Bacteria domain (labelled A, B, & C) and one identified to the Proteobacteria phylum (Fig. 4.5). Unknown Bacteria A is most similar (86% pairwise identity; PI) to Kiritimatiellales, Unknown Bacteria B is most similar (~79% PI) to *Mangrovibacterium* sp., and Unknown Bacteria C is similar to Nitrospiraceae (81% PI) and an “Uncultured *Bradyrhizobium* sp.” (87% PI); the latter *Bradyrhizobium* may be a horizontally transferred *nifH* gene due to its phylogenetic placement away from other *Bradyrhizobium* spp. (Fig. 4.5). The Unknown Proteobacteria is similar (80% PI) to *Vibrio* sp. and the group known as Gamma_1 (84% nucleotide PI; Turk-Kubo et al., 2022). Since the multi-level pattern analysis was based on broad groupings (e.g., “high” vs. “low” temperature, etc.) it was unable to pinpoint ASVs associations with discrete environmental conditions. Hence, we also plotted CLR distributions for each dominant ASV relative to environmental data (Supplemental Fig. 4.15). These results confirm the broader Labrador Sea versus CAA divisions discussed above (e.g., for temperature, salinity, and trace metals), but also showed: (i) the highest fluorescence values were mainly observed when Unknown Desulfobulbaceae *nifH* ASVs were present, (ii) higher nitrite concentrations were observed when *Stutzerimonas stutzeri*, *Vitreoscilla* sp., Unknown Burkholderiales, and a *Bradyrhizobium* sp. were present,

and (iii) several instances of high silicate were observed when *Novosphingobium*/Sphingomonadaceae was present (Supplemental Fig. 4.15). Dominant ASVs with the highest relative abundances and most frequent detections (each in ≥ 27 samples) included *Bradyrhizobium* (ASV93fd9) in the Labrador Sea, *Stutzerimonas stutzeri* (ASVb0b18) in Baffin Bay, and at least three ASVs from the Unknown Desulfobulbaceae in the CAA.

Collectively, these results demonstrate that the Labrador Sea and Baffin Bay were characterized by only a few dominant *nifH* ASVs during our study, whereas the CAA was characterized by a relatively greater number of dominant *nifH* ASVs (Table 4.1). Furthermore, dominant diazotrophs mirrored the biogeographic divisions seen for the entire diazotroph community (Fig. 4.2 and Fig. 4.3). Interestingly, the CAA also contained multiple dominant ASVs with poor taxonomic characterization suggesting that the CAA may be home to several yet-undescribed NCDs that were particularly important within this region of the Arctic Ocean during the time of sampling.

4.4.3 UCYN-A within the Canadian Arctic Gateway

Although rare, cyanobacterial diazotrophs were identified in five samples, and when present, represented 1.8–61% of the *nifH* reads of that sample. These cyanobacterial diazotrophs belonged to *Trichodesmium*, *Pseudanabaena*, and two Unknown Cyanobacteria that are phylogenetically closest to *Chroococcidiopsis* and *Euhalothece* genera (Fig. 4.6a). UCYN-A was also detected at one station in the Baffin Bay (BB1, 60m, large size fraction; Fig. 4.6a), prompting us to further quantify the *nifH* gene for this species within the CAG using qPCR assays. *NifH* gene copy numbers for two UCYN-A ecotypes, A1 and A2, are shown (Fig. 4.6b). Although UCYN-A1 and A2 were detected in the Labrador Sea, Baffin Bay, and the CAA, values were typically very close to the LOQ [$5 \text{ nifH copies L}^{-1}$] (Fig. 4.6b). Neither of the UCYN-A ecotypes were detected at station LS2 within the Labrador Sea (Fig. 4.6), while highest values (but still very low at less than $\sim 30 \text{ nifH copies L}^{-1}$) were observed for both ecotypes at CAA7 at the top of the water column. No diazotrophs were detected in these surface CAA7 samples using *nifH* amplicon sequencing (Fig. 4.2a), suggesting that *nifH* amplification via the widely used nested PCRs and UV imaging approach was not sensitive enough to detect UCYN-A at this station.

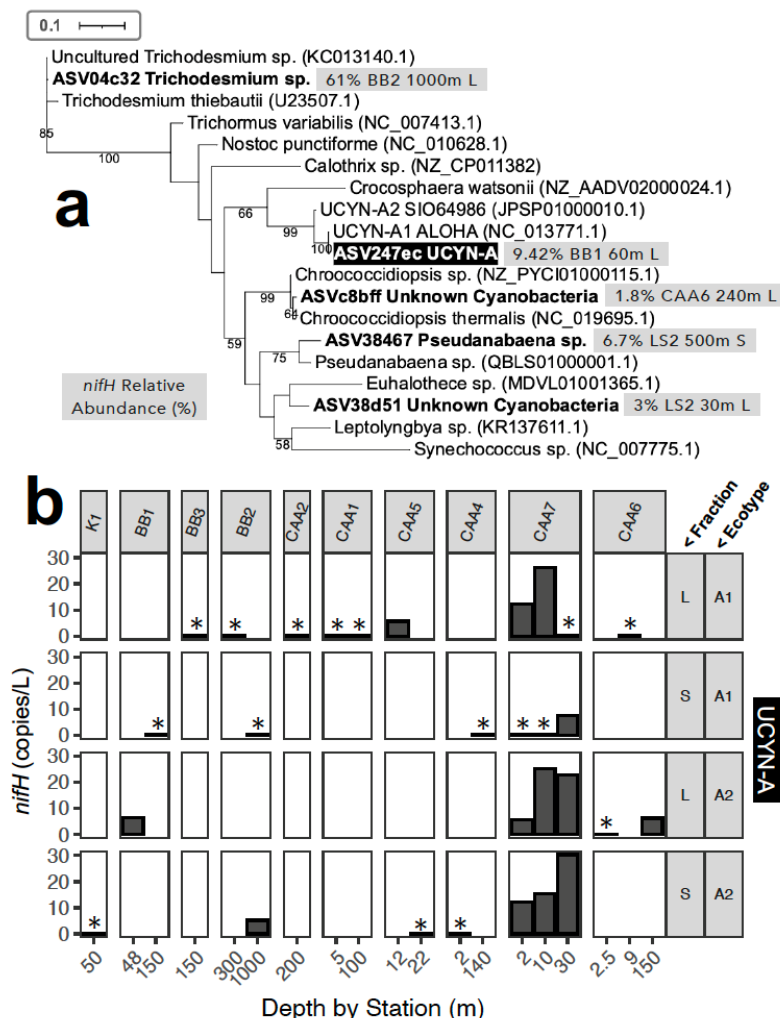


Figure 4.6. Cyanobacterial diazotrophs detected in Canadian Arctic Gateway during GN02 cruise. **(a)** Maximum Likelihood phylogeny of all cyanobacterial diazotroph ASVs detected. Additional sequences in tree that are similar to the ASVs detected (non-bold labels) were collected from NCBI (Acland et al., 2014; Johnson et al., 2008) and Kapili & Dekas (2021). Support values (numbers mid branch) are rapid bootstrap values based on 1,000 replicates. The tree is built using a codon-alignment. Each ASV was only detected in one sample and the percent relative abundance within such samples is shown (grey boxes). UCYN-A was also detected based on its *nifH* (black box). **(b)** Quantitative-PCR data for UCYN-A1 and -A2 ecotypes. Size fractions (Fraction) correspond to L = $\geq 3\mu\text{m}$, and S = 0.2– $3\mu\text{m}$. Values with asterisks (*) are those where UCYN-A was detectable but not quantifiable [$\text{LOQ} = 5 \text{ nifH copies L}^{-1}$].

Furthermore, at BB1 where the UCYN-A *nifH* gene was detected through amplicon sequencing, there were no detectable UCYN-A *nifH* copies via qPCR (Fig. 4.6); this disparity may arise from the greater number of PCR cycles associated with *nifH* sequencing (Zani et al., 2000). Note that the UCYN-A1 ASV sequenced herein would be compatible

with the UCYN-A1 qPCR assay. These data indicate that while UCYN-A was primarily found at CAA7 in the Canadian Arctic Archipelago, it was also detectable throughout the Canadian Arctic Gateway during our study despite being found in the rare fraction of the diazotroph microbiome.

4.4.4 Searching for Dominant Canadian Arctic Gateway Diazotrophs Outside of the CAG & Polar Realm

To assess whether the dominant ASVs identified herein also occurred outside of the Canadian Arctic Gateway, we searched for these *nifH* ASVs sequences within: (i) a nearby fjord in the coastal NWA (data collected herein), (ii) previously published data for the western Arctic Ocean (Shiozaki et al., 2018), (iii) *Tara* samples collected at lower latitudes (Delmont et al., 2021), and (iv) a recently curated and comprehensive catalogue of NCDs based on samples from across the world's oceans (Turk-Kubo et al., 2022).

Seawater in the Bedford Basin displays the expected seasonal cycling of a temperate fjord in the North Atlantic, with lower surface temperatures in the winter and higher surface temperatures in the summer (Fig. 4.7a). In the surface, nutrient concentrations display a well-established seasonal cycle driven by a seasonal-increase in water column stratification that leads to a spring phytoplankton bloom and a drawdown of inorganic nitrogen (shown herein via nitrate; Fig. 4.7a; Li & Dickie, 2001). At the bottom of the fjord (near 60m), salinity and temperature are more seasonally stable with a narrower range of values, although winter mixing can lead to distinct fluctuations (Fig. 4.7a). Known intrusion events bring shelf-waters from the North Atlantic into the deeper waters of the fjord, leading to pulsed increases in salinity and temperature, and a pulsed decrease in nitrate (these events may be seen via dashed lines in Fig. 4.7a; Haas et al., 2021). Aside from the occasional intrusions and daily tidal action, there is a net outward flow of seawater through the fjord (Kerrigan et al., 2017). Seasonally through 2014, and weekly through 2015–2019, we sampled for diazotrophs in the Bedford Basin (BBMP in Fig. 4.1a); this sequencing effort amounted to 81,661 *nifH* ASVs (limited to ASVs observed more than once). Since it is outside the scope of this study, we will not describe in full detail the weekly diazotroph communities of the Bedford Basin. Based on identical DNA sequence identity (PI), approximately ~10% [or 247/2490] of the total ASVs observed in the CAG

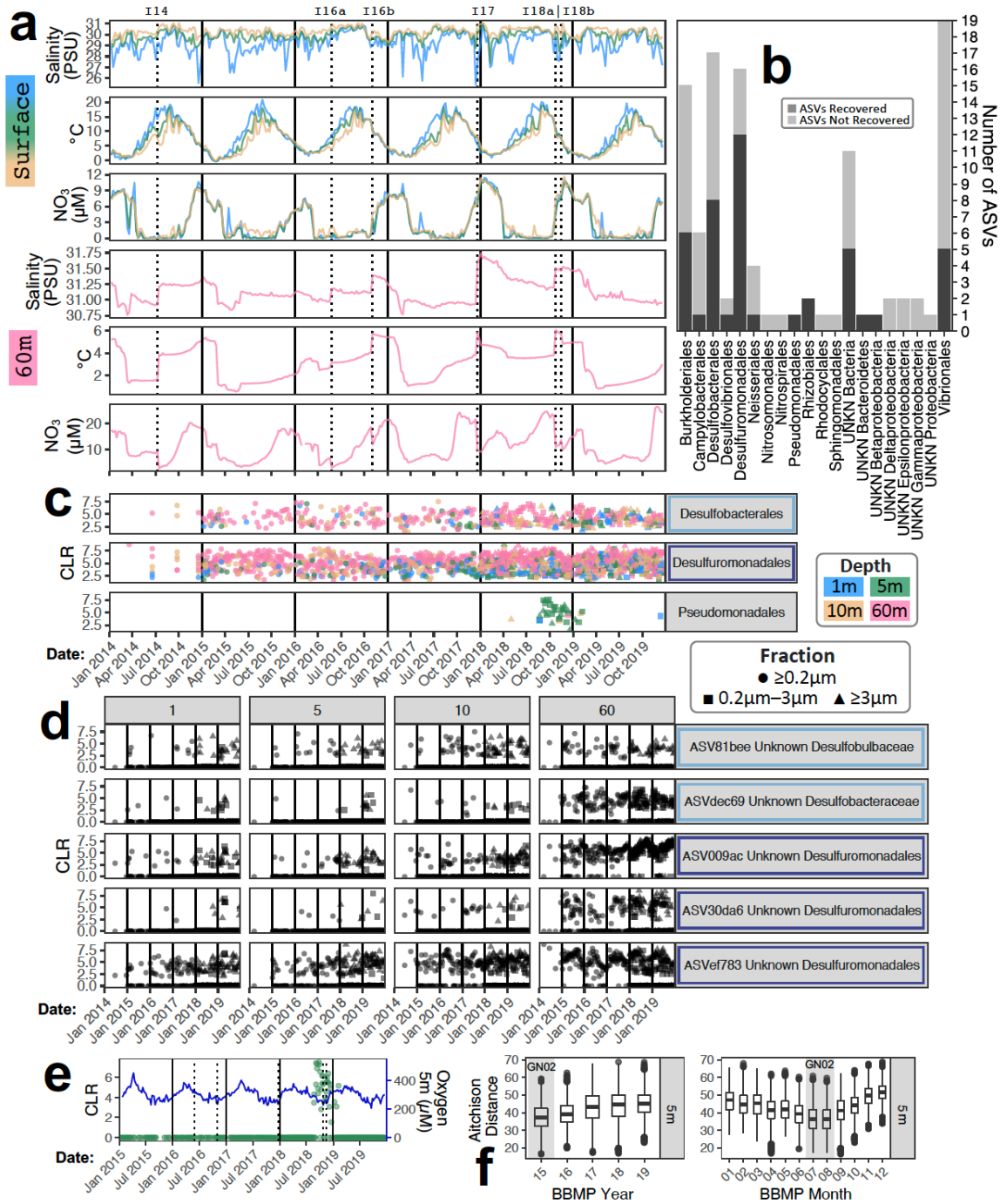


Figure 4.7. Dominant ASVs recovered from the CAG within a 5-yr time series in the NWA. **(a)** surface and deep (60m) ocean conditions of the Bedford Basin time series (2015–2019). Known shelf-water intrusion events are shown as dashed vertical lines (I14–I18b). **(b)** Number of dominant ASVs recovered and not recovered in the time series by taxonomic group. **(c)** CLR values for top two taxonomic groups recovered, as well as a Pseudomonadales ASV that had a distinct temporal pattern. Note only ten samples were collected in 2014 and size fractionation started in 2018, thus doubling observations for the last two years. **(d)** Profiles for ASVs that comprise nearly all observations shown the upper

two panels of (c). (e) Comparison of *S. stutzeri* (Pseudomonadales) versus dissolved oxygen in the Bedford Basin at 5m. (f) Distributions for Aitchison Distances between Bedford Basin 5m and Arctic 5m samples, grey highlighting = year and months when Arctic GN02 cruise occurred (other depths shown in Supplemental Fig. 4.20).

were recovered in the fjord (Supplemental Fig. 4.16); this fraction was much higher for dominant Arctic ASVs at ~42% [or 44/106 ASVs], with the most recovered dominant groups belonging to the Desulfobacterales [8 ASVs] and Desulfuromonadales [12 ASVs] (Fig. 4.7b-c; also see Supplemental Fig. 4.17 for the full set of ASVs). Of these 20 ASVs, five comprised most of the detections (Fig. 4.7d and Supplemental Fig. 4.18). Based on phylogenetic placement (Fig. 4.5), one of the Desulfobacterales is related to *nifH* from *Desulfopila* [ASV81bee], and two of the Desulfuromonadales are most related to the *nifH* from *Geopsychrobacter* sp. [ASV009ac & ASV30da6; 88–89% PI] (Supplemental Table 4.3). These ASVs were generally present throughout each year and were mainly detected at 60m, whereas one of the Unknown Desulfuromonadales [ASVef783] was detected in both surface and deep samples (Fig. 4.7d). Further comparison to the dominant diazotrophs from the western Arctic Ocean given by Shiozaki et al. (2018) indicated four overlapping reference sequences (Supplemental Table 4.4). These all belonged to Desulfobacterales and Desulfuromonadales, however, ‘ASV30da6 Unknown Desulfuromonadales’ was the sole ASV recovered at high frequency within the fjord (Supplemental Table 4.4). The Pseudomonadales group is also of special interest, given its very distinct presence within the fjord at one depth and time period (late 2018 at 5m; Fig. 4.7c). This group was characterized by a single ASV identical to *Stutzerimonas stutzeri* [ASVb0b18] found associated with deeper samples of higher salinity and nutrients and lower oxygen in Baffin Bay (Table 4.1). Interestingly, in the Bedford Basin this *nifH* sequence was present in near opposite conditions; it was restricted to the surface with comparatively lower salinity relative to Baffin Bay and had highest relative abundance/CLR values when nitrate was low, and temperatures were above 10°C (Fig. 4.7 and Supplemental Fig. 4.15; also see Supplemental Fig. 4.19a for easier interpretation at 5m). Other oceanographic data suggests that the presence of this ASV in the Bedford Basin may have been tied to oxygen levels. Although temperature ultimately controls the amount of oxygen that will dissolve at the surface (Supplemental Fig. 4.19b), the ASV matching to *S. stutzeri* was detected mainly when surface dissolved oxygen dropped to concentrations more comparable to Baffin Bay

(Fig. 4.7e versus oxygen in Fig. 4.4a). Although temperatures in the fjord were higher than in the Arctic during the time of sampling, Aitchison Distances calculated across all sample pairs between the two areas (fjord versus Arctic) indicated the highest diazotroph community similarities occurred between the two areas during the month and year in which the Arctic cruise samples were collected (Fig. 4.7f). This trend was consistent across all depths (Supplemental Fig. 4.20). Hence, the Bedford Basin *nifH* timeseries appears to share some overall seasonal similarity to the Canadian Arctic Gateway with respect to its diazotroph community, at least during the mid to late summer; further Arctic sampling is required to determine whether this trend holds true across the year.

Finally, the diazotrophs that were dominant in the CAG did not match any of the *nifH* sequences belonging to the diazotroph MAGs reported by Delmont et al. (2021) with high sequence identity [at 99.69% coverage (cov.) and 100% PI to the ASVs herein]. Even at a lower threshold there was only one MAG [AON_82_MAG_70] that somewhat matched to a single one of our dominant ASVs [ASV41648 Unknown Sphingomonadaceae; most similar to *Novosphingobium*] (>95% cov. & >92% PI; Supplemental Tables 4.3 and 4.4). Such low recoveries suggest that the *nifH* ASVs we identified from the CAG were not well-represented in *Tara Oceans* diazotroph MAGs collected at mainly lower latitudes (Delmont et al., 2021). Recently, a comprehensive catalogue of major NCD groups in the ocean has become available (Turk-Kubo et al., 2022). Matches to ~13% [14/106 ASVs] of the dominant Arctic diazotrophs of the CAG at the time sampled were found within this catalogue based on nucleotide sequences (at >95% cov. and >92% PI; Supplemental Table 4.4). When considering amino acid similarities for the best nucleotide matches to the catalogue instead, this proportion increased to 44% (Supplemental Table 4.4). Dominant CAG ASVs well-recovered by the NCD catalogue included members of the Burkholderiales, Desulfobacterales, Desulfuromonadales, Pseudomonadales, Rhizobiales, and Sphingomonadales (Supplemental Table 4.4). Dominant CAG ASVs poorly recovered by the catalogue (<90% amino acid similarity) were the Campylobacterales, the Nitrosomonadales [ASv9cb21 *Sulfuriferula* sp.], the Unknown Desulfovibrionaceae ASVef8b2, the Unknown Nitrospiraceae ASV5c36e, and two Unknown Deltaproteobacteria [ASVa6c8b and ASVa14c3] (Supplemental Table 4.4). With marginally better representation, the Unknown Bacterial Groups A, B, and C had at

least one member within each group with 92–95% amino acid similarity to a NCD within the catalogue (Supplemental Table 4.4). Overall, the above results taken together suggest that while some dominant diazotrophs in the Canadian Arctic Gateway were characterized by ASVs elsewhere in the western Arctic and coastal NW Atlantic (chiefly, Desulfobacterales and Desulfuromonadales), the majority (>50%) were poorly characterized at the ASV-level by reference sequences earmarked as the most important NCDs known at lower-latitudes or elsewhere in the ocean (Delmont et al., 2021; Turk-Kubo et al., 2022). This finding is further underscored by the fact that several of the groups identified herein can currently only be identified at higher levels of classification based on comparisons to known *nifH* sequences (Supplemental Table 4.3) and current *nifH* taxonomic methods (for e.g., using the tree of Kapili & Dekas, (2021) as was completed herein).

4.5 Discussion

4.5.1 Taxonomy and Predicted Ecology of NCDs within the CAG

Although our results present a snapshot of diazotrophic diversity within the Canadian Arctic Gateway, they capture a period during the year when one would expect a putative ecological niche for diazotrophs. Samples were collected from late summer after the phytoplankton bloom and displayed N:P ratios characteristic of nitrate limitation in the mixed-layer (Schuback et al., 2017). NCDs are mostly heterotrophic bacteria; therefore, while lower levels of DIN in the mixed layer could directly create a niche for diazotrophs, it is also possible that the primary productivity generated by the spring bloom may have supported NCD groups deeper in the water column (for e.g., via DOC/DOM; Bombar et al., 2016; Fonseca-Batista et al., 2019). Although additional temporal sampling is needed to confirm whether the dominant ASVs reported herein are annually reoccurring in the CAG, our results nevertheless contribute to our understanding of diazotrophic diversity within the CAG given the sparsity of previous sampling in this region (reviewed by von Friesen & Riemann, 2020). Our results especially advance the spatial coverage for *nifH* community sequencing within the CAG. Previous work has mainly come from different molecular characterization methods than those applied herein, including *nifH* clone libraries in the Baffin Bay and CAA (Blais et al., 2012), the pyrosequencing of one sample

from the Baffin Bay (Farnelid et al., 2011), and a few metagenomics samples (Salazar et al., 2019). To our knowledge the Labrador Sea has never been investigated via *nifH* amplicon community sequencing.

Previous work has shown less overlap between diazotrophic communities in the Arctic and elsewhere (Blais et al., 2012; Farnelid et al., 2011), a conclusion supported by our data as the majority of the dominant ASVs found herein were poorly represented at the ASV-level by other known NCD *nifH* reference sequences (Turk-Kubo et al., 2022). On a broader scale, our data further supports that most diazotrophs in the Arctic Ocean belong to proteobacterial members (Karlusich et al., 2021) of the phylogenetic Clusters I and III (Blais et al., 2012; Díez et al., 2012; Farnelid et al., 2011; Fernández-Méndez et al., 2016; Shiozaki et al., 2018). While on a narrower scale (e.g., at the genus-level), overlaps between our study and other marine-derived *nifH* signatures previously found in the Arctic included *Bradyrhizobium*, *Vibrio*, *Desulfuromonas*, *Sulfuricurvum*, *Arcobacter*, and Desulfobulbaceae (Blais et al., 2012; Díez et al., 2012; Fernández-Méndez et al., 2016; Jabir et al., 2021; Salazar et al., 2019). The latter suggests that within the broader Clusters I & III, there are consistent genera being recovered via *nifH* amplicon sequencing from within the marine Arctic sector. Note that diazotroph Cluster I generally includes the Mo and V containing nitrogenases, while Cluster III includes the anaerobic diazotrophs (Riemann et al., 2010; Zehr et al., 2003).

Although the potential for lateral gene transfer and gene duplication requires that taxonomic assignments derived solely from *nifH* to be interpreted with some caution (Riemann et al., 2010; Zehr et al., 2003), it is nevertheless still useful to consider the known ecologies for the taxa observed. In the Labrador Sea, we found that important taxa included several ASVs belonging to Burkholderiales, as well as *Vitreoscilla* and *Bradyrhizobium*. The group Burkholderiales has previously been associated to the Mackenzie River and other rivers near the Beaufort Sea (Kellogg et al., 2019; Ortega-Retuerta et al., 2013), as well as sediments of Baffin Bay (Algora et al., 2015), and to the dark ocean more generally (Orcutt et al., 2011). The dominant *Vitreoscilla* recovered were most like *Vitreoscilla filiformis*, an aerobic/microaerophilic species known to grow on a variety of carbon sources (Strohl et al., 1986). Meanwhile, *Bradyrhizobium* is a genus that has been found at lower latitude (e.g., east Indian Ocean; Wu et al., 2021), as well as in the coastal NWA (Tang,

Wang, et al., 2019). Although *Bradyrhizobium* are mainly described from their symbiotic association with legumes, free-living members with genes for oxygen tolerance are becoming more appreciated in the marine environment (Tao et al., 2021). In the Baffin Bay, *S. stutzeri* (previously *Pseudomonas stutzeri*) was important. *Stutzerimonas stutzeri* is a model organism for denitrification and marine cultures of this organism can be isolated from suboxic waters in the Baltic Sea (Bentzon-Tilia et al., 2014; Lalucat et al., 2006). The species *S. stutzeri* is known to fix nitrogen and can also be found terrestrially (Desnoues et al., 2003; Zhang et al., 2019). Others have shown that bottom waters of the Baffin Bay are the site of benthic denitrification (Lehmann et al., 2019). While it is tempting to connect this microbe's presence in Baffin Bay to denitrification processes, *S. stutzeri* occurred mid-depth and just above where benthic denitrification is expected (Lehmann et al., 2019). Although still speculative at this point, our data suggest that oxygen may have influenced the *S. stutzeri* ASV's presence within the CAG (a link to oxygen was also suggested by its appearance within the Bedford Basin time-series). Hence, more research will be required to fully resolve any links to other nitrogen-related pathways or oxygen preferences within the CAG, and to also confirm whether the ASV detected between the CAG and Bedford Basin is indeed from the same organism and not a pattern resultant from a lateral gene transfer between different taxa (Zehr et al., 2003). Similar work will also be needed to fully resolve why members of the Burkholderiales were detected in conjunction with high nitrite levels during our study. In the CAA, several taxonomic groups were important including *Vibrio* sp., Desulfuromonadales, Desulfobacterales, *Sulfuricurvum*, and *Arcobacter*. The genus *Vibrio* is geographically widespread and has been previously shown to increase proportionally to rising temperature within the North Sea (Vezzulli et al., 2012). Members of this genus (e.g., *Vibrio natrigens*) are known to be fast growing and capable of fixing nitrogen under anaerobic conditions (Hoff et al., 2020). The Desulfuromonadales and Desulfobulbaceae are iron and sulfur reducers known to occur in Arctic surface sediments (Jabir et al., 2021), and were the only two groups with identical ASVs recovered in the dominant diazotrophs recently reported in the western Arctic Ocean, where they were likewise associated with sediments (Shiozaki et al., 2018). The fact that these groups were mainly identified at 60 m depth in the Bedford Basin (total depth = 71m) is consistent with a sediment-related origin for these ASVs; hence, these diazotrophs were possibly present

in all three study regions (western Arctic, eastern Arctic, and coastal NW Atlantic) due to their shared association with sediments. A final group, the Campylobacterales, were generally less well-characterized by the reference NCD catalogue (Turk-Kubo et al., 2022) and included *Sulfuricurvum* and *Arcobacter*. Both are typical members of the dark ocean (Orcutt et al., 2011) and previously reported in Arctic sediments (Jabir et al., 2021) and under Arctic sea ice (Fernández-Méndez et al., 2016), suggesting their prevalence within the Arctic marine environment. The recent discovery of ultrasmall *Arcobacter* in the Arctic and including the CAG (Karlusich et al., 2021) is consistent with our view that Arctic-associated diazotrophs within the Campylobacterales are under-studied and should be given further research attention. Note that this ultrasmall *Arcobacter* was not similar to the dominant *Arcobacter* reported herein [~82% nucleotide similarity; ~91% amino acid similarity; Turk-Kubo et al., 2022]. Considering the above lifestyle preferences, future studies should: (i) consider more closely how dissolved oxygen may play a role in diazotroph diversity within the CAG, and (ii) aim to collect more vertical profiles that encompass both the water column and sediments to further elucidate how resuspended sediments shape diazotroph diversity in the Arctic Ocean. One should note that for oxygen the role may be indirect given that dissolved oxygen concentrations will also be a function of temperature (for e.g., Supplemental Fig. 4.19b).

Most of the *nifH* variants detected herein did not share high nucleotide sequence similarity at the ASV-level to known *nifH* sequences from strains represented in NCBI, the dataset of Kapili & Dekas (2021), or the NCD catalogue of Turk-Kubo et al. (2022) [7/106 ASVs with >95% nucleotide identity to sequences in NCBI (Acland et al., 2014) and Kapili & Dekas (2021); 13/106 ASVs with >95% nucleotide identity to the NCD catalogue of Turk-Kubo et al. (2022)] (Supplemental Table 4.4). Furthermore, the CAA also was home to most of the ASVs identified as “Unknown Bacteria” (Fig. 4.2). Taken together, these results point to the Canadian Arctic Gateway, and especially the Canadian Arctic Archipelago, as ocean regions that appear particularly well suited for future cultivation studies seeking to describe novel isolates within lesser-known NCD groups.

4.5.2 UCYN-A and other Cyanobacterial Diazotrophs within the Canadian Arctic Gateway

Our few observations of cyanobacterial diazotrophs in the CAG extend previous observations of their presence in the Arctic. Although others have reported on *Nodularia*-like *nifH* within sea ice (Fernández-Méndez et al., 2016) and *Trichodesmium* in Arctic sea ice brine (Díez et al., 2012), with so few observations, it remains difficult to decipher the environmental parameters and mechanisms shaping the occurrence of cyanobacterial diazotrophs within the Arctic sector. Considering that we detected ASVs likely belonging to photoautotrophic *Trichodesmium* [station BB2], *Chroococidiopsis* [station CAA6], and *Pseudanabaena* [station LS2] relatively deeper in the water column (below 200m), it is possible that these cyanobacterial ASVs may have been transported to such depths. This is consistent with the hypothesis that transport/advection from lower latitudes may be a main source for marine cyanobacterial diazotrophs within the Arctic Ocean (von Friesen & Riemann, 2020; Zehr & Capone, 2021a). In the Labrador Sea deep convection occurs during the winter and this could provide one way for surface phytoplankton cells to be transported to deeper depths (Koelling et al., 2022). Also of relevance is that members of *Chroococidiopsis* (observed in CAA) are known to exist in colder environments (Caiola et al., 1996).

Like the other cyanobacteria detected, UCYN-A were infrequent at the time of our study and qPCR-derived abundances were detected at low levels (≤ 30 copies L⁻¹) within Baffin Bay and the Canadian Arctic Archipelago (with highest values at CAA7). Although it is possible that we may have missed a higher peak in UCYN-A earlier in the season at other stations, plausible reasons for higher UCYN-A abundances at CAA7 include this station's coastal status and its proximity to the western Arctic. Selden et al. (2022) recently showed that UCYN-A growth could be stimulated by coastal upwelling in the nearby Beaufort Sea. If coastal upwelling is a driving factor for UCYN-A growth in the Arctic, then this would be more evident at the coastal CAA sites versus the more pelagic sites in Baffin Bay and the Labrador Sea. The status of CAA7 as one of the two most westward sites sampled within the CAA (CAA6 being the other) also suggests that advection from the western Arctic may have been responsible for the higher UCYN-A abundances at this station. Although CAA6 and CAA7 are both closer to the Beaufort Sea, published work

tracking dissolved lead from the Pacific/Canadian Basin versus the Atlantic (using datasets collected during the same expedition from which our samples originate) point to CAA7 as being more influenced by Canadian Basin waters than Baffin Bay and Atlantic waters (Colombo et al., 2019). Isotopic nitrate tracer studies from the GN02 cruise also showed that transport occurred from west to east on the southern side of Parry Channel within the CAA (Lehmann et al., 2019, 2022; Sherwood et al., 2021). Hence, if advection was driving the presence of UCYN-A at CAA7, it would likely be because it was originating from the western Arctic where it is known to occur (Harding et al., 2018; Shiozaki et al., 2018). Although our findings do not suggest a high abundance for UCYN-A within the CAG at the time sampled, they nevertheless provide important spatial data that can be used to further refine models of global diazotroph distributions where polar data are critically needed (Tang & Cassar, 2019).

4.5.3 Biogeographic Division of Diazotrophs between Labrador Sea and Baffin Bay/CAA

Our analyses showed that a major division occurred between the diazotrophic communities found in the Labrador Sea versus those found at more northern sites within Baffin Bay and the CAA at the time of sampling. Contemporaneous oceanographic data collected during the expedition indicate that samples from the Labrador Sea were mainly distinguished from other samples by their higher temperatures. Although we cannot discount that the Labrador Sea may have been selective to ASVs preferring slightly higher temperatures, it is important to note that samples across the CAG during the time of sampling were all relatively cold at 0–6°C. Consequently, temperature differences may also be reflective of other ocean processes that may be more strongly influencing the separation of the two communities. In support of the latter [i.e., other ocean processes] is the fact that while both areas are considered polar biomes, the Baffin Bay/CAA and the Labrador Sea represent separate biogeochemical provinces corresponding to the Boreal Polar province (BPLR) versus the Atlantic Arctic province (ARCT), respectively (Longhurst, 1995; Reygondeau et al., 2013). BPLR includes the Arctic Ocean and nearby areas (e.g., CAA) that are influenced by its surface waters, while ARCT includes the Labrador Sea, Irminger Sea, and western Greenland Sea (Longhurst, 1995). In the CAG,

this separation can be especially distinguished by the Davis Strait (shallow area between Labrador Sea and Baffin Bay) where the shallowing bathymetry influences major currents (Belkin et al., 2009). For instance, in this area (Fig 4.1b) the West Greenland and Irminger Currents move west/southwest off of the western coast of Greenland, encircling the Labrador Sea and contributing to differences between the two regions (more northern Baffin Bay versus Labrador Sea; Lacour et al., 2015; Lehmann et al., 2019; Wu et al., 2012). Also relevant is that the Baffin Bay had dissimilarity to the CAA (Fig. 4.2). Unlike the CAA, the water column of the Baffin Bay is structured such that its surface, mid-layer (for e.g., at 700m), and bottom waters, are more so influence by Arctic waters, Atlantic waters, and isolated waters, respectively (Colombo et al., 2020; Lehmann et al., 2019). At present, the influence of currents moving through the Labrador Sea (and consequently the effects this has on the water masses present) has been studied with a greater emphasis on phytoplankton (for e.g., see Fragoso et al. (2016) and Lacour et al. (2015)). Our results would argue that in future, similar attention should be given to how these physical features may shape the diazotrophic communities within the region. In this regard, more strategic sampling over time, as well as from the Labrador and West Greenland Currents (versus the central Labrador Sea which was collected herein) would help place further spatiotemporal limits on the NCD division that we have identified for the CAG.

4.6 Conclusions

Non-cyanobacterial diazotrophs exist under a variety of conditions within the ocean (for e.g., the euphotic and aphotic zones, particles, and oxygen minimum zones; Moisaner et al., 2017; Pajares & Ramos, 2019), yet less is known about their diversity and community dynamics within the Arctic (von Friesen & Riemann, 2020). Given the urgency of climate-driven changes occurring within this region of the ocean (for e.g., decreasing in sea ice; Fernández-Gómez et al., 2018; Zehr & Capone, 2020) and the potential for diazotrophs to contribute to Arctic nitrogen fixation (Sipler et al., 2017), it is now pressing that we establish detailed baseline data for marine diazotrophs in the Arctic. Our findings help address this need and show that in this area of the Arctic there is a previously unrecognized biogeographic separation that can occur between diazotrophs within the Labrador Sea and the Baffin Bay/CAA—a separation that appears reflective of the physico-chemical features

that shape the water masses within this region of ocean (Longhurst, 1995; Reygondeau et al., 2013). Across the Arctic, non-cyanobacterial diazotrophs were predominant and associated to Clusters I and III. While we did find UCYN-A within the CAG, thus confirming its broader presence within the Arctic on the eastern side of the Canadian Arctic Archipelago, it was present only at low levels according to our qPCR assay. There is some indication that UCYN-A may have been advected to station CAA7 (one of the most westward sites we examined within the CAA), and if this is the case, then it would have originated closer to the Canadian Basin/Pacific. Within Clusters I and III, we identified dominant diazotrophs for each region within the CAG (for example, those belonging to *Bradyrhizobium* and *Vitreoscilla* in the Labrador Sea and *S. stutzeri* in Baffin Bay). Although our results cannot directly demonstrate that nitrogen-fixation was occurring, they nevertheless provide a baseline for future studies that may seek to reconcile nitrogen fixation rates with historical diazotroph occurrences and relative abundances in the region. ASVs identical to those found in our study were recovered from recently published data from the western Arctic (Shiozaki et al., 2018) and from the coastal NWA (via our own coastal time-series), and were found to belong to the Desulfobulbaceae and Desulfuromonadales (western Arctic & NWA) and to *S. stutzeri* (NWA only). Our results support the hypothesis that the former two groups may have been shared between the three regions because of their association to resuspended sediments. Overall, the majority of the dominant ASVs we identified were poorly reflected at the ASV-level in the known reference sequences for major NCDs groups globally and at lower latitudes (Delmont et al., 2021; Turk-Kubo et al., 2022). Taken together, our findings highlight that the Canadian Arctic Gateway may be a prime location for future studies on NCDs and lesser-known taxonomic groups that fall within this fraction of the ocean's microbiome.

4.7 Acknowledgements

GEOTRACES is gratefully thanked for providing the oceanographic data used in this Chapter. The GEOTRACES 2021 Intermediate Data Product (IDP2021) represents an international collaboration and is endorsed by the Scientific Committee on Oceanic Research (SCOR). The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to the IDP2021.

Furthermore, N. Lehmann is especially thanked for assistance with molecular sampling during the Arctic GEOTRACES GN02 expedition. The Bedford Basin data collected and presented herein was made possible through ongoing collaboration with the Bedford Institute of Oceanography, who is thanked for providing oceanographic data and weekly water samples. The Ocean Frontier Institute is also thanked for providing both funding (awarded to J. LaRoche) and technical support (through the OFI technical pool, which assists with weekly Bedford Basin sampling). This project was also funded through a Discovery grant (awarded to J. LaRoche) and a CGS-Doctoral award (granted to B. Robicheau) from the National Sciences and Engineering Research Council of Canada, and through a Canada Foundation for Innovation Grant (to J. LaRoche). Killam predoctoral and NS Graduate scholarships additionally supported B. Robicheau.

4.8 Author Contributions

The authorship order for a future journal submission of this Chapter will be: Brent M. Robicheau, Jennifer Tolman, Sonja Rose, Dhvani Desai, and Julie LaRoche.

B. Robicheau organized and performed quantitative molecular lab work (with input from J. Tolman), conducted all data analyses, and wrote the entire text. J. Tolman organized DNA sequencing and assisted with initial data processing. S. Rose assisted with molecular lab work. D. Desai helped establish bioinformatics protocols. J. LaRoche established the initial experimental design and provided input on the text and data interpretations.

4.9 Supplemental Information

4.9.1 Supplemental Text

4.9.1.1 Supplemental Methods 4.S1: Additional R Packages used during Analyses

The following R packages (R Core Team, 2021; RStudio Team, 2021) were implemented herein: *tidyverse* (Wickham et al., 2019), *readr* (Wickham et al., 2022), *reshape2* (Wickham, 2007), *cowplot* (Wilke, 2020), *scales* (Wickham & Seidel, 2022), *metR* (Campitelli, 2021), *BiocManager* (Morgan, 2022), *ggords* (Wang, 2017), *tibble* (Müller & Wickham, 2022), and *data.table* (Dowle & Srinivasan, 2021).

4.9.2 Supplemental Tables for Chapter 4

Table 4.2. Significance of environmental parameters calculated using *envfit* (Oksanen et al., 2022). Abbreviations: Temp = Temperature, Sal = Salinity, Fluo = Fluorescence, Oxygen = Dissolved Oxygen, Si = Silicate. This analysis is associated with Fig. 4.3.

	PC1	PC2	r2	Pr(>r)	
Depth	0.86908	-0.49467	0.1698	0.001	***
Temp	0.99776	-0.06686	0.4062	0.001	***
Sal	0.75456	-0.65623	0.4403	0.001	***
NO₃	0.3294	-0.94419	0.1998	0.001	***
Oxygen	0.21247	0.97717	0.1367	0.001	***
Fluo	-0.25412	0.96717	0.1116	0.003	**
PO₄	-0.33233	-0.94316	0.0676	0.01	**
Si	-0.60899	-0.79318	0.0362	0.082	.
NO₂	-0.53265	-0.84634	0.0242	0.165	

Significance codes: 0 | '***' = 0.001 | '**' = 0.01 | '*' = 0.05 | '.' = 0.1 | ' ' = 1

Permutation: free

Number of permutations: 999

Table 4.3. Additional alignment scores for dominant *nifH* ASVs in the CAG based on comparison to *nr/nt* database in NCBI (Acland et al., 2014) via BLAST (Johnson et al., 2008) and Kapili & Dekas (2021) via standalone BLAST (Altschul et al., 1990).

Maximum Relative Abundance	Total Samples ASV is Present	Total Reads Per ASV	Final Name given in Phylogeny and Analyses	Top NCBI <i>nr/nt</i> BLAST Hits (Coverage Pairwise Identity), Organism Name(s) for better matches	Kapili & Dekas (2021) Top BLAST Hits Coverage PI	Taxon names for top matches in Kapili & Dekas (2021) <i>nifH</i> sequence collection (those used from their reference <i>nifH</i> tree)
21.19	9	753	ASV00023 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	100 86.2	Desulfopila_sp
11.26	12	673	ASV009ac Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 88.6	Geopsychrobacter_electrodiphilus
11.42	25	881	ASV027a0 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 87.4	Desulfuromonas_sp
47.45	6	1939	ASV03ba1 Unknown Burkholderiales	99 93.52, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	99.69 89.8	Vitreoscilla, Solimonas, Burkholderiales
10.36	4	600	ASV0a013 Unknown Bacteria	No Better Match than Kapili and Dekas collection	79.38 73.5	Okeania_hirsuta, Hydrocoleum_sp
81.09	8	3483	ASV0c8df Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 83.4	Pelobacter_sp
5.87	7	477	ASV1019a Desulfuromonas sp.	No Better Match than Kapili and Dekas collection	100 85.6	Desulfuromonas_sp
26.60	5	1358	ASV10d8a Unknown Burkholderiales	100 93.85, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	100 92.9	Azohydromonas_lata
99.90	3	2086	ASV12006 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	100 84.1	Desulfopila_sp
26.60	15	2592	ASV14093 Vitreoscilla sp.	No Better Match than Kapili and Dekas collection	100 93.8	Vitreoscilla_filiformis
6.57	3	319	ASV18d78 Unknown Bacteroidia	No Better Match than Kapili and Dekas collection	94.46 75.6	Alkaliphilus_metaliredigens
5.81	14	276	ASV18e60 Vibrio sp.	100 89.85, Vibrio sp. STUT-A11	100 86.9	Vibrio_sp
19.78	6	1384	ASV19e73 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	99.69 79.1	delta_proteobacterium
6.82	3	165	ASV1a2c0 Vibrio sp.	No Better Match than Kapili and Dekas collection	100 83.4	Vibrio_tritonus
14.19	3	145	ASV1ff6e Unknown Gammaproteobacteria	99 77.91, Klebsiella sp. RM1-2	96.62 79.2	Vibrio_sp
62.67	18	4559	ASV20751 Unknown Bacteria	No Better Match than Kapili and Dekas collection	100 87.7	Kiritimatiellales_bacterium
19.85	3	477	ASV222f2 Unknown Burkholderiales	99 93.83, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	99.69 91.4	Burkholderiales_bacterium
1.52	11	125	ASV27104 Bradyrhizobium sp.	100 98.15, Bradyrhizobium sp. CCBAU 101065	100 97.8	Bradyrhizobium_sp
4.71	7	428	ASV2dce4 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	62.46 75.7	Methylococcus_sp
11.62	3	618	ASV30da6 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 88.3	Geopsychrobacter_electrodiphilus
23.52	7	1040	ASV36a96 Unknown Bacteria	No Better Match than Kapili and Dekas collection	99.69 86.7	Kiritimatiellales_bacterium
57.22	17	3910	ASV37f70 Unknown Burkholderiales	99 93.52, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	99.69 91.0	Burkholderiales_bacterium
18.83	8	1043	ASV39410 Unknown Burkholderiales	100 100, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	100 92	Ideonella_dechloratans
4.56	8	289	ASV3aa29 Vibrio sp.	No Better Match than Kapili and Dekas collection	100 82.8	Vibrio_sp
6.87	3	187	ASV3e02a Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	98.77 79.8	Desulfovibrio_frigidus
1.13	3	71	ASV3f2a0 Sulfurospirillum sp.	No Better Match than Kapili and Dekas collection	99.69 88.3	Sulfurospirillum_arcachonense
6.21	7	263	ASV41648 Unknown Sphingomonadaceae	No Better Match than Kapili and Dekas collection	100 94.2	Novosphingobium_sp
61.29	8	899	ASV4650f Vibrio sp.	No Better Match than Kapili and Dekas collection	100 85.6	Vibrio_sp
37.71	14	1907	ASV46e90 Vibrio sp.	No Better Match than Kapili and Dekas collection	99.69 85.2	Vibrio_sp
12.25	5	383	ASV4911a Vibrio sp.	100 89.23, Vibrio sp. STUT-A11	100 86.9	Vibrio_sp
8.02	13	801	ASV4abe3 Vibrio sp.	100 89.54, Vibrio sp. STUT-A11	100 87.2	Vibrio_sp
22.17	18	1872	ASV523fe Vitreoscilla sp.	No Better Match than Kapili and Dekas collection	100 94.5	Vitreoscilla_filiformis
54.39	3	2953	ASV54ae9 Pelobacter sp.	No Better Match than Kapili and Dekas collection	98.77 78.8	Celerinatantimonas_diazotrophica
10.67	13	548	ASV55e6e Arcobacter sp.	No Better Match than Kapili and Dekas collection	100 87.2	Arcobacter_sp
31.09	7	1498	ASV5b667 Vitreoscilla sp.	No Better Match than Kapili and Dekas collection	100 94.8	Vitreoscilla_filiformis
7.61	9	1203	ASV5c36e Unknown Nitrospiraceae	99 83.38, Uncultured Bradyrhizobium sp.	60.62 74.9	Sporomusa_sphaeroides
14.04	4	395	ASV5f845 Unknown Burkholderiales	100 93.54, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	100 92.6	Azohydromonas_lata
6.36	3	215	ASV60788 Vibrio sp.	No Better Match than Kapili and Dekas collection	100 86.5	Amphritea_atlantica
10.41	12	734	ASV61c77 Unknown Burkholderiales	100 94.15, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	100 92.6	Azohydromonas_lata
31.70	27	2129	ASV62490 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	62.15 75.9	Desulfotomaculum_ferrireducens
19.88	4	592	ASV63ee1 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 87.7	Desulfuromonas_sp
25.17	3	1058	ASV67b6f Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 87.7	Desulfuromonas_sp
9.97	3	273	ASV67b96 Unknown Bacteria	No Better Match than Kapili and Dekas collection	100 84.9	Kiritimatiellales_bacterium
1.53	3	75	ASV6a08e Uliginosibacterium sp.	No Better Match than Kapili and Dekas collection	100 87.1	Betaproteobacteria_bacterium
1.92	10	132	ASV6cb4d Unknown Burkholderiales	No Better Match than Kapili and Dekas collection	100 92.3	Ideonella_sp
5.92	4	327	ASV6ce95 Unknown Bacteria	No Better Match than Kapili and Dekas collection	88.31 74	Lablibacter_marinus
18.66	6	919	ASV6d089 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	99.69 79.8	delta_proteobacterium
19.48	8	737	ASV6d6c1 Vitreoscilla sp.	No Better Match than Kapili and Dekas collection	100 93.5	Vitreoscilla_filiformis
48.00	6	3155	ASV6fc48 Unknown Burkholderiales	100 94.15, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	100 92.6	Ideonella_dechloratans
5.04	5	303	ASV7062b Vibrio sp.	No Better Match than Kapili and Dekas collection	99.69 86.4	Vibrio_sp
17.55	21	1361	ASV708e2 Vibrio sp.	100 84.92, Vibrio natriegens strain PWH3a	99.69 84.0	Vibrio_tritonus
9.40	3	208	ASV70924 Unknown Burkholderiales	100 91.69, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	100 89.5	Vitreoscilla_filiformis, Ideonella_sp
18.97	11	691	ASV71ce2 Vibrio sp.	No Better Match than Kapili and Dekas collection	99.69 83.1	Vibrio_diazotrophicus
92.06	3	3783	ASV728c6 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	98.77 85.7	Desulfuromonas_soudanensis
27.41	3	873	ASV78e87 Unknown Epsilonproteobacteria	100 88, Sulfurimonas sp. B2	100 86.8	Arcobacter_anaerophilus
24.19	5	1722	ASV7e886 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 84.4	Pelobacter_carbinolicus
94.14	4	1164	ASV7fb2c Unknown Burkholderiales	No Better Match than Kapili and Dekas collection	100 91.4	Burkholderiales_bacterium
69.49	16	6893	ASV81bee Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	61.54 77.5	Desulfotomaculum_aeronauticum
16.56	3	687	ASV88dee Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	98.46 77.1	Halodesulfovibrio_aestuarii, Desulfovibrio_desulfuricans

Table continues onto next page...

Maximum Relative Abundance	Total Samples ASV is Present	Total Reads Per ASV	Final Name given in Phylogeny and Analyses	Top NCBI nr/nt BLAST Hits (Coverage Pairwise Identity), Organism Name(s) for better matches	Kapili & Dekas (2021) Top BLAST Hits Coverage PI	Taxon names for top matches in Kapili & Dekas (2021) <i>nifH</i> sequence collection (those used from their reference <i>nifH</i> tree)
10.18	8	372	ASV8ed64 Unknown Epsilonproteobacteria	No Better Match than Kapili and Dekas collection	100 86.8	Arcobacter_anaerophilus
6.39	4	290	ASV8f6ef Vibrio	No Better Match than Kapili and Dekas collection	100 85.3	Vibrio_sp
98.57	27	13359	ASV93fd9 Bradyrhizobium sp.	100 99.38, Bradyrhizobium denitrificans LMG 8443	100 98.8	Bradyrhizobium_sp
69.01	10	7765	ASV942de Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	99.69 76.5	Deltaproteobacteria_bacterium
6.25	3	364	ASV94fbf Arcobacter sp.	No Better Match than Kapili and Dekas collection	100 87.4	Arcobacter_sp
13.06	3	518	ASV995a9 Unknown Betaproteobacteria	100 93.54, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	100 93.2	Azohydromonas_lata
9.40	13	829	ASV9cb21 Sulfuriferula sp.	99 85.45, Sulfuriferula nivalis	100 81.3	Pseudolabrys_sp, Ideonella_sp
2.92	7	352	ASVa14c3 Unknown Deltaproteobacteria	No Better Match than Kapili and Dekas collection	100 89.2	Deltaproteobacteria_bacterium
7.29	5	233	ASVa2940 Unknown Bacteria	No Better Match than Kapili and Dekas collection	68.92 73.6	Acidithiobacillus_ferrovorum
4.52	18	445	ASVa5810 Unknown Bacteria	No Better Match than Kapili and Dekas collection	100 79	Mangrovibacterium_diazotrophicum
5.55	7	472	ASVa6c8b Unknown Deltaproteobacteria	No Better Match than Kapili and Dekas collection	100 89.5	Deltaproteobacteria_bacterium
25.46	3	430	ASVa71b5 Unknown Bacteria	No Better Match than Kapili and Dekas collection	94.15 71.7	Desulforhopalus_sp
68.30	7	1394	ASVa8726 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	100 84.0	Desulfopila_sp
9.05	3	165	ASVab855 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	100 83.5	Desulfopila_sp
100.00	29	46450	ASVb0b18 Stutzerimonas stutzeri	100 91.69, Pseudomonas oryzae KCTC 32247	100 100	Pseudomonas_stutzeri, Pseudomonas_sp
2.81	7	272	ASVb20cf Unknown Proteobacteria	No Better Match than Kapili and Dekas collection	99.38 80.1	Vibrio_sp, Vibrio_natriegens
14.98	3	406	ASVb22ca Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 84.1	Pelobacter_carbinolicus
76.79	20	5728	ASVb431e Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	98.77 84.7	Desulfopila_aestuarii
5.68	5	148	ASVb4bd6 Vibrio sp.	No Better Match than Kapili and Dekas collection	100 84.7	Vibrio_sp
13.55	6	852	ASVb9003 Unknown Burkholderiales	99 93.83, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	99.69 91.4	Burkholderiales_bacterium, Aquabacterium_sp
34.79	12	7552	ASVbbc22 Unknown Bacteria	No Better Match than Kapili and Dekas collection	100 81	Nitrospiraceae_bacterium
4.34	8	336	ASVbbdcf Unknown Bacteria	No Better Match than Kapili and Dekas collection	99.38 75.9	Candidatus_Galacturoniibacter_soehngeni
20.32	6	673	ASVc07e Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 83.8	Pelobacter_carbinolicus
8.24	8	481	ASVc2a07 Vibrio sp.	No Better Match than Kapili and Dekas collection	100 82.8	Vibrio_sp
3.39	11	493	ASVc5adf Sulfuricurvum sp.	No Better Match than Kapili and Dekas collection	100 94.5	Sulfuricurvum_sp
97.00	3	4707	ASVc8025 Unknown Burkholderiales sp.	99 93.21, Rheinheimera hassiensis E48, Sphingomonas paucimobilis, Stenotrophomonas maltophilia KNUC170	99.69 91.4	Burkholderiales_bacterium
17.32	3	817	ASVc8d56 Unknown Desulfovibrionaceae	No Better Match than Kapili and Dekas collection	98.77 76.4	Acetobacterium_dehalogenans
4.09	18	520	ASVca89c Vibrio sp.	100 85.37, Agarivorans sp. B27047, Agarivorans albus JCM 21469	100 84.7	Agarivorans_albus
5.23	5	281	ASVcb436 Vibrio sp.	No Better Match than Kapili and Dekas collection	99.69 84.9	Vibrio_sp
3.87	9	386	ASVccdd5 Unknown Helicobacteraceae	No Better Match than Kapili and Dekas collection	100 81.5	Sulfurospirillum_arcachonense
76.77	3	1579	ASVcfb81 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	98.77 85.4	Desulfuromonas_soudanensis
8.76	11	766	ASVd5064 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	62.46 75.2	Methylomicrobium_sp
10.06	4	460	ASVd58b4 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	100 79.6	delta_proteobacterium
23.64	7	1380	ASVd59b4 Unknown Desulfobulbaceae	No Better Match than Kapili and Dekas collection	94.46 73.3	Acetobacterium_tundrae
1.30	3	35	ASVd74b8 Vibrio sp.	No Better Match than Kapili and Dekas collection	99.69 82.8	Vibrio_diazotrophicus
12.09	5	723	ASVd7f6f Unknown Burkholderiales	100 94.15, Stenotrophomonas maltophilia KNUC170, Sphingomonas paucimobilis	100 92.3	Ideonella_dechloratans, Azohydromonas_lata
32.38	3	1953	ASVdec69 Unknown Desulfobacteraceae	No Better Match than Kapili and Dekas collection	99.08 74.2	Pseudodesulfobivrio_profundus
15.42	19	1363	ASVe1a8a Vibrio sp.	100 89.23, Vibrio sp. STUT-A11	100 87.5	Vibrio_sp
29.84	11	1216	ASVe569b Unknown Bacteria	No Better Match than Kapili and Dekas collection	100 86.8	Kiritimatiales_bacterium
14.86	6	956	ASVe90eb Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 83.5	Pelobacter_carbinolicus
6.12	3	266	ASVeca8e Arcobacter sp.	No Better Match than Kapili and Dekas collection	100 86.8	Arcobacter_sp
7.25	6	366	ASVed3c9 Hydrogenophaga sp.	99 95.98, Ideonella dechloratans strain CCLUG 30977	99.69 95.4	Hydrogenophaga_flava
19.08	16	2289	ASVef783 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 84.1	Pelobacter_carbinolicus
9.31	25	2013	ASVef8b2 Unknown Desulfovibrionaceae	No Better Match than Kapili and Dekas collection	98.77 76.8	Acetobacterium_dehalogenans
29.85	6	2708	ASVefce1 Unknown Desulfuromonadales	No Better Match than Kapili and Dekas collection	100 85.9	Malonomonas_rubra
10.23	4	607	ASVF31c4 Unknown Gammaproteobacteria	No Better Match than Kapili and Dekas collection	98.77 87.9	Sulfurivermis_fontis
5.07	10	394	ASVfb17c Vibrio sp.	No Better Match than Kapili and Dekas collection	99.69 83.6	Vibrio_sp

Table 4.4. Alignment scores for best matches between CAG *nifH* ASV sequences versus the dominant *nifH* OTUs reported for the western Arctic Ocean (Shiozaki et al., 2018), as well as to the *nifH* from low-latitude MAGs reported by Delmont et al. (2021), and to the catalogue of well-known NCDs reported by Turk-Kubo et al. (2022). Yes and No scores indicate whether the ASV in our study was recovered within the other datasets aforementioned.

ASV Information		Recovery Statistics: n = Number of Sequences ASVs are being compared against					Turk-Kubo et al. 2022 DNA Catalog			
Final ASV Name Used in Tree and Analyses	In Bedford Basin (at 100% PI, 100% coverage)	Delmont et al. 2021 [Identical ASVs] 0/49 nifH MAGs Delmont et al. 2021 (n = 49); 95% coverage, 100% PI	Delmont et al. 2021 [Lower Threshold] 1/49 nifH MAGs Delmont et al. 2021 (n = 49); 95% coverage >92% PI	Shiozaki et al. 2018 [Identical ASVs] 4 ASVs Recovered Shiozaki et al. 2018 (n = 45); 99.69% coverage, 100% PI	Overall Status as Recovered in NCD catalogue? 47/106 = 44% ASVs recovered as belonging to major NCD groups, based on 97% amino acid similarity	Turk-Kubo et al. 2022 DNA Collection (n = 189); Sorted by E value; Name of Top Alignment Hit	Similar Sequences Recovered at >95% Cov. & >92% IDT = 14/106 CAG nifH ASVs = 13%	Similarity (%) for match when converted to Amino Acids	Turk-Kubo et al. 2022, smaller set of NCDs only reported using amino acids (n = 16); >97%	
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
					Turk-Kubo et al. 2022 DNA Catalog					
ASV0002a Unknown Desulfobulbaceae	No	No	No	No	Yes	DOCY3 / TARA_PSW_86_MAG_00080	66.15% cov / 99.38% cov	90.2% PI / 80.3% PI	100% & 91.59%	94 (12/200) 90.97% cov 86.15% PI
ASV009ac Unknown Desulfuromonadales	Yes	No	No	No	Yes	ALV82198.1 & Chuk-Shio-otu1	98.77% cov	88.8% PI	99.99%	<97% AA similarity
ASV027a0 Unknown Desulfuromonadales	Yes	No	No	Yes; Otu0004	Yes	NB3-Pelobacter	98.46% cov	95.6% PI	99.99%	<97% AA similarity
ASV0c8f8 Unknown Desulfuromonadales	Yes	No	No	No	Yes	Chuk-Shio-otu1 & ALV82198.1	100% cov	98.2% PI	100%	<97% AA similarity
ASV12006 Unknown Desulfobulbaceae	Yes	No	No	No	Yes	DOCY3 / P4_Loesch	67.38% cov / 98.77% cov	90.4% PI / 74.9% PI	100% & 86.92%	94 (12/200) 90.97% cov 86.15% PI
ASV139c73 Unknown Desulfobulbaceae	No	No	No	No	Yes	No Hits	No Hits	No Hits	N/A	94 (12/200) 90.97% cov 86.15% PI
ASV21704 Bradiyrhizobium sp.	Yes	No	No	No	Yes	ALV82265.1 & MH144511	100% cov	94.5% PI	99.99%	98 (12/200) 90.97% cov 86.15% PI
ASV2d64 Unknown Desulfobulbaceae	Yes	No	No	No	Yes	TARA_PSE_93_MAG_00025	93.85% cov	78.7% PI	93.07%	94 (12/200) 90.97% cov 86.15% PI
ASV30d66 Unknown Desulfuromonadales	Yes	No	No	Yes; Otu0023	Yes	ALV82198.1 & Chuk-Shio-otu1 & Chuk-Shio-otu4	98.77% cov	88.8% PI	99.99%	<97% AA similarity
ASV37770 Unknown Burkholderiales	No	No	No	No	Yes	ACH99083.1	98.77% cov	88.2% PI	99.99%	<97% AA similarity
ASV3e02a Unknown Desulfobulbaceae	No	No	No	No	Yes	No Hits	No Hits	No Hits	N/A	94 (12/200) 90.97% cov 86.15% PI
ASV41648 Unknown Sphingomonadales	No	No	Yes; AON_82_MAG_70; 100% 93.6%	No	Yes	TARA_AON_82_MAG_00070	100% cov	93.6% PI	98.13%	94 (12/200) 90.97% cov 86.15% PI
ASV61c77 Unknown Burkholderiales	No	No	No	No	Yes	ACH99083.1	96.62% cov	88.3% PI	100%	<97% AA similarity
ASV62490 Unknown Desulfobulbaceae	No	No	No	No	Yes	TARA_PSW_86_MAG_00080	99.69% cov	80.1% PI	91.59%	94 (12/200) 90.97% cov 86.15% PI
ASV63ee1 Unknown Desulfuromonadales	Yes	No	No	No	Yes	Chuk-Shio-otu4	98.77% cov	88.8% PI	99.99%	<97% AA similarity
ASV6766f Unknown Desulfuromonadales	Yes	No	No	No	Yes	Chuk-Shio-otu4	98.77% cov	88.5% PI	99.99%	<97% AA similarity
ASV6c6dd Unknown Burkholderiales	No	No	No	No	Yes	ACH33106.1	99.69% cov	90.1% PI	100%	94 (12/200) 90.97% cov 86.15% PI
ASV6d89 Unknown Desulfobulbaceae	No	No	No	No	Yes	No Hits	No Hits	No Hits	N/A	94 (12/200) 90.97% cov 86.15% PI
ASV6f648 Unknown Burkholderiales	Yes	No	No	No	Yes	EVHVF	99.38% cov	87.3% PI	99.33%	<97% AA similarity
ASV70e2b Unknown Vibrrio sp.	No	No	No	No	Yes	8693316	100% cov	88% PI	99.15%	<97% AA similarity
ASV70924 Unknown Burkholderiales	Yes	No	No	No	Yes	Antarc-Shio-SV003	100% cov	86.2% PI	99.33%	<97% AA similarity
ASV728c6 Unknown Desulfuromonadales	Yes	No	No	No	Yes	Chuk-Shio-otu4	100% cov	98.5% PI	100%	<97% AA similarity
ASV76886 Unknown Desulfuromonadales	No	No	No	No	Yes	NB3-Pelobacter	99.69% cov	99.7% PI	100%	<97% AA similarity
ASV776c3 Unknown Burkholderiales	No	No	No	No	Yes	ACH99083.1	98.77%	84.50%	99.99%	<97% AA similarity
ASV81bee Unknown Desulfobulbaceae	Yes	No	No	No	Yes	DOCY3 / TARA_PSW_86_MAG_00080	67.08% cov / 99.38% cov	90.8% PI / 81.8% PI	100% & 91.59%	94 (12/200) 90.97% cov 86.15% PI
ASV8d6e Unknown Desulfobulbaceae	Yes	No	No	Yes; Otu0008	Yes	No Hits	No Hits	No Hits	N/A	94 (12/200) 90.97% cov 86.15% PI
ASV93f9 Bradiyrhizobium sp.	Yes	No	No	No	Yes	CAL79071.1	100% cov	98.2% PI	100%	<97% AA similarity
ASV942de Unknown Desulfobulbaceae	Yes	No	No	No	Yes	DOCY3	67.38% cov	91.3% PI	100%	94 (12/200) 90.97% cov 86.15% PI
ASV98726 Unknown Desulfobulbaceae	No	No	No	No	Yes	DOCY3 / P4_Loesch	67.38% cov / 98.77% cov	90.0% PI / 75.2% PI	100% 86.92%	94 (12/200) 90.97% cov 86.15% PI
ASV98855 Unknown Desulfobulbaceae	No	No	No	No	Yes	DOCY3 / P4_Loesch	67.38% cov / 98.77% cov	90.9% PI / 74.6% PI	100% 86.92%	94 (12/200) 90.97% cov 86.15% PI
ASV9b018 Stutzerimonas stutzeri	Yes	No	No	No	Yes	BAL354 & BAL361	99.69% cov	100% PI	100%	<97% AA similarity
ASV9d2ca Unknown Desulfuromonadales	Yes	No	No	No	Yes	NB3-Pelobacter	99.69% cov	100% PI	100%	<97% AA similarity
ASV9d31e Unknown Desulfobulbaceae	Yes	No	No	No	Yes	TARA_PSW_86_MAG_00080	97.54% cov	81.1% PI	93.33%	94 (12/200) 90.97% cov 86.15% PI
ASV9d666 Unknown Vibrrio sp.	No	No	No	No	Yes	8693316	100% cov	88.9% PI	99.99%	<97% AA similarity
ASV9d903 Unknown Burkholderiales	No	No	No	No	Yes	ACH99083.1	98.77% cov	85.1% PI	99.99%	<97% AA similarity
ASV9df7e Unknown Desulfuromonadales	Yes	No	No	Yes; Otu0001	Yes	ALV82198.1 & Chuk-Shio-otu1	100% cov	99.7% PI	100%	<97% AA similarity
ASV9e025 Unknown Burkholderiales sp.	No	No	No	No	Yes	ACH99083.1	98.77% cov	85.1% PI	99.99%	<97% AA similarity
ASV9fc81 Unknown Desulfuromonadales	Yes	No	No	No	Yes	Chuk-Shio-otu4	100% cov	98.5% PI	100%	<97% AA similarity
ASV9f064 Unknown Desulfobulbaceae	No	No	No	No	Yes	TARA_PSE_93_MAG_00025	93.85% cov	79.9% PI	93.07%	94 (12/200) 90.97% cov 86.15% PI
ASV9f884 Unknown Desulfobulbaceae	Yes	No	No	No	Yes	DOCY3	64.31% cov	91.9% PI	100%	94 (12/200) 90.97% cov 86.15% PI
ASV9f954 Unknown Desulfobulbaceae	No	No	No	No	Yes	No Hits	No Hits	No Hits	N/A	94 (12/200) 90.97% cov 86.15% PI
ASV9f76f Unknown Burkholderiales	No	No	No	No	Yes	ACH99083.1	96.62% cov	88.6% PI	100%	<97% AA similarity
ASV9e90b Unknown Desulfuromonadales	No	No	No	No	Yes	ALV82198.1 & Chuk-Shio-otu1	100% cov	99.4% PI	100%	<97% AA similarity
ASV9d3c9 Hydrogenophaga sp.	No	No	No	No	Yes	AFD32190.1	99.69% cov	90.7% PI	99.15%	<97% AA similarity
ASV9f783 Unknown Desulfuromonadales	Yes	No	No	No	Yes	ALV82198.1 & Chuk-Shio-otu1	100% cov	100% PI	100%	<97% AA similarity
ASV9fca1 Unknown Desulfuromonadales	Yes	No	No	No	Yes	Chuk-Shio-otu4	100% cov	96.3% PI	100%	<97% AA similarity
ASVf31c4 Unknown Gammaproteobacteria	No	No	No	No	Yes	Gamma_2	98.46% cov	90% PI	99.33%	<97% AA similarity
ASV03ba1 Unknown Burkholderiales	Yes	No	No	No	No	CAL79071.1	99.38% cov	84.3% PI	95.33%	<97% AA similarity
ASV0a013 Unknown Bacteria	No	No	No	No	No	8693316	63.69% cov	75.9% PI	79.71%	<97% AA similarity
ASV1019a Desulfuromonas sp.	No	No	No	No	No	CB914H4	99.69% cov	88% PI	96.30%	<97% AA similarity
ASV1008a Unknown Burkholderiales	Yes	No	No	No	No	Antarc-Shio-SV003	100% cov	87.4% PI	94.44%	<97% AA similarity
ASV14093 Vitreoscilla sp.	Yes	No	No	No	No	TARA_AON_82_MAG_00263	100% cov	88.7% PI	94.44%	<97% AA similarity
ASV18f79 Unknown Bacteroidia	Yes	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASV18e60 Unknown Vibrrio sp.	No	No	No	No	No	P7_Loesch	98.77% cov	89.4% PI	93.46%	<97% AA similarity

Table continues onto next page...

ASV Information		Recovery Statistics: n = Number of Sequences ASVs are being compared against					Turk-Kubo et al. 2022 DNA Catalog			
Final ASV Name Used in Tree and Analyses	In Bedford Basin (at 100% PI, 100% coverage)	Delmont et al. 2021 [Identical ASVs]	Delmont et al. 2021 [Lower Threshold]	Shiozaki et al. 2018 [Identical ASVs]	Overall Status as Recovered in NCD catalogue?	Turk-Kubo et al. 2022 DNA Catalog				
		0/49 nHf MAGs Delmont et al. 2021 (n = 49); ≥99.69% coverage, 100% PI	1/49 nHf MAGs Delmont et al. 2021 (n = 49); >95% coverage >92% PI	4 ASVs Recovered Shiozaki et al. 2018 (n = 45); ≥99.69% coverage, 100% PI	47/106 = 44% ASVs recovered as belonging to major NCD groups, based on 97% amino acid similarity	Turk-Kubo et al. 2022 DNA Collection (n = 189), Sorted by E value; Name of Top Alignment Hit	Similar Sequences Recovered at ≥95% Cov. & >92% ID7 = 14/106 CAG nHf ASVs = 13%	Nucleotide Pairwise Identity (PI)	Similarity (%) for match when converted to Amino Acids	Turk-Kubo et al. 2022, smaller set of NCDs only reported using amino acids (n = 16); ≥97%
ASV1a2c0 <i>Vibrio</i> sp.	No	No	No	No	No	8693316	99.69% cov	86.4% PI	94.44%	<97% AA similarity
ASV1ff6e Unknown Gammaproteobacteria	No	No	No	No	No	HIPOt_1	58.77% cov	83.8% PI	98.41%	<97% AA similarity
ASV20751 Unknown Bacteria	Yes	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASV22272 Unknown Burkholderiales	No	No	No	No	No	CA_79071.1	99.38% cov	84.6% PI	95.33%	<97% AA similarity
ASV36a8e Unknown Bacteria	Yes	No	No	No	No	PSE_id-1140575	97.54% cov	79.5% PI	84.76%	<97% AA similarity
ASV39410 Unknown Burkholderiales	Yes	No	No	No	No	Antaic-Shio-SV003	100% cov	86.5% PI	95.37%	<97% AA similarity
ASV3aa29 <i>Vibrio</i> sp.	No	No	No	No	No	P7_Loesch	98.46% cov	86.0% PI	91.51%	<97% AA similarity
ASV3f2a0 <i>Sulfurospirillum</i> sp.	No	No	No	No	No	8173703	99.69% cov	86.7% PI	91.67%	<97% AA similarity
ASV4650f <i>Vibrio</i> sp.	Yes	No	No	No	No	Azo_Seaver & Vib_Sever	99.08% cov	88% PI	95.33%	<97% AA similarity
ASV46e90 <i>Vibrio</i> sp.	No	No	No	No	No	TARA_PSE_93_MAG_00027	99.69% cov	86% PI	93.52%	<97% AA similarity
ASV4921a <i>Vibrio</i> sp.	Yes	No	No	No	No	P7_Loesch	98.77% cov	89.4% PI	93.46%	<97% AA similarity
ASV44ba3 <i>Vibrio</i> sp.	Yes	No	No	No	No	P7_Loesch	98.77% cov	89.8% PI	93.46%	<97% AA similarity
ASV523fe <i>Vitreoscilla</i> sp.	No	No	No	No	No	TARA_AON_82_MAG_00263	100% cov	89.6% PI	94.44%	<97% AA similarity
ASV54ae9 <i>Pelobacter</i> sp.	No	No	No	No	No	ABP37981.1	97.85% cov	80.9% PI	86.67%	<97% AA similarity
ASV55e6e <i>Arcobacter</i> sp.	Yes	No	No	No	No	Arc_Seaver	99.69% cov	83.7% PI	89.81%	<97% AA similarity
ASV5667 <i>Vitreoscilla</i> sp.	Yes	No	No	No	No	TARA_AON_82_MAG_00263	100% cov	89.3% PI	94.44%	<97% AA similarity
ASV5c3e Unknown Nitrospiraceae	No	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASV5f845 Unknown Burkholderiales	Yes	No	No	No	No	Antaic-Shio-SV003	100% cov	87.7% PI	94.44%	<97% AA similarity
ASV60788 <i>Vibrio</i> sp.	No	No	No	No	No	Npac-Grad-denov6 & TARA_PSE_93_MAG_00126	89.54% cov	84.9% PI	94.85%	<97% AA similarity
ASV67b9e Unknown Bacteria	No	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASV6a08e <i>Uliginosibacterium</i> sp.	No	No	No	No	No	AFD32190.1	99.69% cov	82.7% PI	93.46%	<97% AA similarity
ASV6ce95 Unknown Bacteria	No	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASV6d6c1 <i>Vitreoscilla</i> sp.	No	No	No	No	No	Antaic-Shio-SV003 & TARA_AON_82_MAG_00263	100% cov	88.3% PI	96.3% & 94.44%	<97% AA similarity
ASV708e2 <i>Vibrio</i> sp.	No	No	No	No	No	8693316	100% cov	85.5% PI	94.44%	<97% AA similarity
ASV71ce2 <i>Vibrio</i> sp.	No	No	No	No	No	P7_Loesch	98.46% cov	84.4% PI	91.51%	<97% AA similarity
ASV78e87 Unknown Epsilonproteobacteria	No	No	No	No	No	8693316	51.69% cov	85.2% PI	92.73%	<97% AA similarity
ASV8ed64 Unknown Epsilonproteobacteria	No	No	No	No	No	8173703	100% cov	85.5% PI	90.74%	<97% AA similarity
ASV8f6ef <i>Vibrio</i> sp.	No	No	No	No	No	Azo_Seaver & Vib_Sever	99.08% cov	87.4% PI	95.33%	<97% AA similarity
ASV9fff <i>Arcobacter</i> sp.	No	No	No	No	No	Arc_Seaver	99.69% cov	83.0% PI	88.89%	<97% AA similarity
ASV995a9 Unknown Betaproteobacteria	Yes	No	No	No	No	TARA_AON_82_MAG_00263	100% cov	89.3% PI	96.30%	<97% AA similarity
ASV9a21 <i>Sulfuriferus</i> sp.	No	No	No	No	No	TARA_PSE_93_MAG_00003	99.38% cov	78.4% PI	89.72%	<97% AA similarity
ASVa14c3 Unknown Deltaproteobacteria	No	No	No	No	No	Npac-Shio-otu00004 & Sof-Sao-SV003	60.92% cov	74.7% PI	75.38%	<97% AA similarity
ASVa2940 Unknown Bacteria	Yes	No	No	No	No	8693316	15.69% cov	94.1% PI	100.00%	<97% AA similarity
ASVa5810 Unknown Bacteria	Yes	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASVa6c8 Unknown Deltaproteobacteria	No	No	No	No	No	Cluster-3	99.69% cov	75.9% PI	79.44%	<97% AA similarity
ASVa71d5 Unknown Bacteria	No	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASVb2dcf Unknown Proteobacteria	No	No	No	No	No	Gamma_1	98.46% cov	84.1% PI	94.34%	<97% AA similarity
ASVb2c2 Unknown Bacteria	No	No	No	No	No	CAA31666.1	60.62% cov	84.9% PI	92.31%	<97% AA similarity
ASVbbcf Unknown Bacteria	No	No	No	No	No	TARA_AON_82_MAG_00083	99.69% cov	83.3% PI	95.37%	<97% AA similarity
ASVc2a07 <i>Vibrio</i> sp.	No	No	No	No	No	ABP37981.1	100% cov	84.4% PI	92.59%	<97% AA similarity
ASVc5adf <i>Sulfuricum</i> sp.	No	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASVc65e Unknown Desulfobiontrionaceae	No	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASVca8c <i>Vibrio</i> sp.	No	No	No	No	No	8693316	100% cov	86.5% PI	96.30%	<97% AA similarity
ASVcb36 <i>Vibrio</i> sp.	No	No	No	No	No	TARA_PSE_93_MAG_00027	99.69% cov	83.7% PI	93.52%	<97% AA similarity
ASVcdd5 Unknown Helicobacteraceae	No	No	No	No	No	P6_Loesch	94.15% cov	79.9% PI	81.90%	<97% AA similarity
ASVd74b8 <i>Vibrio</i> sp.	No	No	No	No	No	P7_Loesch	98.46% cov	84.1% PI	91.51%	<97% AA similarity
ASVdec69 Unknown Desulfobacteraceae	Yes	No	No	No	No	TARA_PSW_86_MAG_00010	99.08% cov	74.2% PI	81.31%	<97% AA similarity
ASVe1a8a <i>Vibrio</i> sp.	Yes	No	No	No	No	P7_Loesch	98.77% cov	89.4% PI	93.46%	<97% AA similarity
ASVe5b8 Unknown Bacteria	Yes	No	No	No	No	SCW_Cole_denov6	95.69% cov	85.2% PI	94.17%	<97% AA similarity
ASVea3e <i>Arcobacter</i> sp.	No	No	No	No	No	TARA_PSE_93_MAG_00078	100% cov	86.9% PI	92.59%	<97% AA similarity
ASVf8b2 Unknown Desulfobiontrionaceae	Yes	No	No	No	No	No Hits	No Hits	No Hits	N/A	<97% AA similarity
ASVf17c <i>Vibrio</i> sp.	Yes	No	No	No	No	Azo_Seaver & Vib_Sever	99.08% cov	84.8% PI	95.33%	<97% AA similarity

Table 4.5. Individual ASV scores for multi-level pattern analyses. Scores for each site grouping category are shown per dominant ASV. Analysis is limited to those ASVs recovered in ≥ 7 Canadian Arctic Gateway (GN02 cruise) samples. Only statistically significant p -values ≤ 0.05 (at $\alpha = 0.05$) were used to assign an ASV to a given category of samples. Please see the caption of Table 4.1 for a full description of the site categories. Ocean data are from *IDP2021* (GEOTRACES Intermediate Data Product Group, 2021) and Colombo et al. (2020, 2021, 2022). Highlighted rows (grey) provide summaries of the unhighlighted rows (white); such summaries are given for ASVs that belong to overlapping taxonomic groups.

ASV Name	Region	S	D	T	Sa	F	O ₂	NO ₂	NO ₃	PO ₄	Si	Di-Fe	TP-Fe	TP-V
Bradyrhizobium spp.	K	Sm/-	De/-	H	H/-	-	H	-	L/-	L	L	L?/-	L?/-	L?/-
Bradyrhizobium sp. - ASV27104	K	-	De	H	-	-	H	-	L	L	L	-	-	-
Bradyrhizobium sp. - ASV93f49	K	Sm	-	H	H	-	H	-	-	L	L	L?	L?	L?
UNKN Sphingomonadaceae - ASV41648	-	Sm	-	-	-	-	-	-	L	-	-	-	-	-
Sulfuriferula sp. - ASV9cb21	CAA	-	Su	L	L	-	H	-	L	L	L	-	H	H
UNKN Burkholderiales spp.	K/LS	Sm/-	-	H	H/-	-	H/-	-	-	L/-	L/-	L/-	-	-
UNKN Burkholderiales - ASV37f70	K/LS	-	-	H	H	-	H	-	-	-	L	L	-	-
UNKN Burkholderiales - ASV39410	K	Sm	-	H	-	-	H	-	-	L	-	-	-	-
UNKN Burkholderiales - ASV61c77	K	-	-	H	-	-	H	-	-	-	L	-	-	-
UNKN Burkholderiales - ASV6cb4d	K/LS	-	-	H	-	-	-	-	-	-	-	-	-	-
Vitreoscilla sp.	K/LS	Sm/-	-	H	-	-	H	-	-	-	L/-	L/-	L?/-	L?/-
Vitreoscilla sp. - ASV14093	K/LS	-	-	H	-	-	H	-	-	-	L	L	L?	L?
Vitreoscilla sp. - ASV523fe	K/LS	Sm	-	H	-	-	H	-	-	-	L	-	-	-
Vitreoscilla sp. - ASV5b667	K/LS	-	-	H	-	-	H	-	-	-	-	-	-	-
Vitreoscilla sp. - ASV6d6c1	K/LS	-	-	H	-	-	H	-	-	-	L	-	-	-
UNKN Deltaproteobacteria spp.	-	Sm	Su	-	-	-	H/-	-	L	L	-	-	H	H
UNKN Deltaproteobacteria - ASVa14c3	-	Sm	Su	-	-	-	H	-	L	L	-	-	H	H
UNKN Deltaproteobacteria - ASVa6c8b	-	Sm	Su	-	-	-	-	-	L	L	-	-	H	H
UNKN Desulfobulbaceae spp.	CAA/-	Lr/-	-	L/-	L/-	H/-	-	H/-	-	-	H/-	H/-	-	-
UNKN Desulfobulbaceae - ASV00023	-	-	-	-	L	-	-	-	-	-	-	-	-	-
UNKN Desulfobulbaceae - ASV2dce4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UNKN Desulfobulbaceae - ASV62490	CAA	Lr	-	L	-	-	-	H	-	-	-	H	-	-
UNKN Desulfobulbaceae - ASV81bee	CAA	Lr	-	L	L	-	-	-	-	-	-	-	-	-
UNKN Desulfobulbaceae - ASV942de	-	Lr	-	-	-	-	-	-	-	-	H	-	-	-
UNKN Desulfobulbaceae - ASVa8726	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UNKN Desulfobulbaceae - ASVb431e	CAA	Lr	-	L	-	H	-	-	-	-	-	H	-	-
UNKN Desulfobulbaceae - ASVd5064	CAA	-	-	L	-	-	-	-	-	-	-	-	-	-
UNKN Desulfobulbaceae - ASVd59b4	-	Lr	-	-	-	-	-	-	-	-	-	-	-	-
UNKN Desulfovibrionaceae - ASVef8b2	CAA	Sm	-	L	-	-	-	-	-	-	-	H	-	-
Desulfuromonas sp. - ASV1019a	-	-	-	-	-	-	-	-	-	-	H	-	-	-
UNKN Desulfuromonadales spp.	BB/CAA/-	Lr/-	-	L/-	L/-	H/-	-	-	-	-	-	H/-	H/-	H/-
UNKN Desulfuromonadales - ASV009ac	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UNKN Desulfuromonadales - ASV027a0	BB/CAA	Lr	-	L	-	-	H	-	-	-	-	-	H	H
UNKN Desulfuromonadales - ASV0c8df	-	-	-	-	L	-	-	-	-	-	-	-	L	L
UNKN Desulfuromonadales - ASVef783	CAA	-	-	L	L	-	H	-	-	-	-	-	-	-
Arcobacter sp. - ASV55e6e	CAA	-	De	L	-	-	L	H	H	H	-	H	-	-
Sulfuricurvum sp. - ASVc5adf	CAA	-	Su	L	L	-	H	-	L	L	L	-	H	H
UNKN Epsilonproteobacteria - ASV8ed64	-	Sm	-	-	-	-	L	L	H	-	H	-	-	-
UNKN Helicobacteraceae - ASVccdd5	-	-	De	-	-	-	L	-	H	H	-	H	-	-
Stutzerimonas stutzeri - ASVb0b18	BB	-	De	-	H	-	L	-	H	H	-	-	L	L
Vibrio spp.	CAA/-	Sm/-	De/-	L/-	H/-	H/-	L/-	H/-	H/-	H/-	-	H/H?/-	H?/-	H?/-
Vibrio sp. - ASV18e60	CAA	Sm	De	L	-	-	-	-	H	H	-	-	-	-
Vibrio sp. - ASV3aa29	-	-	-	-	-	-	-	H	-	-	-	H	-	-
Vibrio sp. - ASV4650f	-	-	-	-	-	-	L	-	-	-	-	-	-	-
Vibrio sp. - ASV46e90	CAA	Sm	-	L	-	H	-	-	-	-	-	-	-	-
Vibrio sp. - ASV4abe3	CAA	-	-	L	-	-	-	H	-	-	-	-	-	-
Vibrio sp. - ASV708e2	CAA	-	-	L	-	-	-	H	-	-	-	-	H?	H?
Vibrio sp. - ASV71ce2	CAA	-	De	L	-	-	L	-	H	H	-	H	-	-
Vibrio sp. - ASVc2a07	-	-	De	-	-	-	-	H	-	H	-	H	-	-
Vibrio sp. - ASVca89c	CAA	-	De	L	H	-	-	-	H	-	-	H	-	-
Vibrio sp. - ASVe1a8a	CAA	-	-	L	-	-	-	H	-	-	-	H?	-	-
Vibrio sp. - ASVfb17c	CAA	-	De	L	-	-	L	-	H	H	-	H	-	-
UNKN Nitrospiraceae - ASV5c36e	CAA	-	Su	-	-	-	H	-	L	L	L	-	H	H
Unknown Bacteria spp.	CAA/-	Lr/Sm/-	Su/De/-	L/-	-	H/-	L/-	L/H/-	L/H/-	L/-	L/-	H?/-	H/H?/-	H/H?/-
UNKN Bacteria A - ASV20751	CAA	Lr	Su	L	-	-	H	-	L	-	-	-	-	-
UNKN Bacteria A - ASV36a96	-	-	-	-	-	-	-	L	-	-	-	-	-	-
UNKN Bacteria A - ASVe569b	CAA	-	De	L	-	-	-	-	H	H	-	-	H?	H?
UNKN Bacteria B - ASVa5810	CAA	Sm	De	L	-	-	-	-	H	-	-	H?	-	-
UNKN Bacteria B - ASVbbdf	-	-	De	-	-	-	-	-	H	-	-	-	-	-
UNKN Bacteria C - ASVbc22	CAA	-	Su	L	-	-	H	-	L	L	L	-	H	H
UNKN Proteobacteria - ASVb20cf	-	-	Su	-	-	-	H	-	L	L	-	-	H	H

4.9.3 Supplemental Figures for Chapter 4

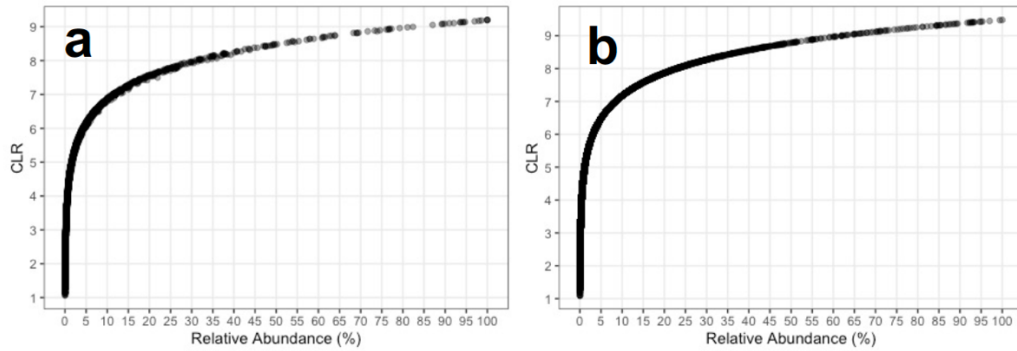


Figure 4.8. Relationship between CLR values and relative abundances calculated for each ASV within each sample. **(a)** Relationship for values in Arctic GN02 *nifH* dataset. **(b)** Relationship for values in Bedford Basin *nifH* dataset [limited to ASVs in more than two samples].

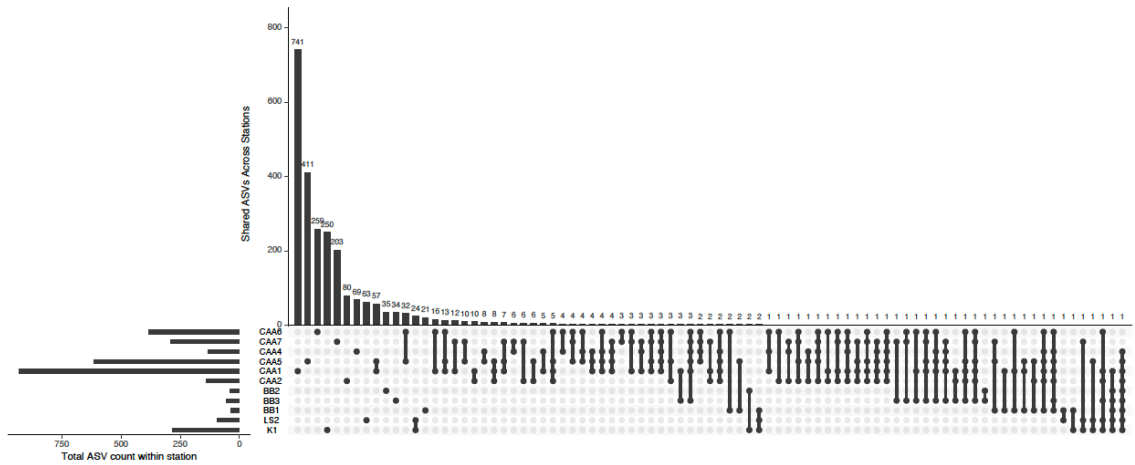


Figure 4.9. Upset plot showing the number ASVs unique to each sample and shared across sample in the Arctic *nifH* GN02 dataset. Top bar plot is count of shared ASVs for each grouping of samples (shown by bottom dots and lines). The total number of ASVs within each station is also shown (bar plot on left).

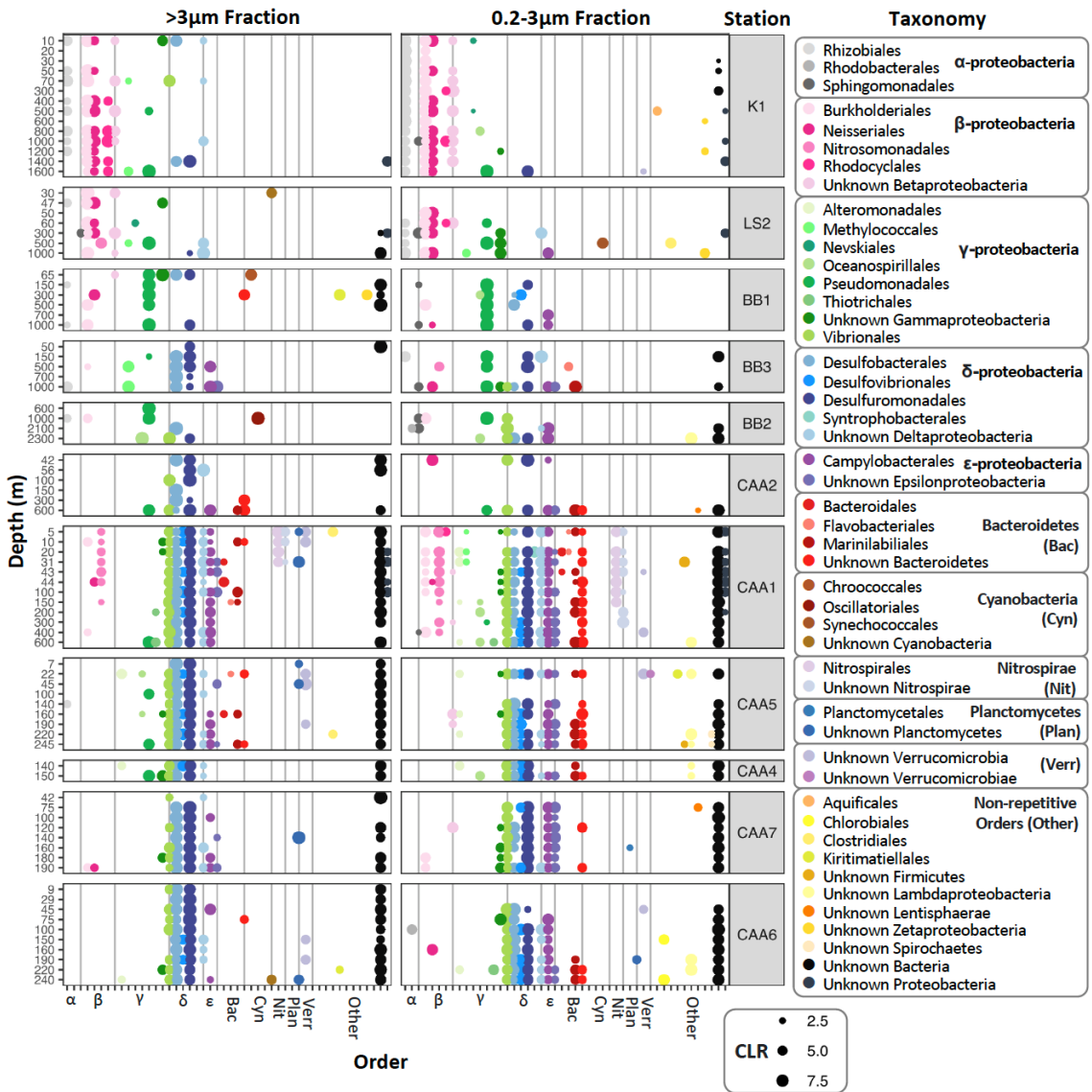


Figure 4.10. Major diazotroph Orders present in Canadian Arctic Gateway during GN02 GEOTRACES Expedition. Values are centered-log ratios (CLR) for individual ASVs. To increase readability data are over-plotted for individual ASVs, therefore dots shown only convey the maximum CLR value reached for a taxonomic group at a given station and depth (i.e., it is possible that lower CLR values may be masked by larger CLR values at any given station + depth within each taxonomic group). Hence, these data should only be used to interpret areas where maximal CLR values were observed per taxonomic Order. For each ASV if the taxonomic Order is unresolved, the next highest taxonomic rank was used (these are groups labeled “Unknown ...”). The x-axis from left to right follows the legend list from top to bottom and ASVs have been grouped by their taxonomic classes.

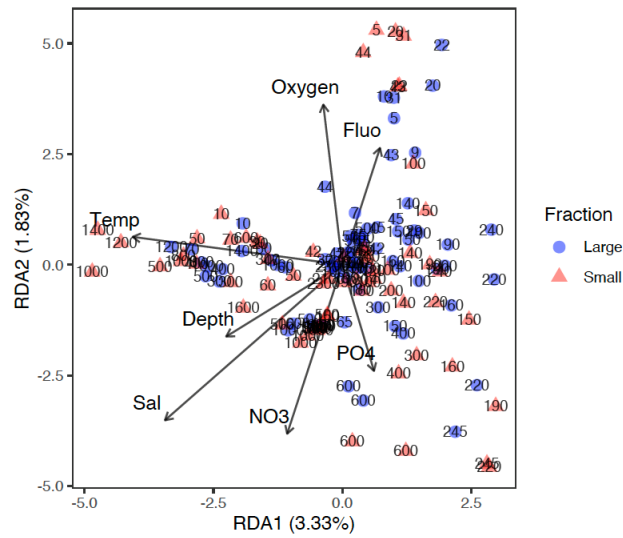


Figure 4.11. RDA analysis presented in Fig. 4.3, but instead samples are visualized by size fraction. For size fractions: Large = $\geq 3\mu\text{m}$, and Small = $0.2\text{--}3\mu\text{m}$. See caption of Fig. 4.3 for abbreviations.

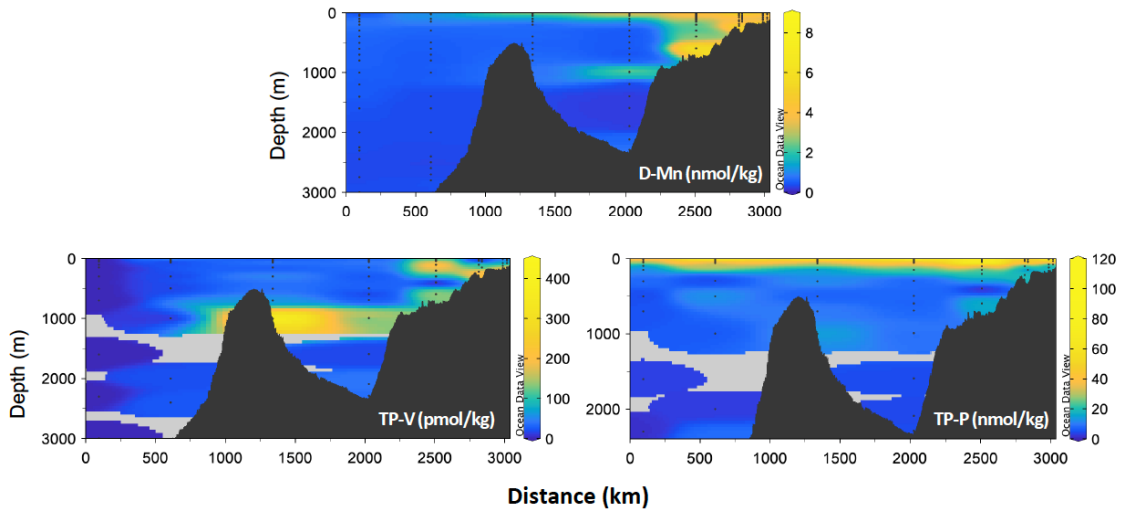


Figure 4.12. Additional oceanographic data from the Canadian Arctic Gateway GN02 expedition. Data are from IDP2021 dataset (GEOTRACES Intermediate Data Product Group, 2021) and Colombo et al. (2020, 2021, 2022). Data shown include dissolved manganese (D-Mn), total particulate vanadium (TP-V), and total particulate phosphorus (TP-P). Dark ribbons at bottom of graphs outline bottom depths based on *ETOPO1* dataset (Amante & Eakins, 2009). Stations are organized according to their distance from the outer Labrador Sea.

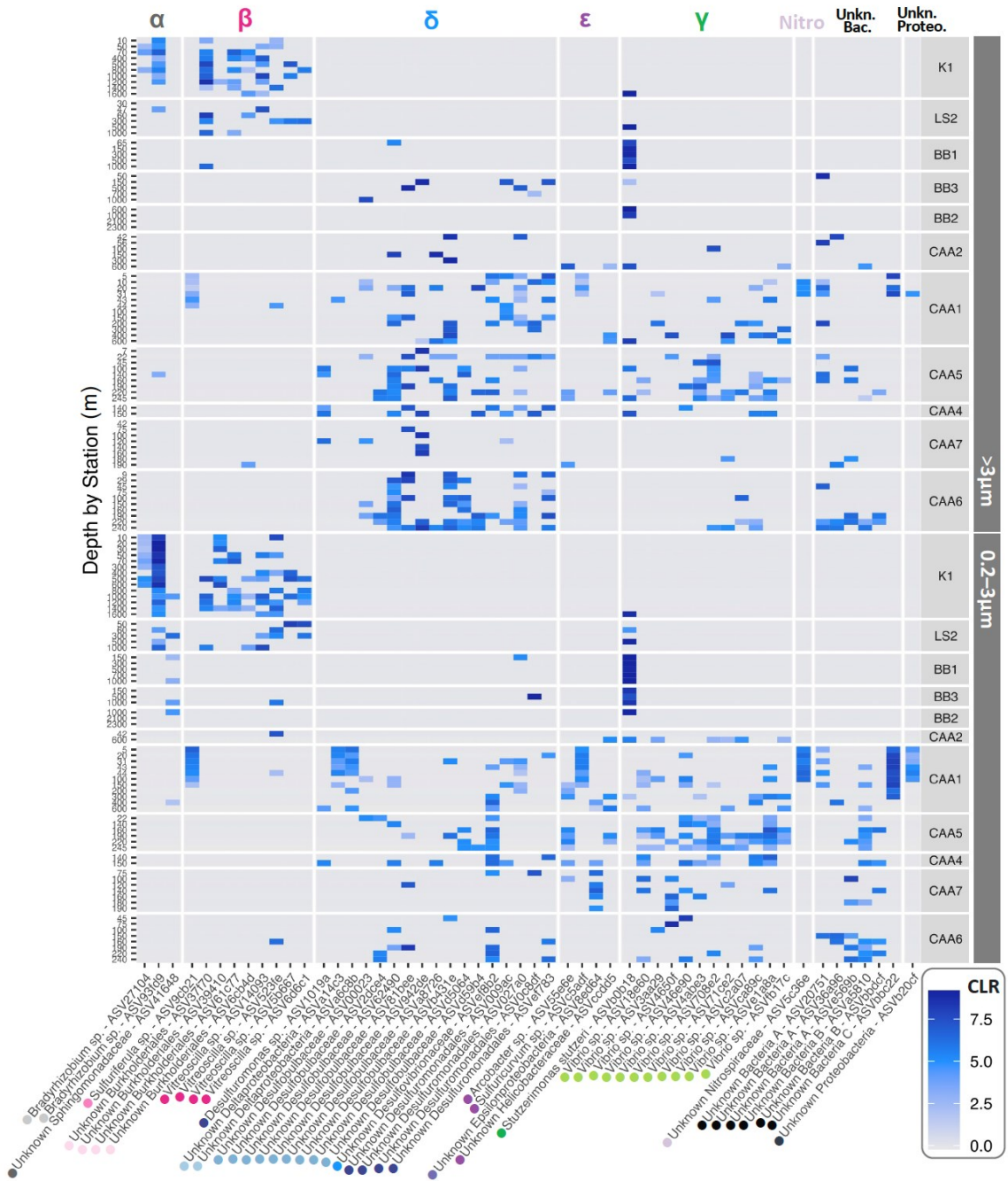


Figure 4.13. Individual spatial profiles (based on CLR values) for dominant *nifH* ASVs detected in the Canadian Arctic Gateway during GN02 expedition and present in ≥ 7 samples. Nitro = Nitrospirae, Unkn. Bac. = Unknown Bacteria, and Unkn. Proteo. = Unknown Proteobacteria. Green symbols indicate proteobacterial groups. See Fig. 4.5 taxonomic Class coloration for interpreting colored dots on ASV names.

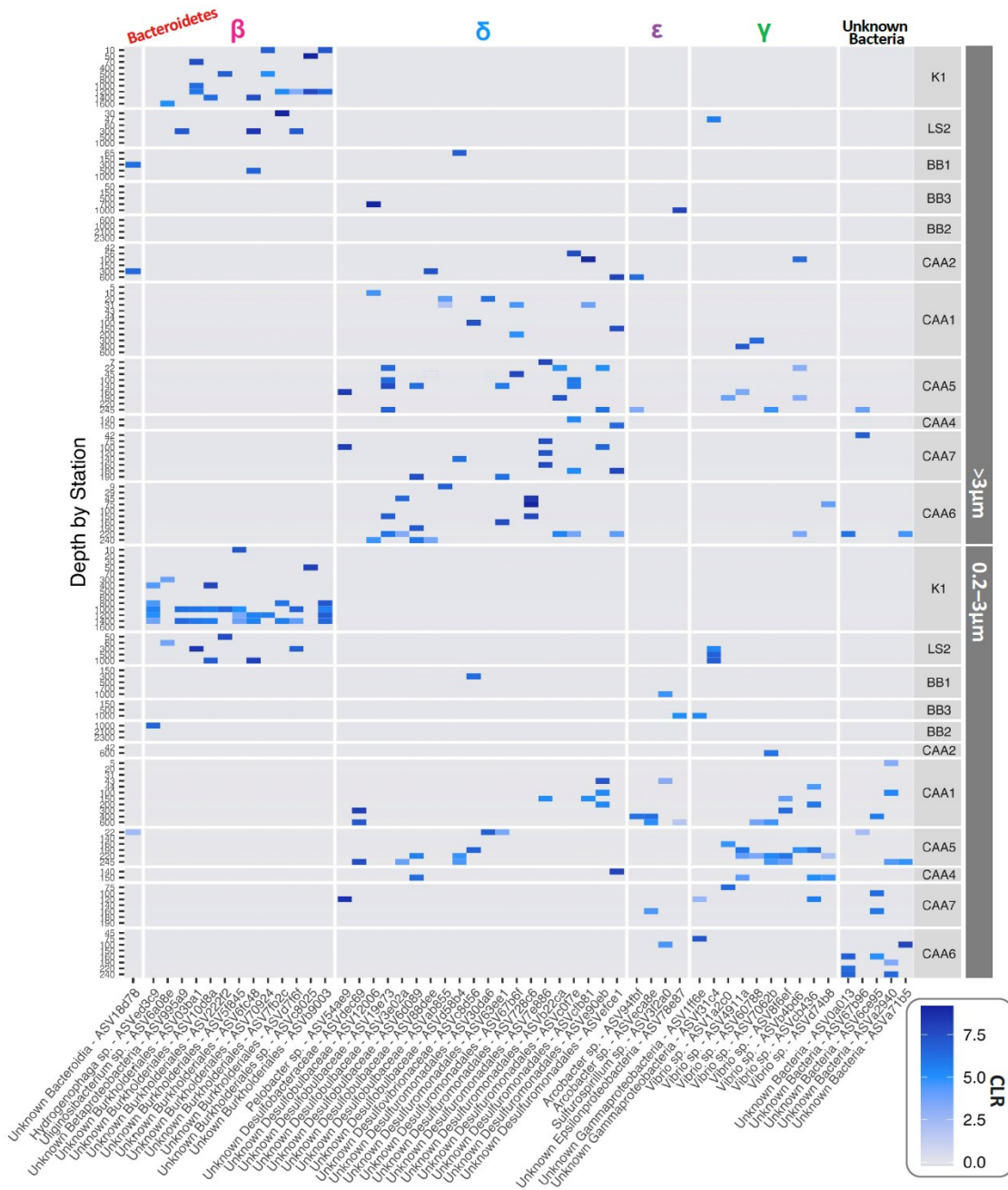


Figure 4.14. Individual spatial profiles (based on CLR values) for dominant *nifH* ASVs detected in the Canadian Arctic Gateway during GN02 expedition and present in <7 samples. Greek symbols indicate proteobacterial groups. The x-axis is grouped into taxonomic classes.

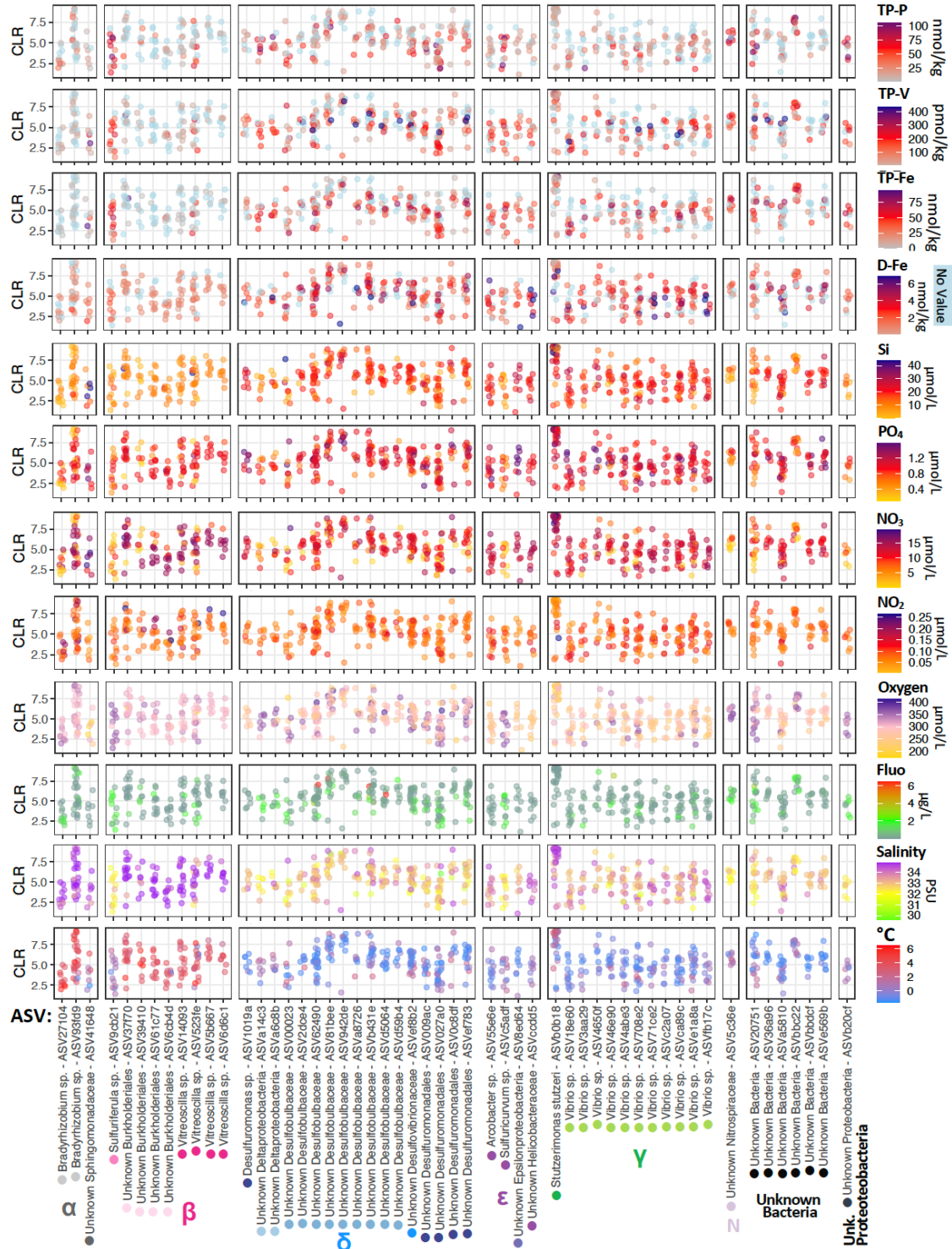


Figure 4.15. Raw CLR data distributions for dominant ASVs in the CAG compared to each other according to oceanographic data. For ASVs in ≥ 7 samples. Environmental data are from *IDP2021* (GEOTRACES Intermediate Data Product Group, 2021) and Colombo et al. (2020, 2021, 2022). Greek letters denote proteobacterial groups. Abbreviations: Nitrospirae (N), dissolved (D-), Silicate (Si), dissolved O₂ (Oxygen), Fluorescence (Fluo), total particulate (TP-). Colored dots correspond to Class (See Fig. 4.5).

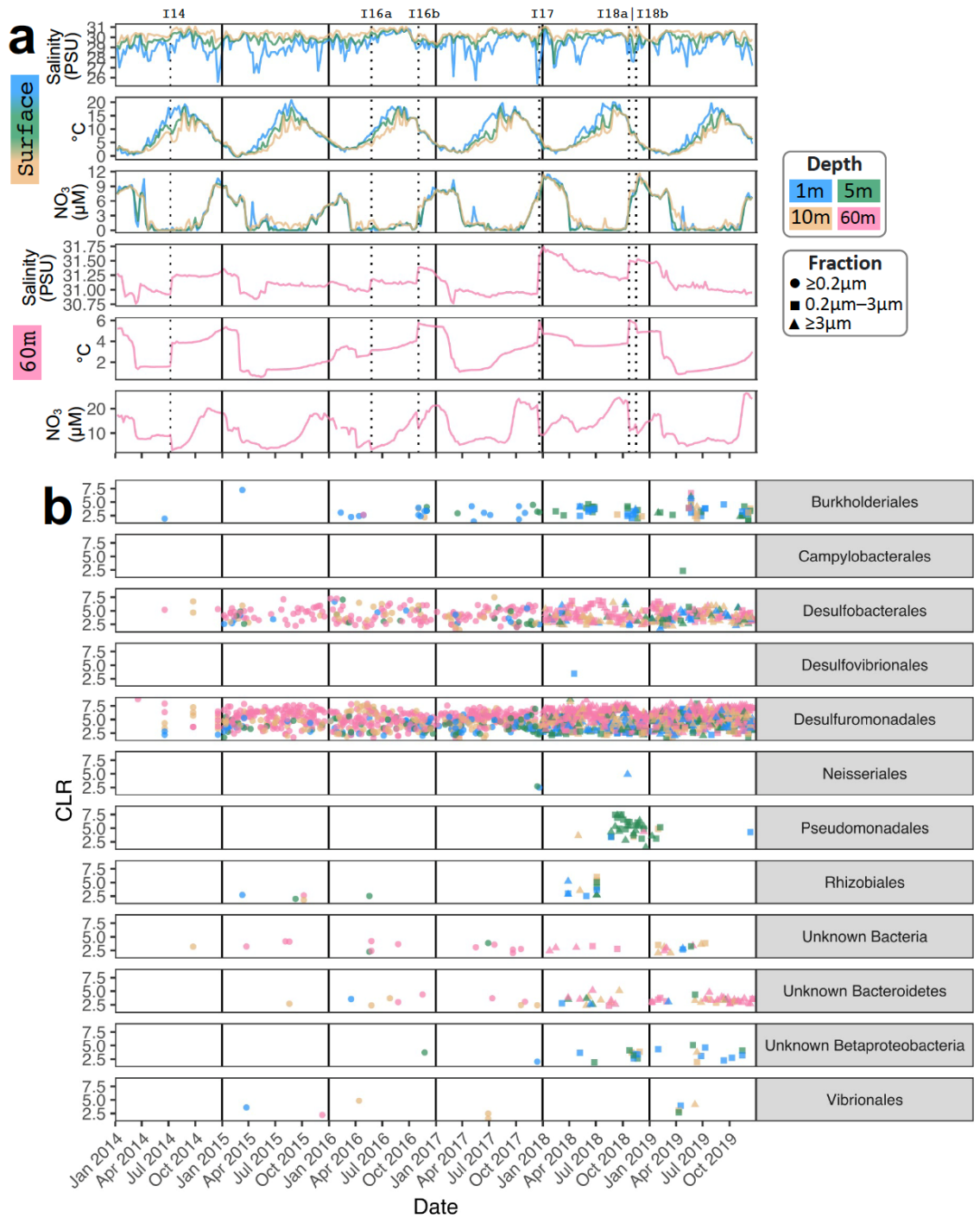


Figure 4.17. All dominant ASVs recovered from the Canadian Arctic Gateway within a 5-yr time series in the coastal NW Atlantic organized by taxonomic Order. **(a)** surface and deep (60m) ocean conditions in the Bedford Basin time series replotted from main text to allow for easier comparison. Dashed lines = known intrusion events. **(b)** CLR values for all dominant ASVs falling within the taxonomic Orders recovered. Panel B is an expanded version of Fig. 4.7c.

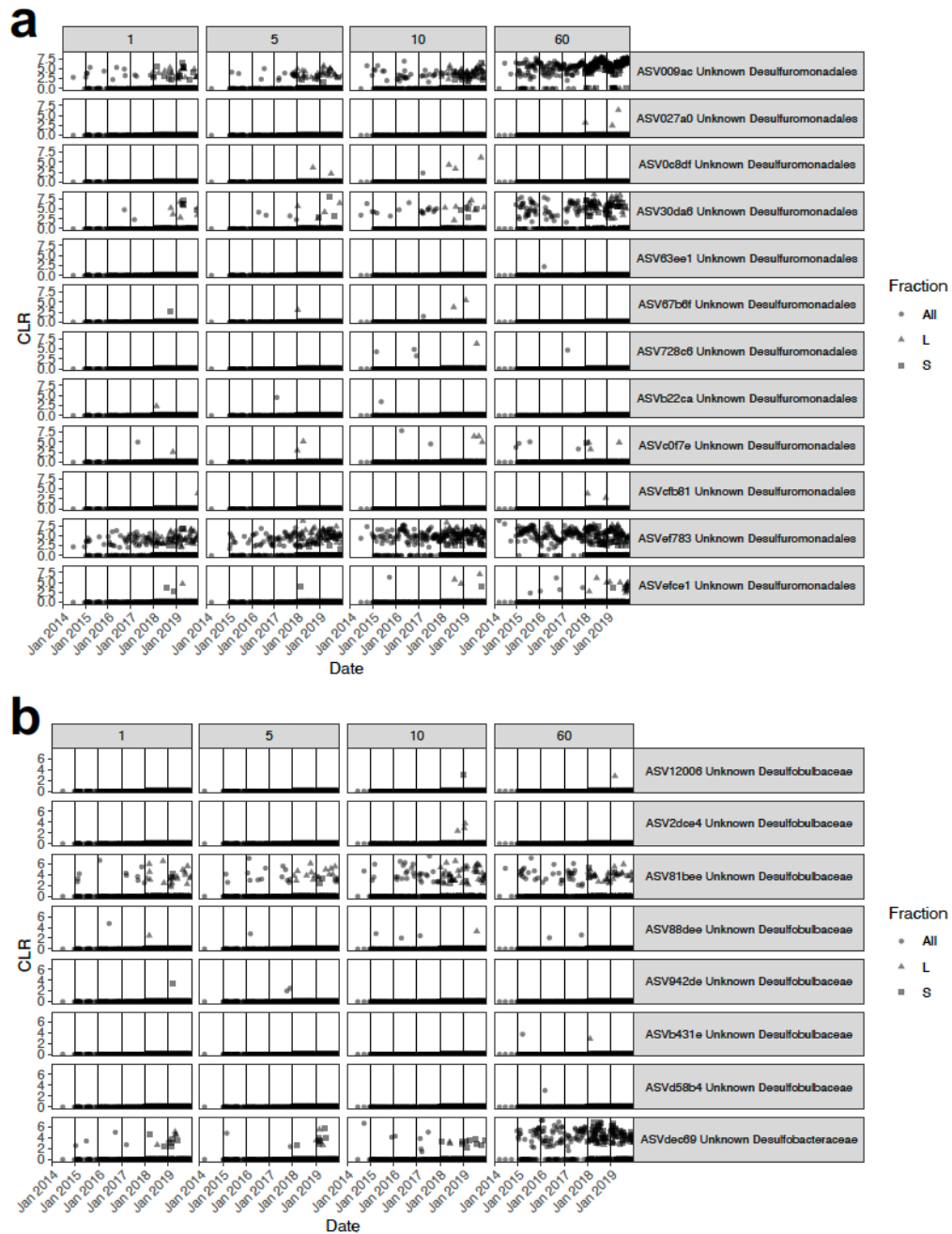


Figure 4.18. Individual temporal profiles (for CLR values) for all dominant Unknown Desulfuromonadales **(a)** and Unknown Desulfobulbaceae **(b)** ASVs from the Canadian Arctic Gateway recovered within the Bedford Basin times series. Size fractions (Fraction) are: All = $\geq 3\mu\text{m}$, L = $\geq 3\mu\text{m}$, and S = 0.2– $3\mu\text{m}$. Note size fractionation only began at start of January 2018. Note that these plots provide an expanded view of Fig. 4.7d.

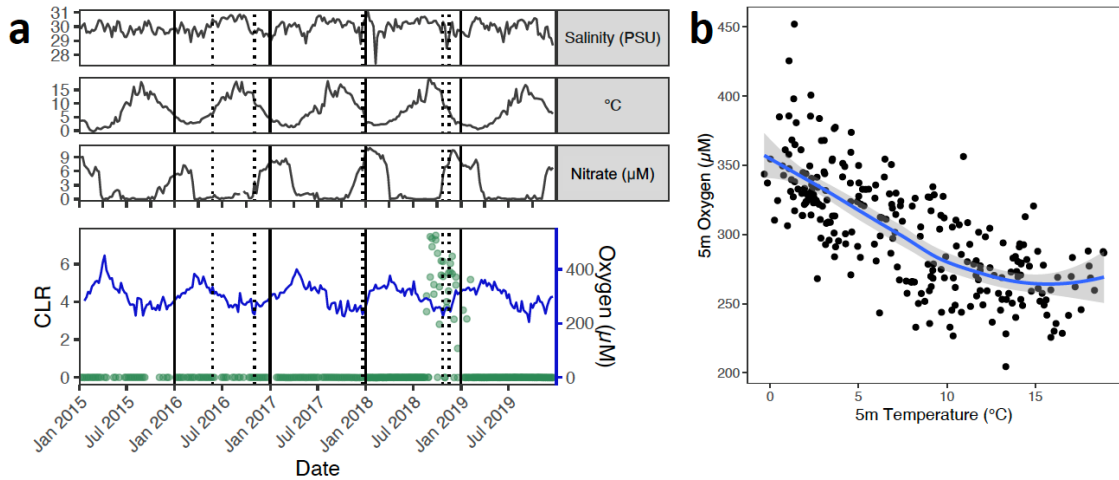


Figure 4.19. Dominant *Stutzerimonas stutzeri* in Bedford Basin and correlation with oxygen. **(a)** Individual temporal profile (based on CLR values) for ASVb0b18 *Stutzerimonas stutzeri* within the Bedford Basin time series aligned to 5m ocean data for the Bedford Basin. Dashed lines are shelf-water intrusion events at 60m. **(b)** The relationship between dissolved oxygen and temperature values within the Bedford Basin.

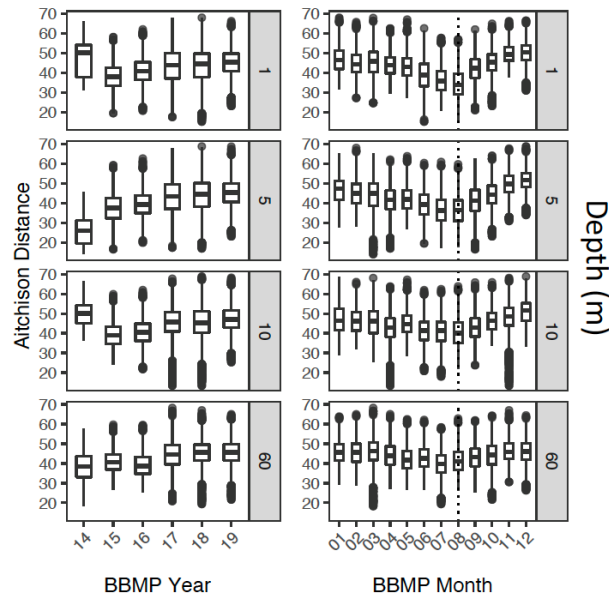


Figure 4.20. Distributions for Aitchison Distances calculated between all Bedford Basin *nifH* samples and all Arctic *nifH* samples (both size fractions) and divided out by year (left panels) and by month (right panels). The GN02 cruise occurred in July–August 2015 (dashed line). Note that 2014 data are shown in the left panels, however, there are only 10 samples for this year during the Bedford Basin time series, therefore the boxplot for 2014 does not represent a complete comparison and should not be included in the interpretation [it remains in the figure so that there is a record of the calculations].

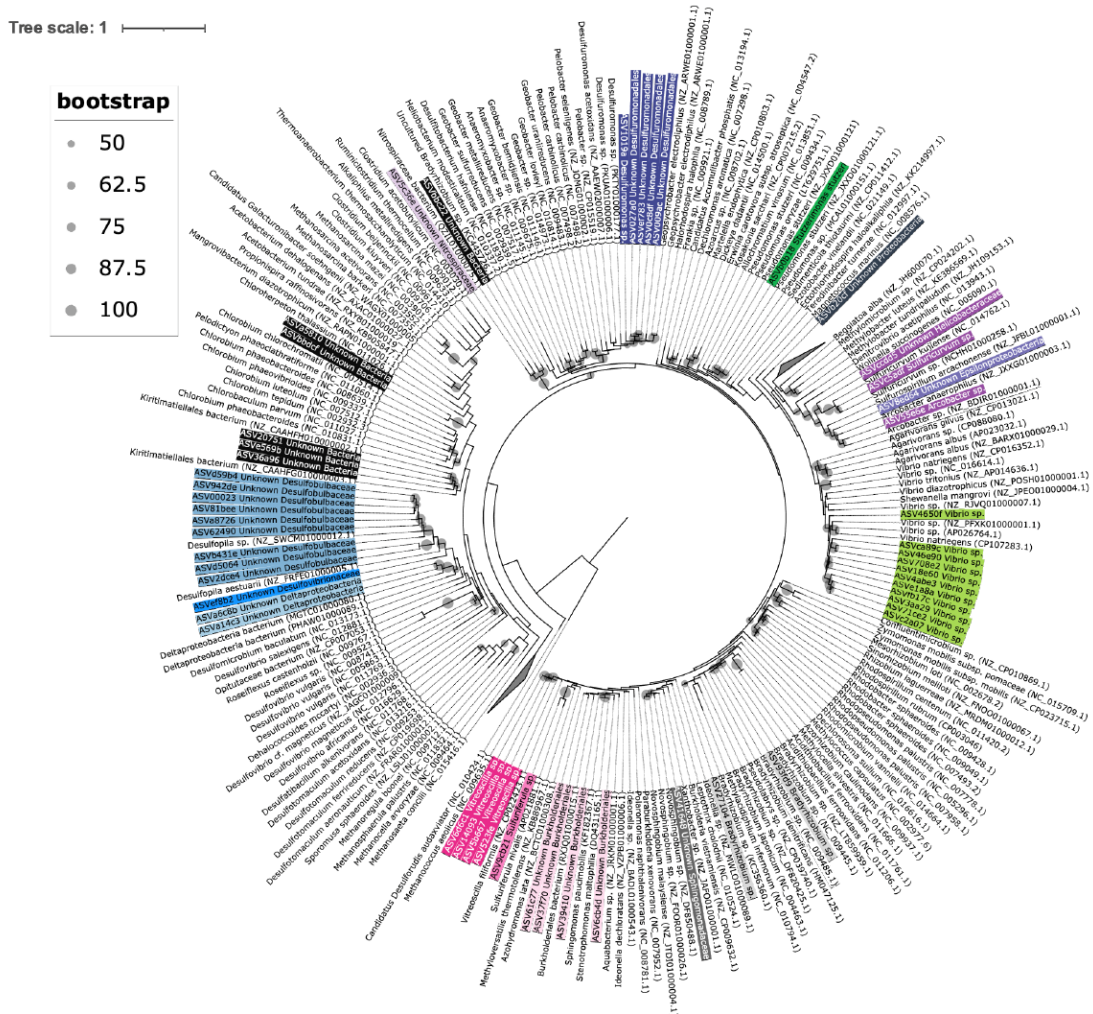


Figure 4.21. Branch lengths for Maximum Likelihood phylogeny presented in the main text. Colored branch labels are dominant *nifH* ASVs that occurred in the Canadian Arctic Gateway during the GN02 expedition. This is the same tree as that in Fig. 4.5, however, here the topology of the tree is shown with branch lengths included in the analysis. Please see the caption of Fig. 4.5 for more complete tree reconstruction details. Bootstrap values (bootstrap) are for 1000 replicates.

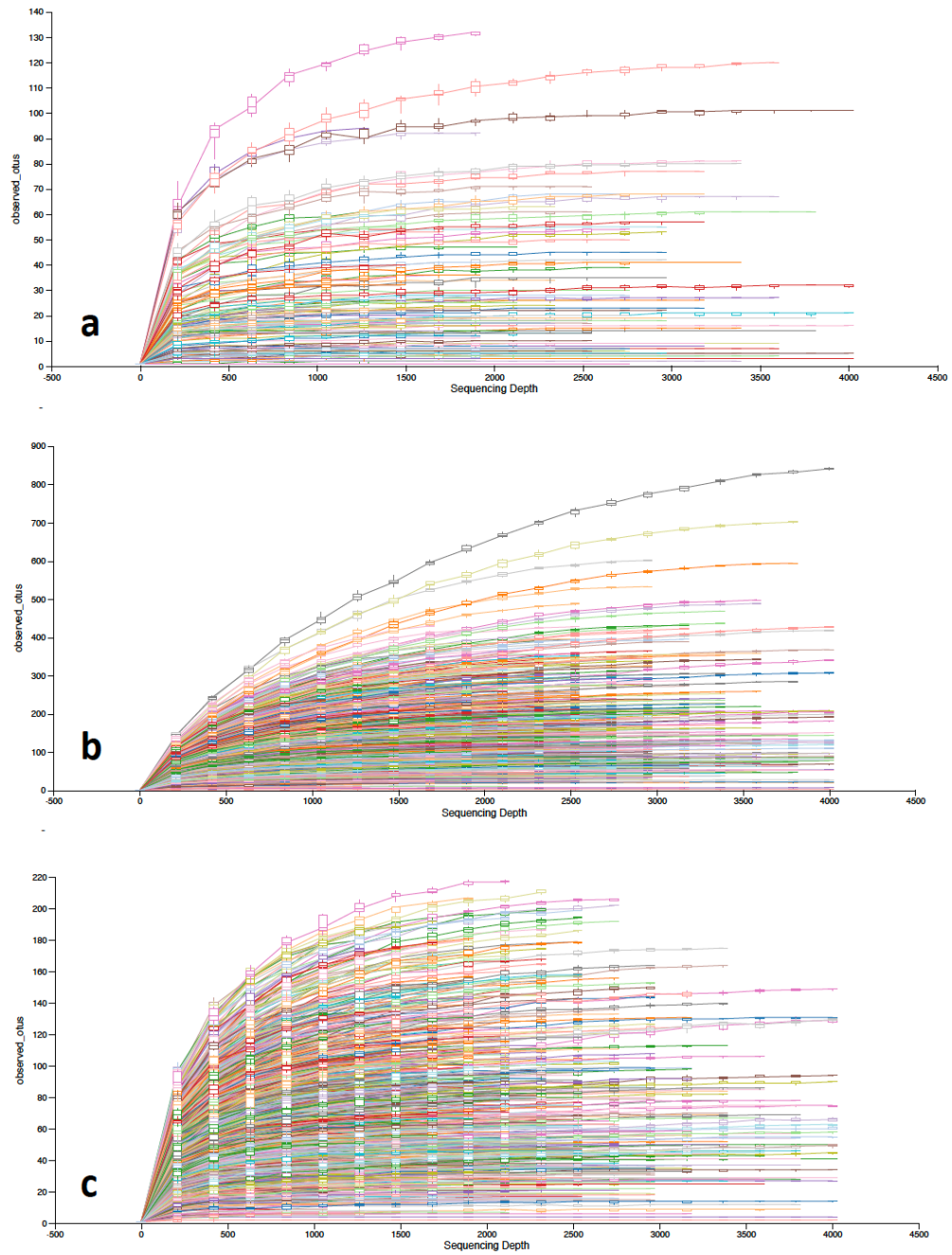


Figure 4.22. Rarefaction curves for raw sequencing data. Individual samples are colored, and plots were generated as part of *QIIME2 View* (Bolyen et al., 2019). **(a)** *nifH* samples from Arctic GN02 cruise. Also shown are *nifH* samples from Bedford Basin time series filtered at ASVs occurring in at least **(b)** two samples or **(c)** 25 samples.

CHAPTER 5

Conclusions

5.1 Overview of Thesis Conclusions

At its core this thesis has significantly advanced our understanding of the diversity and community dynamics of phytoplankton and diazotrophs within the NWA and Arctic Ocean sectors, thereby achieving its overarching objective. Although Chapters 2, 3, and 4 covered rather broad and distinct research contributions (phytoplankton, UCYN-A, and Arctic diazotrophs) they are all connected through the reoccurring approach of using higher-resolution spatiotemporal datasets, along with detailed forays into the intraspecific layers of microbial diversity, to help us better understand life in the ocean microbiome. Chapter 2 provided a prime example of using chloroplast *16S* rRNA genetic markers to resolve phytoplankton within an ocean time series. Specifically this Chapter advanced our understanding of phytoplankton by: (i) revealing the chloroplast genetic signatures of species that have strong seasonal cycles within the coastal NWA (e.g., *Eutreptiella pomquetensis* in the spring), (ii) establishing that there is overlap between the phytoplankton communities of the Bedford Basin time series and the nearby Scotian Shelf, and (iii) demonstrating that the known positive relationship between smaller phytoplankton cell abundances and temperature for the NWA can vary in the Bedford Basin based on anomalous temperatures (Li et al., 2006). Overall, Chapter 2 contributes a rich multi-year weekly molecular dataset that can be used to ask further questions about phytoplankton in the coastal NWA. This utility was demonstrated in Chapter 3, where using a network analysis, I was able to identify the likely presence of the UCYN-A/host consortium by identifying co-occurring haptophyte *16S* rRNA chloroplast and UCYN-A *nifH* ASVs (Thompson et al., 2012); specifically, in the Bedford Basin, *Braarudosphaera bigelowii* & UCYN-A2 co-occurred and *Chrysochromulina* & UCYN-A1 co-occurred. Keeping with the concept of exploring intraspecific levels of diversity, in Chapter 3, I examined mutations associated with individual UCYN-A ASVs over the period of three years, hence

identifying microevolutionary patterns. The findings presented in Chapter 3 allowed me to conclude that much of the rare and transient diversity observed within UCYN-A ecotypes may result from neutral mutational variation (Yang & Bielawski, 2000). While completing the research for this dissertation it became apparent that the qPCR assay typically used to enumerate *nifH* gene copies for UCYN-A2 (that of Thompson et al., (2014)) could be improved to reduce its cross-reactivity to the UCYN-A1 ecotype. This improved assay was successfully designed in Chapter 3 and then applied to weekly eDNA samples from the Bedford Basin. Results from this chapter established that within a coastal ecosystem the occurrence of UCYN-A1 and -A2 ecotypes could be highly variable from year-to-year (sometimes co-occurring and sometimes discretely occurring over a matter of weeks). Overall, findings from Chapter 3 advance our ability to interpret intraspecific diversity within the globally important UCYN-A (Farnelid et al., 2016; Martínez-Pérez et al., 2016) and emphasizes the need to consider both spatial and temporal patterns as we strive to continue refining the ecological definitions of UCYN-A ecotypes. Chapter 4, like Chapters 2–3, expands our understanding of the ocean microbiome by sampling the environment at higher resolutions, but in this case spatially by focusing on diazotrophs in the under-sampled Canadian Arctic Gateway (von Friesen & Riemann, 2020). In Chapter 4, I characterized the diversity and environmental conditions associated with ASVs belonging to dominant diazotrophs observed within the Labrador Sea, Baffin Bay, and the Canadian Arctic Archipelago (dominant ASVs all fell within diazotrophic Cluster I [Molybdenum and some Vanadium containing nitrogenases] and Cluster III [containing anaerobes])(Riemann et al., 2010; Zehr et al., 2003)). At the genus-level some of the dominant ASVs matched genera previously identified using *nifH* within the Arctic Ocean (e.g., *Bradyrhizobium* and *Arcobacter* (Blais et al., 2012; Fernández-Méndez et al., 2016; Karlusich et al., 2021)), but at the ASV-level, more than 50% of the dominant ASVs had rather low sequence similarity to important *nifH* signatures known from elsewhere in the ocean (Delmont et al., 2021; Turk-Kubo et al., 2022). Overall, these results indicate that the Canadian Arctic Gateway may provide a prime location to continue studying a large fraction of undescribed diazotrophs falling within the NCD guild. In Chapter 4, I also detected very low levels of UCYN-A throughout the Arctic sites examined; these results are important as they advance our understanding of UCYN-A distribution within the Arctic

Ocean by demonstrating that this species can be found east of the Chuckchi and Beaufort Seas (and hence outside of where UCYN-A has mainly been studied in the Arctic previously; Harding et al., 2018; Shiozaki et al., 2018). As a final component, Chapter 4 described the broader diazotroph community composition observed for the Canadian Arctic Gateway and revealed that during the 2015 GEOTRACES expedition the diazotroph community was biogeographically separated between the Labrador Sea and more northern Baffin Bay/Canadian Arctic Archipelago sites. Although more work will be needed to determine if this biogeographic separation reoccurs annually, Chapter 4 provides an initial account of this discovered pattern.

In the end, this thesis presents a more detailed picture of: (i) phytoplankton coastal dynamics within the NWA, (ii) UCYN-A intraspecific community patterns, and (iii) the diversity of marine diazotrophs that can occur in the Arctic Ocean. In the section that follows, I will briefly provide some examples of how the set of high-resolution observations offered by this thesis can help inform future studies.

5.2 Opportunities for Future Research

The total microbial diversity on Earth is immense, yet much of it remains undescribed and uncultured (De Vargas et al., 2015; Locey & Lennon, 2016; Pedrós-Alió, 2006; Pedrós-Alió & Manrubia, 2016). This feature is reflected in Chapter 2 and Chapter 4 by the numerous dominant plankton and diazotrophs that were identified for the coastal NWA and Arctic Ocean, but for which no species names could be attributed using current methods for assigning taxonomies. Consequently, a major research opportunity that remains is the culturing of regionally dominant phytoplankton and diazotrophs that this dissertation has brought to the forefront. Additional culturing efforts also have the potential to further our understand of the metabolisms and physiologies of these under-described organisms (Lewis et al., 2020). In fact, one could argue that the multi-year time series presented herein afford us an opportunity to not only know more about which uncultured organisms may be worth pursuing for laboratory isolation, but also when and where we can expect to find them in the coastal NWA and CAG.

Reflecting on all chapters presented in this dissertation, three microbes stand out as particularly interesting for future culturing pursuits. The first is the source of the

Eutreptiella pomquetensis chloroplast signature that occurred repeatedly in the fall [cASV18] at the edge of the Scotian Shelf. At present there is only one cultivated strain of *E. pomquetensis* and it is known to grow only at temperatures that are far colder than that observed herein at the Scotian Shelf edge (McLachlan et al., 1994). Given the importance of *E. pomquetensis* in the spring within the Bedford Basin (Chapter 2), the fall waters of the Scotian Shelf represent a prime opportunity to study the ecology of this phytoplankton (or at minimum its chloroplasts) in warmer waters.

The second microbe of interest is UCYN-A and its host. If Chapter 3 teaches us anything about how we can advance efforts for isolating UCYN-A ecotypes from the coastal NWA, it is that the dominance of individual ecotypes for this species can alternate across the summer to early fall period within this region. For instance, if one were to use summer and fall Bedford Basin seawater to further attempt the cultivation of UCYN-A, sampling at weekly intervals could be advantageous for targeting separate UCYN-A ecotypes. In this regard, weekly enrichment culturing efforts that focus in on time-periods when one can expect to find UCYN-A in the Bedford Basin (Chapter 3) seem the most promising for moving forward with culturing efforts in the north Atlantic. For example, I have recently had mixed success with simply using weekly summer-fall seawater from the Bedford Basin and adding Fe, PO₄, and vitamins to enrich samples for diazotrophs, and hence ecotypes of UCYN-A (Fig. 5.1a; see Langlois et al. (2012) for an example of enriching for diazotrophs using iron and phosphate). Although this strategy (full methods described in caption of Fig. 5.1) seems promising for reducing the taxonomic complexity within a given laboratory sample, it inevitably provides its own unique set of challenges given that other microbes will likely co-isolate with UCYN-A (note that the number of ASVs goes from typically >200 *16S* rRNA ASVs in the natural Bedford Basin community to <100 ASVs in the enrichment cultures). Possible reasons for associations in enrichment cultures would be varied (just like in natural communities) and could include for example cross-feeding or overlapping growth preferences (Fuhrman et al., 2015). This said, the *in situ* *16S* rRNA time series generated in Chapter 2 may prove useful as a starting point for helping to decipher whether co-isolating taxa also co-occur in nature (such co-presences are initially suggested in Fig. 5.1b). Overall, the example provided in Fig. 5.1 demonstrates how one can use data and trends from Chapters 2 and 3 as an opportunity to pursue the

isolation of UCYN-A ecotypes from the NWA in future. It is also important to note that for the Bedford Basin time series frozen Glycerol-TE samples are collected each week to preserve living cells. Equipped with the spatiotemporal patterns presented in Chapters 2 and 3, this catalogue of frozen samples represents yet another resource for working backwards into the past to try and isolate microbes (or even ecotypes) of importance.

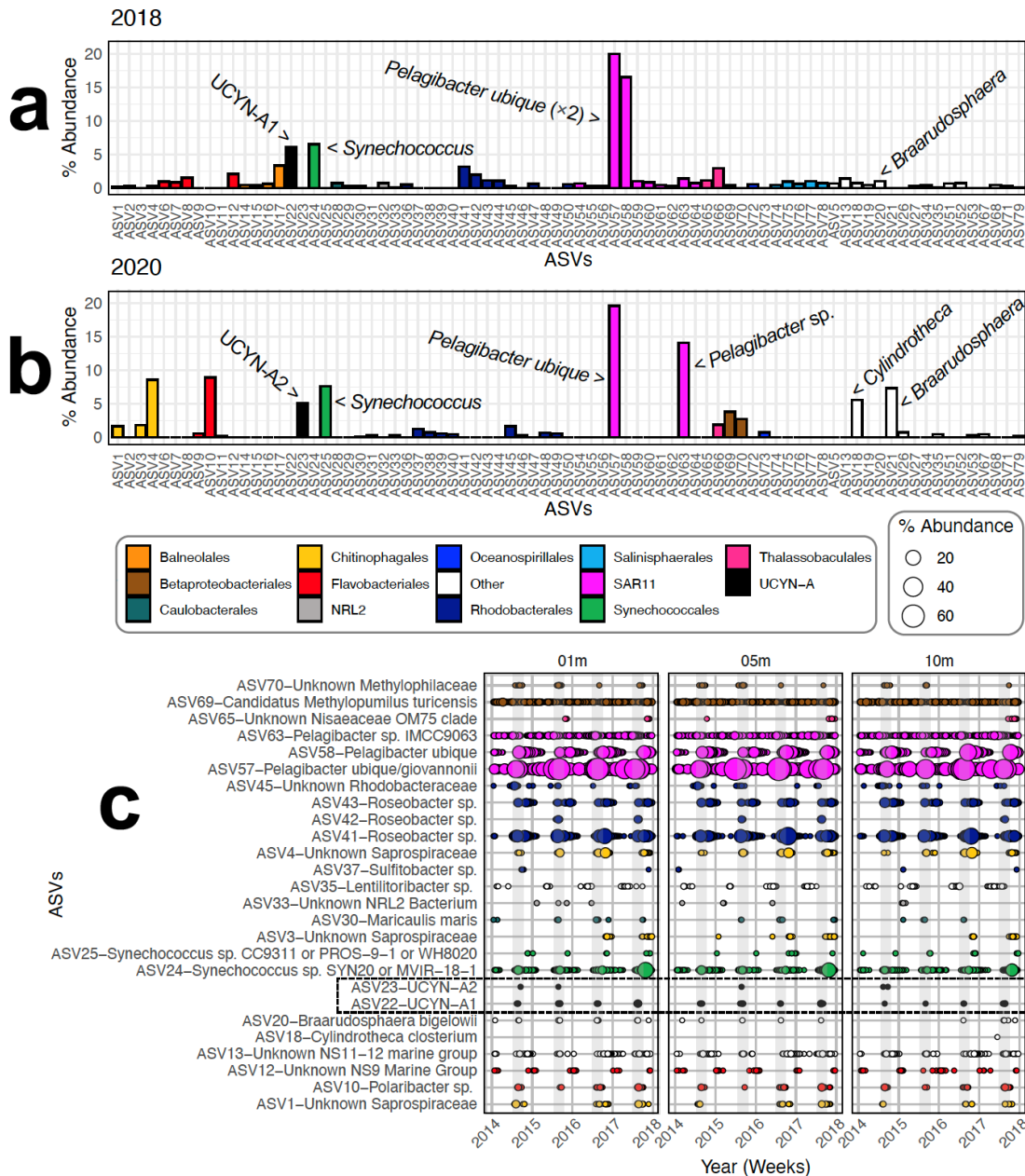


Figure 5.1. Example of using thesis results and *16S* rRNA datasets to inform additional culturing experiments. Bedford Basin 2018 and 2020 surface water was incubated with 2nM Fe, 200nM PO₄ (2018) or 400 nM PO₄ (2020), and 0.5 mL/L vitamins (2020 only) during the late summer–early fall to target UCYN-A (time-period informed from Chapter

3) [flasks obtained from Greiner Bio-One, Austria]. For 2018, after 13 weeks at 15°C with a 12h/12h light/dark cycle, the *16S* rRNA V6-V8 microbial community was characterized based on the relative abundances of ASVs (**panel a**). For 2020, after 2 weeks at the same temperature and light regime as above, we sorted 2,000 cells attributed to UCYN-A via fluorescence activated cell-sorting (BD Influx Cell Sorter; USA) [achieved by screening cell populations with qPCR assays from Chapter 3]. Sorted cells were grown for another 8.5 weeks in ~10mL of 0.2µm filtered Bedford Basin seawater used as the growth medium (water originated from the same day as the original sample). Also, the 2020 sorted cells were further enriched with 2nM Fe, 200nM PO₄, and 0.5mL/L vitamins after 0 weeks, 2 weeks, and 4 weeks of incubation. At the end of the 8.5 weeks, the *16S* rRNA V6-V8 microbial community was characterized (**panel b**). Multi-year *in situ* patterns for some of the cultured ASVs can be seen in the weekly Bedford Basin *16S* rRNA V6-V8 data obtained from Chapter 2 (**panel c**) [includes all *16S* rRNA data without chloroplasts singled out]. ASVs missing from the time series were likely rare and uncaptured by the eDNA sequencing. *16S* rRNA data were processed using the same methods given in Chapter 2, except the *SILVA 132.99* database (Quast et al., 2013) was also used for assigning taxonomies.

A final microbe of interest is the dominant Unknown Proteobacterial lineage [ASVb20cf] that was observed in the CAA region within Chapter 4. This *nifH* sequence was most like the Gamma 1 proteobacterial NCDs at ~94% amino acid similarity (Turk-Kubo et al., 2022). The Gamma 1 NCD belongs to a phylogenetic diazotroph clade that comprises Gamma 1, 2, & P and have previously been studied in the ocean using qPCR assays (Halm et al., 2012; Langlois et al., 2015; Langlois et al., 2008). Further comparison to the sequences and phylogenies provided in the gammaproteobacterial diazotroph study of Langlois et al. (2015) also indicates that the Unknown Proteobacterial ASV above [ASVb20cf] has the highest sequence similarity to an uncultured *nifH* clone that belongs to the Gamma 1, 2, & P clade [at 99% coverage and 94% DNA similarity to NCBI accession GQ426265.1 (Hamilton et al., 2011; Sayers et al., 2022)], hence further supporting that this ASV is most similar to the aforementioned clade. The Gamma 1, 2, & P group is closely related to the uncultured Gamma 4 and Gamma A NCDs that are widespread in the ocean (Langlois et al., 2015; Turk-Kubo et al., 2022). Consequently, the Canadian Arctic Archipelago may represent a good place for obtaining a polar isolate that is closely related to these well-known gammaproteobacterial diazotroph groups.

Besides culturing efforts, there are other avenues for future research. Although it was not pursued herein, additional research is needed to continue gathering information about nitrogen fixation in the CAG and Arctic Ocean more broadly (Sipler et al., 2017; von

Friesen & Riemann, 2020). For example, while the results of Chapter 4 demonstrate that microbes containing the *nifH* gene for dinitrogen reductase are present in the CAG, future transcriptomic and proteomics work would help to determine whether these *nifH* signatures are indeed active. Likewise, more nitrogen fixation rate measurements would help advance our interpretation of previous reports showing small fixation rates within the Baffin Bay (Blais et al., 2012). In addition to the above, this thesis also presents critical baseline information that can be used to study the ongoing effects of climate change on seasonal phytoplankton and diazotroph communities in the coastal NWA and diazotroph biogeography in the CAG (Benway et al., 2019; Hutchins & Fu, 2017). In this regard, it would be important to continue sampling for eDNA in the Bedford Basin time series and in the CAG with the goal of developing a better predictive picture of how the microbial communities are varying over time and in response to changing environmental conditions.

On a final note, the findings on UCYN-A microdiversity presented in Chapter 3 also calls into question whether or not the rare intraspecific microbial diversity (or ‘microdiversity’) that is made visible through the generation of ASVs is meaningful or not with respect to understanding the ecology of the organisms that they characterise. Unlike OTUs, which collapse diversity typically at ~97% similarity, ASVs present the microbial ecologist with sequences of very high DNA similarity to decipher (Nearing et al., 2018). In the case of UCYN-A in the Bedford Basin it was clear that the rare microdiversity for this species was ephemeral at our study site (present in one year and then gone the next year); however, the question remains as to how this pattern may change across other taxonomic groups and geographic regions (for e.g., in seasonally stratified waters versus more permanently stratified central ocean gyres). Using single cell genomics others have shown that another ubiquitous marine cyanobacterium, *Prochlorococcus*, can have 100s of co-existing subpopulations, suggesting that in some groups the displayed microdiversity is likely less transient (Kashtan et al., 2014). Interestingly, a study on ASV microdiversity recently completed for *Prochlorococcus* subpopulations in the North Pacific (Thompson et al., 2021), also showed that like the UCYN-A in our study only a few *Prochlorococcus* ASVs (in their case two ASVs) represented most of the sequencing reads across samples; yet consistent with single-cell genomics, 1000s of other rare *Prochlorococcus* ASVs were also present (based on the *ITS* rRNA gene). These findings like those presented in Chapter

3, indicate that it will be both interesting and important to study the microdiversity of other species over annual cycles to examine how ASV-level patterns play out across taxa. Furthermore, it would be worthwhile to acquire and examine single cell genomes for even more microbial groups to assess what other differences might exist in the overall genomes of different, but still closely related, ASVs. As others have suggested (for e.g., Ahlgren et al., (2019)), one of the key benefits of such fine scale microdiversity studies is that they may help us uncover the scale of diversity that is needed to resolve the ecologies of various microbial groups.

BIBLIOGRAPHY

- Acland, A., Agarwala, R., Barrett, T., et al. (2014). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 42(D1), D7–D17.
- Agawin, N. S. R., Benavides, M., Busquets, A., Ferriol, P., Stal, L. J., & Arístegui, J. (2014). Dominance of unicellular cyanobacteria in the diazotrophic community in the Atlantic Ocean. *Limnology and Oceanography*, 59(2), 623–637.
- Agawin, N. S. R., Duarte, C. M., Agustí, S., & Vaqué, D. (2004). Effect of N:P ratios on response of Mediterranean picophytoplankton to experimental nutrient inputs. *Aquatic Microbial Ecology*, 34(1), 57–67.
- Ahlgren, N. A., Perelman, J. N., Yeh, Y. C., & Fuhrman, J. A. (2019). Multi-year dynamics of fine-scale marine cyanobacterial populations are more strongly explained by phage interactions than abiotic, bottom-up factors. *Environmental Microbiology*, 21(8), 2948–2963.
- Ahlgren, N. A., & Rocap, G. (2012). Diversity and distribution of marine *Synechococcus*: Multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Frontiers in Microbiology*, 3(213), 1–24.
- Algora, C., Vasileiadis, S., Wasmund, K., Trevisan, M., Krüger, M., Puglisi, E., & Adrian, L. (2015). Manganese and iron as structuring parameters of microbial communities in Arctic marine sediments from the Baffin Bay. *FEMS Microbiology Ecology*, 91(6), 1–13.
- Altenburger, A., Blossom, H. E., Garcia-Cuetos, L., Jakobsen, H. H., Carstensen, J., Lundholm, N., Hansen, P. J., Moestrup, & Haraguchi, L. (2020). Dimorphism in cryptophytes—The case of *Teleaulax amphioxeia*/*Plagioselmis prolunga* and its ecological implications. *Science Advances*, 6(37), 1–8.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.
- Amante, C., & Eakins, B. W. (2009). ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. In *NOAA Technical Memorandum NESDIS NGDC-24*.

- Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*, 2(2), 1–7.
- Anderson, R. F., Mawji, E., Cutter, G. A., Measures, C. I., & Jeandel, C. (2014). GEOTRACES: Changing the way we explore ocean chemistry. *Oceanography*, 27(1), 50–61.
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. www.bioinformatics.babraham.ac.uk/projects/fastqc/
- Ardyna, M., Babin, M., Gosselin, M., Devred, E., Rainville, L., & Tremblay, J. É. (2014). Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophysical Research Letters*, 41(17), 6207–6212.
- Azam, F., Fenchel, T., Field, J., 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.
- Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37–48.
- Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., & Stamatakis, A. (2019). EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Systematic Biology*, 68(2), 365–369.
- Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., & Follows, M. J. (2010). Patterns of diversity in marine phytoplankton. *Science*, 327(5972), 1509–1511.
- Belkin, I. M., Cornillon, P. C., & Sherman, K. (2009). Fronts in large marine ecosystems. *Progress in Oceanography*, 81(1–4), 223–236.
- Benavides, M., Bonnet, S., Berman-Frank, I., & Riemann, L. (2018). Deep into oceanic N₂ fixation. *Frontiers in Marine Science*, 5(108), 1–4.
- Bendif, E. M., Nevado, B., Wong, E. L. Y., Hagino, K., Probert, I., Young, J. R., Rickaby, R. E. M., & Filatov, D. A. (2019). Repeated species radiations in the recent evolution of the key marine phytoplankton lineage *Gephyrocapsa*. *Nature Communications*, 10(1), 1–9.

- Bentzon-Tilia, M., Farnelid, H., Jürgens, K., & Riemann, L. (2014). Cultivation and isolation of N₂-fixing bacteria from suboxic waters in the Baltic Sea. *FEMS Microbiology Ecology*, *88*(2), 358–371.
- Benway, H. M., Lorenzoni, L., White, A. E., Fiedler, B., Levine, N. M., Nicholson, D. P., DeGrandpre, M. D., Sosik, H. M., Church, M. J., O'Brien, T. D., Leinen, M., Weller, R. A., Karl, D. M., Henson, S. A., & Letelier, R. M. (2019). Ocean time series observations of changing marine ecosystems: An era of integration, synthesis, and societal applications. *Frontiers in Marine Science*, *6*(393), 1–22.
- Bertilsson, S., Berglund, O., Karl, D. M., & Chisholm, S. W. (2003). Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. *Limnology and Oceanography*, *48*(5), 1721–1731.
- Biller, S. J., Berube, P. M., Lindell, D., & Chisholm, S. W. (2015). *Prochlorococcus*: the structure and function of collective diversity. *Nature Reviews Microbiology*, *13*(1), 13–27.
- Blais, M., Tremblay, J.-É., Jungblut, A. D., Gagnon, J., Martin, J., Thaler, M., Lovejoy, C., Blais, M., Tremblay, J.-É., Jungblut, A. D., Gagnon, J., Martin, J., Thaler, M., & Lovejoy, C. (2012). Nitrogen fixation and identification of potential diazotrophs in the Canadian Arctic. *Global Biogeochemical Cycles*, *26*(GB3022), 1–13.
- Boeuf, D., Humily, F., & Jeanthon, C. (2014). Diversity of arctic pelagic bacteria with an emphasis on photoheterotrophs: A review. *Biogeosciences*, *11*(12), 3309–3322.
- Bolaños, L. M., Choi, C. J., Worden, A. Z., Baetge, N., Carlson, C. A., & Giovannoni, S. (2021). Seasonality of the microbial community composition in the North Atlantic. *Frontiers in Marine Science*, *8*(624164), 1–16.
- Bolaños, L. M., Karp-Boss, L., Choi, C. J., Worden, A. Z., Graff, J. R., Haëntjens, N., Chase, A. P., Della Penna, A., Gaube, P., Morison, F., Menden-Deuer, S., Westberry, T. K., O'Malley, R. T., Boss, E., Behrenfeld, M. J., & Giovannoni, S. J. (2020). Small phytoplankton dominate western North Atlantic biomass. *The ISME Journal*, *14*(7), 1663–1674.
- Bolyen, E., Rideout, J. R., Dillon, M. R., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857.

- Bombar, D., Heller, P., Sanchez-Baracaldo, P., Carter, B. J., & Zehr, J. P. (2014). Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *The ISME Journal*, 8(12), 2530–2542.
- Bombar, D., Paerl, R. W., & Riemann, L. (2016). Marine non-cyanobacterial diazotrophs: Moving beyond molecular detection. *Trends in Microbiology*, 24(11), 916–927.
- Bonachela, J. A., Klausmeier, C. A., Edwards, K. F., Litchman, E., & Levin, S. A. (2016). The role of phytoplankton diversity in the emergent oceanic stoichiometry. *Journal of Plankton Research*, 38(4), 1021–1035.
- Bork, P., Bowler, C., De Vargas, C., Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans studies plankton at Planetary scale. *Science*, 348(6237), 873.
- Boyce, D. G., Lewis, M. R., & Worm, B. (2010). Global phytoplankton decline over the past century. *Nature*, 466(7306), 591–596.
- Brand, L. E., Campbell, L., & Bresnan, E. (2012). *Karenia*: The biology and ecology of a toxic genus. *Harmful Algae*, 14, 156–178.
- Braun, P. D., Schulz-Vogt, H. N., Vogts, A., & Nausch, M. (2018). Differences in the accumulation of phosphorus between vegetative cells and heterocysts in the cyanobacterium *Nodularia spumigena*. *Scientific Reports*, 8(1), 1–6.
- Breitbarth, E., Oschlies, A., & LaRoche, J. (2007). Physiological constraints on the global distribution of *Trichodesmium* - effect of temperature on diazotrophy. *Biogeosciences*, 4(1), 53–61.
- Brown, M. B. (1975). 400: A method for combining non-independent, one-sided tests of significance. *Biometrics*, 31(4), 987–992.
- Brownrigg, R. (2018). mapdata: extra map databases, R package version 2.3.0, Original S code by RA Becker & AR Wilks.
- Buttigieg, P. L., Fadeev, E., Bienhold, C., Hehemann, L., Offre, P., & Boetius, A. (2018). Marine microbes in 4D — using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Current Opinion in Microbiology*, 43, 169–185.

- Cabello, A. M., Cornejo-Castillo, F. M., Raho, N., Blasco, D., Vidal, M., Audic, S., de Vargas, C., Latasa, M., Acinas, S. G., & Massana, R. (2016). Global distribution and vertical patterns of a prymnesiophyte–cyanobacteria obligate symbiosis. *The ISME Journal*, *10*(3), 693–706.
- Cabello, A. M., Turk-Kubo, K. A., Hayashi, K., Jacobs, L., Kudela, R. M., & Zehr, J. P. (2020). Unexpected presence of the nitrogen-fixing symbiotic cyanobacterium UCYN-A in Monterey Bay, California. *Journal of Phycology*, *56*(6), 1521–1533.
- Caiola, M. G., Billi, D., & Friedmann, E. I. (1996). Effect of desiccation on envelopes of the cyanobacterium *Chroococidiopsis* sp. (Chroococcales). *European Journal of Phycology*, *31*(1), 97–105.
- Calbet, A., & Landry, M. R. (2004). Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnology and Oceanography*, *49*(1), 51–57.
- Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), 2639–2643.
- Campbell, K., Mundy, C. J., Gosselin, M., Landy, J. C., Delaforge, A., & Rysgaard, S. (2017). Net community production in the bottom of first-year sea ice over the Arctic spring bloom. *Geophysical Research Letters*, *44*(17), 8971–8978.
- Campitelli, E. (2020). *ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'* (R package version 0.4.1).
- Campitelli, E. (2021). *metR: Tools for Easier Analysis of Meteorological Fields* (0.12.0). R package. <https://doi.org/10.5281/zenodo.2593516>
- Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A. F., Carpenter, E. J., Capone, C. J., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., & Michaels, A. F. (2005). Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, *19*(2), 1–17.
- Caputo, A., Nylander, J. A. A., & Foster, R. A. (2019). The genetic diversity and evolution of diatom-diazotroph associations highlights traits favoring symbiont integration. *FEMS Microbiology Letters*, *366*(2), 1–11.

- Carpenter, E. J., Janson, S., Boje, R., Pollehne, F., & Chang, J. (1995). The dinoflagellate *Dinophysis norvegica*: biological and ecological observations in the Baltic Sea. *European Journal of Phycology*, 30, 1–9.
- Cerdan-Garcia, E., Baylay, A., Polyviou, D., Woodward, E. M. S., Wrightson, L., Mahaffey, C., Lohan, M. C., Moore, C. M., Bibby, T. S., & Robidart, J. C. (2021). Transcriptional responses of *Trichodesmium* to natural inverse gradients of Fe and P availability. *The ISME Journal*, 16(4), 1055–1064.
- Chassot, E., Bonhommeau, S., Dulvy, N. K., Mélin, F., Watson, R., Gascuel, D., & Le Pape, O. (2010). Global marine primary production constrains fisheries catches. *Ecology Letters*, 13(4), 495–505.
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*, 38(23), 5315–5316.
- Cheung, S., Zehr, J. P., Xia, X., Tsurumoto, C., Endo, H., Nakaoka, S., Mak, W., Suzuki, K., & Liu, H. (2021). Gamma4: a genetically versatile Gammaproteobacterial *nifH* phylotype that is widely distributed in the North Pacific Ocean. *Environmental Microbiology*, 23(8), 4246–4259.
- Chien, Y. T., & Zinder, S. H. (1996). Cloning, functional organization, transcript studies, and phylogenetic analysis of the complete nitrogenase structural genes (*nifHDK2*) and associated genes in the archaeon *Methanosarcina barkeri* 227. *Journal of Bacteriology*, 178(1), 143–148.
- Chisholm, S. W., Olson, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B., & Welschmeyer, N. A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature*, 334(6180), 340–343.
- Choi, C. J., Bachy, C., Jaeger, G. S., Poirier, C., Sudek, L., Sarma, V. V. S. S., Mahadevan, A., Giovannoni, S. J., & Worden, A. Z. (2017). Newly discovered deep-branching marine plastid lineages are numerically rare but globally distributed. *Current Biology*, 27(1), R15–R16.
- Choi, C. J., Jimenez, V., Needham, D. M., Poirier, C., Bachy, C., Alexander, H., Wilken, S., Chavez, F. P., Sudek, S., Giovannoni, S. J., & Worden, A. Z. (2020). Seasonal and geographical transitions in Eukaryotic phytoplankton community structure in the Atlantic and Pacific Oceans. *Frontiers in Microbiology*, 11(542372), 1–21.

- Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., & Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Applied and Environmental Microbiology*, 71(9), 5362–5370.
- Clayton, S., Dutkiewicz, S., Jahn, O., & Follows, M. J. (2013). Dispersal, eddies, and the diversity of marine phytoplankton. *Limnology and Oceanography: Fluids and Environments*, 3(1), 182–197.
- Cline, M. S., Smoot, M., Cerami, E., et al. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols*, 2(10), 2366–2382.
- Cohan, F. M., & Perry, E. B. (2007). A systematics for discovering the fundamental units of bacterial diversity. *Current Biology*, 17(10), R373–R386.
- Colombo, M., Jackson, S. L., Cullen, J. T., & Orians, K. J. (2020). Dissolved iron and manganese in the Canadian Arctic Ocean: On the biogeochemical processes controlling their distributions. *Geochimica et Cosmochimica Acta*, 277, 150–174.
- Colombo, M., Li, J., Rogalla, B., Allen, S. E., & Maldonado, M. T. (2022). Particulate trace element distributions along the Canadian Arctic GEOTRACES section: shelf-water interactions, advective transport and contrasting biological production. *Geochimica et Cosmochimica Acta*, 323, 183–201.
- Colombo, M., Rogalla, B., Li, J., Allen, S. E., Orians, K. J., & Maldonado, M. T. (2021). Canadian Arctic Archipelago shelf-ocean interactions: A major Iron source to Pacific derived waters transiting to the Atlantic. *Global Biogeochemical Cycles*, 35(10), 1–17.
- Colombo, M., Rogalla, B., Myers, P. G., Allen, S. E., & Orians, K. J. (2019). Tracing dissolved lead sources in the Canadian Arctic: insights from the Canadian Geotraces program. *ACS Earth and Space Chemistry*, 3(7), 1302–1314.
- Comeau, A. M., Douglas, G. M., & Langille, M. G. I. (2017). Microbiome Helper: a custom and streamlined workflow for microbiome research. *MSystems*, 2(1), 1–11.
- Comeau, A. M., Li, W. K. W., Tremblay, J. É., Carmack, E. C., & Lovejoy, C. (2011). Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLOS ONE*, 6(11), 1–12.
- Conover, R., & Mayzaud, P. (1984). Utilization of phytoplankton by zooplankton during the spring bloom in a Nova Scotia inlet. *Can. J. Fish. Aquat. Sci.*, 41, 232–244.

- Conover, S. A. M. (1975). Nitrogen utilization during spring blooms of marine phytoplankton in Bedford Basin, Nova Scotia, Canada. *Marine Biology*, 32(3), 247–261.
- Conroy, B. J., Steinberg, D. K., Song, B., Kalmbach, A., Carpenter, E. J., & Foster, R. A. (2017). Mesozooplankton graze on cyanobacteria in the Amazon River plume and western tropical North Atlantic. *Frontiers in Microbiology*, 8(1436), 1–15.
- Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33(18), 2938–2940.
- Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sánchez-Baracaldo, P., Lima-Mendez, G., Hingamp, P., Alberti, A., Sunagawa, S., Bork, P., de Vargas, C., Raes, J., Bowler, C., Wincker, P., Zehr, J. P., Gasol, J. M., Massana, R., & Acinas, S. G. (2016). Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nature Communications*, 7(11071), 1–9.
- Cornejo-Castillo, F. M., Muñoz-Marín, M. del C., Turk-Kubo, K. A., Royo-Llonch, M., Farnelid, H., Acinas, S. G., & Zehr, J. P. (2019). UCYN-A3, a newly characterized open ocean sublineage of the symbiotic N₂-fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa*. *Environmental Microbiology*, 21(1), 111–124.
- Cram, J. A., Chow, C. E. T., Sachdeva, R., Needham, D. M., Parada, A. E., Steele, J. A., & Fuhrman, J. A. (2015). Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *The ISME Journal*, 9(3), 563–580.
- Crawford, A., Shore, J., & Shan, S. (2022). Measurement of tidal currents using an autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering*, 47(2), 282–294.
- Cullen, J., Doolittle, W., Levin, S., & Li, W. (2007). Patterns and prediction in microbial oceanography. *Oceanography*, 20(2), 34–46.
- Czech, L., Barbera, P., & Stamatakis, A. (2020). Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics*, 36(10), 3263–3265.

- Dalsgaard, T., Thamdrup, B., & Canfield, D. E. (2005). Anaerobic ammonium oxidation (anammox) in the marine environment. *Research in Microbiology*, *156*(4), 457–464.
- Damm, E., Helmke, E., Thoms, S., Schauer, U., Nöthig, E., Bakker, K., & Kiene, R. P. (2010). Methane production in aerobic oligotrophic surface water in the central Arctic Ocean. *Biogeosciences*, *7*(3), 1099–1108.
- Daniels, C. J., Poulton, A. J., Esposito, M., Paulsen, M. L., Bellerby, R., St John, M., & Martin, A. P. (2015). Phytoplankton dynamics in contrasting early stage North Atlantic spring blooms: Composition, succession, and potential drivers. *Biogeosciences*, *12*(8), 2395–2409.
- Dasilva, C. R., Li, W. K. W., & Lovejoy, C. (2014). Phylogenetic diversity of eukaryotic marine microbial plankton on the Scotian Shelf Northwestern Atlantic Ocean. *Journal of Plankton Research*, *36*(2), 344–363.
- De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology*, *90*(12), 3566–3574.
- De Vargas, C., Audic, S., Henry, N., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1–11.
- de Vries, A., & Ripley, B. (2016). *ggdendro: Create dendrograms and tree diagrams using 'ggplot2'* (R package version 0.1-20).
- de Vries, A., & Ripley, B. D. (2022). *ggdendro: Create dendrograms and tree diagrams using "ggplot2"* (version 0.1.23). R package.
- Decelle, J., Romac, S., Stern, R. F., Bendif, E. M., Zingone, A., Audic, S., Guiry, M. D., Guillou, L., Tessier, D., Le Gall, F., Gourvil, P., Dos Santos, A. L., Probert, I., Vaultot, D., de Vargas, C., & Christen, R. (2015). PhytoREF: a reference database of the plastidial *16S* rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Molecular Ecology Resources*, *15*(6), 1435–1445.
- Dekas, A. E., Poretsky, R. S., & Orphan, V. J. (2009). Deep-Sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science*, *326*(5951), 422–426.

- Delmont, T. O., Pierella Karlusich, J. J., Veseli, I., Fuessel, J., Eren, A. M., Foster, R. A., Bowler, C., Wincker, P., & Pelletier, E. (2021). Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *The ISME Journal*, 16(4), 927–936.
- Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö. C., Lee, S. T., Rappé, M. S., MacLellan, S. L., Lüscher, S., & Eren, A. M. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology*, 3(7), 804–813.
- Deming, J. W. (2002). Psychrophiles and polar regions. *Current Opinion in Microbiology*, 5(3), 301–309.
- Desnoves, N., Lin, M., Guo, X., Ma, L., Carreño-Lopez, R., & Elmerich, C. (2003). Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice. *Microbiology*, 149(8), 2251–2262.
- DFO Canada. (2006). *AZMP Bulletin PMZA*.
- Díez, B., Bergman, B., Pedrós-Alió, C., Antó, M., & Snoeijs, P. (2012). High cyanobacterial nifH gene diversity in Arctic seawater and sea ice brine. *Environmental Microbiology Reports*, 4(3), 360–366.
- Dinasquet, J., Ortega-Retuerta, E., Lovejoy, C., & Obernosterer, I. (2018). Editorial: Microbiology of the rapidly changing polar environments. *Frontiers in Marine Science*, 5(154), 1–3.
- Ding, C., Wu, C., Li, L., Pujari, L., Zhang, G., & Sun, J. (2021). Comparison of diazotrophic composition and distribution in the South China Sea and the Western Pacific Ocean. *Biology*, 10(555), 1–22.
- Djurhuus, A., Closek, C. J., Kelly, R. P., Pitz, K. J., Michisaki, R. P., Starks, H. A., Walz, K. R., Andruszkiewicz, E. A., Olesin, E., Hubbard, K., Montes, E., Otis, D., Muller-Karger, F. E., Chavez, F. P., Boehm, A. B., & Breitbart, M. (2020). Environmental DNA reveals seasonal shifts and potential interactions in a marine community. *Nature Communications*, 11(1), 1–9.
- Doney, S. C. (2010). The growing human footprint on coastal and open-Ocean biogeochemistry. *Science*, 328(5985), 1512–1516.

- Dowle, M., & Srinivasan, A. (2021). *data.table: Extension of `data.frame`* (R package version 1.14.2).
- Ducklow, H. W., Doney, S. C., & Steinberg, D. K. (2008). Contributions of long-term research and time-series observations to marine ecology and biogeochemistry. *Annual Review of Marine Science*, *1*, 279–302
- Dunnington, D. (2022). *ggspatial: Spatial Data Framework for ggplot2* (R package version 1.1.6). <https://paleolimbot.github.io/ggspatial/>
- Dutkiewicz, S., Cermeno, P., Jahn, O., Follows, M. J., Hickman, A. A., Taniguchi, D. A. A., & Ward, B. A. (2020). Dimensions of marine phytoplankton diversity. *Biogeosciences*, *17*(3), 609–634.
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLOS Computational Biology*, *7*(10), 1–16.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.
- Edlund, M. B., & Stoermer, E. F. (1993). Resting spores of the freshwater diatoms *Acanthoceras* and *Urosolenia*. *Journal of Paleolimnology*, *9*(1), 55–61.
- Edwards, A., Cameron, K. A., Cook, J. M., Debbonaire, A. R., Furness, E., Hay, M. C., & Rassner, S. M. E. (2020). Microbial genomics amidst the Arctic crisis. *Microbial Genomics*, *6*, 1–20.
- El-Swais, H., Dunn, K. A., Bielawski, J. P., Li, W. K. W., & Walsh, D. A. (2015). Seasonal assemblages and short-lived blooms in coastal north-west Atlantic Ocean bacterioplankton. *Environmental Microbiology*, *17*(10), 3642–3661.
- Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., & Sogin, M. L. (2013). Oligotyping: differentiating between closely related microbial taxa using *16S* rRNA gene data. *Methods in Ecology and Evolution*, *4*(12), 1111–1119.
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.

- Falkowski, P. G. (1994). The role of phytoplankton photosynthesis in global biogeochemical cycles*. *Photosynthesis Research*, 39, 235–258.
- Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., & Taylor, F. J. R. (2004). The evolution of modern eukaryotic phytoplankton. *Science*, 305(5682), 354–360.
- Farnelid, H., Andersson, A. F., Bertilsson, S., Al-Soud, W. A., Hansen, L. H., Sørensen, S., Steward, G. F., Hagström, Å., & Riemann, L. (2011). Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLOS ONE*, 6(4), 1–9.
- Farnelid, H., Turk-Kubo, K., del Carmen Munoz-Marin, M., & Zehr, J. P. (2016). New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquatic Microbial Ecology*, 77(3), 125–138.
- Farnelid, H., Turk-Kubo, K., Ploug, H., Ossolinski, J. E., Collins, J. R., Van Mooy, B. A. S., & Zehr, J. P. (2018). Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *The ISME Journal*, 13(1), 170–182.
- Farnelid, H., Turk-Kubo, K., & Zehr, J. P. (2021). Cell sorting reveals few novel prokaryote and photosynthetic picoeukaryote associations in the oligotrophic ocean. *Environmental Microbiology*, 23(3), 1469–1480.
- Faust, K., & Raes, J. (2016). CoNet app: inference of biological association networks using Cytoscape. *F1000Research*, 5, 1519.
- Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., & Huttenhower, C. (2012). Microbial Co-occurrence Relationships in the Human Microbiome. *PLOS Computational Biology*, 8(7), e1002606.
- Fay, P. (1992). Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiological Reviews*, 56(2), 340–373.
- Fernández-Gómez, B., Díez, B., Polz, M. F., Arroyo, J. I., Alfaro, F. D., Marchandon, G., Sanhueza, C., Fariás, L., Trefault, N., Marquet, P. A., Molina-Montenegro, M. A., Sylvander, P., & Snoeijs-Leijonmalm, P. (2018). Bacterial community structure in a sympagic habitat expanding with global warming: brackish ice brine at 85–90 °N. *The ISME Journal*, 13(2), 316–333.

- Fernández-Méndez, M., Turk-Kubo, K. A., Buttigieg, P. L., Rapp, J. Z., Krumpen, T., Zehr, J. P., & Boetius, A. (2016). Diazotroph diversity in the sea ice, melt ponds, and surface waters of the eurasian basin of the Central Arctic Ocean. *Frontiers in Microbiology*, 7(1884), 1–18.
- Fernández, A., Mouriño, B., Mouriño-Carballido, M., Bode, A., Varela, M., & Marã Nón, E. (2010). Latitudinal distribution of *Trichodesmium* spp. and N₂ fixation in the Atlantic Ocean. *Biogeosciences*, 7, 3167–3176.
- Fletcher-Hoppe, C., Yeh, Y.-C., Raut, Y., Weissman, J. L., & Fuhrman, J. A. (2022). Symbiotic diazotrophic UCYN-A strains co-occurred with El Niño, relaxed upwelling, and varied eukaryotes over 10 years off Southern California Bight. *BioRxiv*, <https://doi.org/10.1101/2022.11.07.514914>.
- Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., & Martiny, A. C. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America*, 110(24), 9824–9829.
- Fonseca-Batista, D., Li, X., Riou, V., Michotey, V., Deman, F., Fripiat, F., Guasco, S., Brion, N., Lemaitre, N., Tonnard, M., Gallinari, M., Planquette, H., Planchon, F., Sarthou, G., Elskens, M., Laroche, J., Chou, L., & Dehairs, F. (2019). Evidence of high N₂ fixation rates in the temperate northeast Atlantic. *Biogeosciences*, 16(5), 999–1017.
- Fragoso, G. M., Poulton, A. J., Yashayaev, I. M., Head, E. J. H., Stinchcombe, M. C., & Purdie, D. A. (2016). Biogeographical patterns and environmental controls of phytoplankton communities from contrasting hydrographical zones of the Labrador Sea. *Progress in Oceanography*, 141, 212–226.
- Frias-Lopez, J., Thompson, A., Waldbauer, J., & Chisholm, S. W. (2009). Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environmental Microbiology*, 11(2), 512–525.
- Fuhrman, J. A., Cram, J. A., & Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology*, 13(3), 133–146.
- Gaby, J. C., & Buckley, D. H. (2012). A comprehensive evaluation of PCR primers to amplify the *nifH* gene of nitrogenase. *PLOS ONE*, 7(7), 1–12.

- Gaby, J. C., & Buckley, D. H. (2014). Database update: a comprehensive aligned *nifH* gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database*, 2014(bau001), 1–8.
- Galand, P. E., Casamayor, E. O., Kirchman, D. L., & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), 22427–22432.
- Gallon, J. R. (1981). The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms. *Trends in Biochemical Sciences*, 6(C), 19–23.
- Gallon, J. R. (1992). Reconciling the incompatible: N₂ fixation and O₂. *New Phytologist*, 122(4), 571–609.
- Gammelsrød, T., Østerhus, S., & Godøy, Ø. (1992). Decadal variations of ocean climate in the Norwegian Sea observed at Ocean Station “Mike” (66°N 2°E). *ICES Mar. Sei. Symp*, 195, 68–75.
- Gast, R. J., Moran, D. M., Dennett, M. R., & Caron, D. A. (2007). Kleptoplasty in an Antarctic dinoflagellate: caught in evolutionary transition? *Environmental Microbiology*, 9(1), 39–45.
- Gentry, R. R., Froehlich, H. E., Grimm, D., Kareiva, P., Parke, M., Rust, M., Gaines, S. D., & Halpern, B. S. (2017). Mapping the global potential for marine aquaculture. *Nature Ecology & Evolution*, 1(9), 1317–1324.
- Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K. W., & Walsh, D. A. (2014). Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *The ISME Journal*, 8(6), 1301–1313.
- GEOTRACES Intermediate Data Product Group. (2021). The GEOTRACES Intermediate Data Product 2021 (IDP2021). NERC EDS British Oceanographic Data Centre NOC. <https://doi.org/10.5285/cf2d9ba9-d51d-3b7c-e053-8486abc0f5fd>
- Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van De Peer, Y., Vandamme, P., Thompson, F. L., Swings, J. (2005). Opinion: re-evaluating prokaryotic species. *Nature Reviews Microbiology*, 3, 733–739.

- Gérikas Ribeiro, C., Lopes dos Santos, A., Marie, D., Pereira Brandini, F., & Vault, D. (2018). Small eukaryotic phytoplankton communities in tropical waters off Brazil are dominated by symbioses between Haptophyta and nitrogen-fixing cyanobacteria. *The ISME Journal*, *12*, 1360–1374.
- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, *345*(6270), 60–63.
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*(2224), 1–6.
- Gradoville, M. R., Cabello, A. M., Wilson, S. T., Turk-Kubo, K. A., Karl, D. M., & Zehr, J. P. (2021). Light and depth dependency of nitrogen fixation by the non-photosynthetic, symbiotic cyanobacterium UCYN-A. *Environmental Microbiology*, *23*(8), 4518–4531.
- Gradoville, M. R., Farnelid, H., White, A. E., Turk-Kubo, K. A., Stewart, B., Ribalet, F., Ferrón, S., Pinedo-Gonzalez, P., Armbrust, E. V., Karl, D. M., John, S., & Zehr, J. P. (2020). Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnology and Oceanography*, *65*(8), 1858–1875.
- Guidi, L., Chaffron, S., Bittner, L., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, *532*(7600), 465–470.
- Guillou, L., Chrétiennot-Dinet, M. J., Medlin, L. K., Claustre, H., Loiseaux-de Goër, S., & Vault, D. (1999). *Bolidomonas*: a new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *Journal of Phycology*, *35*(2), 368–381.
- Haas, S., Robicheau, B. M., Rakshit, S., Tolman, J., Algar, C. K., LaRoche, J., & Wallace, D. W. R. (2021). Physical mixing in coastal waters controls and decouples nitrification via biomass dilution. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(18), 1–10.
- Hagino, K., Onuma, R., Kawachi, M., & Horiguchi, T. (2013). Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae). *PLOS ONE*, *8*(12), 1–11.

- Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., Laroche, J., D'Hondt, S., & Kuypers, M. M. M. (2012). Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *The ISME Journal*, *6*(6), 1238–1249.
- Hamilton, T. L., Boyd, E. S., & Peters, J. W. (2011). Environmental constraints underpin the distribution and phylogenetic diversity of *nifH* in the Yellowstone geothermal complex. *Microbial Ecology*, *61*(4), 860–870.
- Harding, K., Turk-Kubo, K. A., Sipler, R. E., Mills, M. M., Bronk, D. A., & Zehr, J. P. (2018). Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(52), 13371–13375.
- Harris, R. (2010). The L4 time-series: the first 20 years. *Journal of Plankton Research*, *32*(5), 577–583.
- Harrison, P. J., Zingone, A., Mickelson, M. J., Lehtinen, S., Ramaiah, N., Kraberg, A. C., Sun, J., McQuatters-Gollop, A., & Jakobsen, H. H. (2015). Cell volumes of marine phytoplankton from globally distributed coastal data sets. *Estuarine, Coastal and Shelf Science*, *162*, 130–142.
- Hartmann, M., Zubkov, M. V., Scanlan, D. J., & Lepère, C. (2013). In situ interactions between photosynthetic picoeukaryotes and bacterioplankton in the Atlantic Ocean: Evidence for mixotrophy. *Environmental Microbiology Reports*, *5*(6), 835–840.
- Henke, B. A., Turk-Kubo, K. A., Bonnet, S., & Zehr, J. P. (2018). Distributions and abundances of sublineages of the N₂-fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa* (UCYN-A) in the New Caledonian coral lagoon. *Frontiers in Microbiology*, *9*(554), 1–14.
- Hernández-Salmerón, J. E., & Moreno-Hagelsieb, G. (2020). Progress in quickly finding orthologs as reciprocal best hits: comparing blast, last, diamond and MMseqs2. *BMC Genomics*, *21*(1), 1–9.
- Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T. J., & Barlow, R. G. (2008). An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sensing of Environment*, *112*(6), 3153–3159.
- Hoarfrost, A., Nayfach, S., Ladau, J., Yooseph, S., Arnosti, C., Dupont, C. L., & Pollard, K. S. (2019). Global ecotypes in the ubiquitous marine clade SAR86. *The ISME Journal*, *14*(1), 178–188.

- Hoff, J., Daniel, B., Stukenberg, D., Thuronyi, B. W., Waldminghaus, T., & Fritz, G. (2020). *Vibrio natriegens*: an ultrafast-growing marine bacterium as emerging synthetic biology chassis. *Environmental Microbiology*, 22(10), 4394–4408.
- Holligan, P. M., Fernández, E., Aiken, J., Balch, W. M., Boyd, P., Burkill, P. H., Finch, M., Groom, S. B., Malin, G., Muller, K., Purdie, D. A., Robinson, C., Trees, C. C., Turner, S. M., & van der Wal, P. (1993). A biogeochemical study of the coccolithophore, *Emiliana huxleyi*, in the North Atlantic. *Global Biogeochemical Cycles*, 7(4), 879–900.
- Hunter-Cevera, K. R., Neubert, M. G., Olson, R. J., Solow, A. R., Shalapyonok, A., & Sosik, H. M. (2016). Physiological and ecological drivers of early spring blooms of a coastal phytoplankter. *Science*, 354(6310), 326–329.
- Hutchins, D. A., & Capone, D. G. (2022). The marine nitrogen cycle: new developments and global change. *Nature Reviews Microbiology* 2022, 1–14.
- Hutchins, D. A., & Fu, F. (2017). Microorganisms and ocean global change. *Nature Microbiology*, 2(6), 1–11.
- Irion, S., Christaki, U., Berthelot, H., L’Helguen, S., & Jardillier, L. (2021). Small phytoplankton contribute greatly to CO₂-fixation after the diatom bloom in the Southern Ocean. *The ISME Journal*, 15(9), 2509–2522.
- Jabir, T., Vipindas, P. V., Krishnan, K. P., & Mohamed Hatha, A. A. (2021). Abundance and diversity of diazotrophs in the surface sediments of Kongsfjorden, an Arctic fjord. *World Journal of Microbiology and Biotechnology*, 37(3), 1–15.
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36, W5–W9.
- Kapili, B. J., & Dekas, A. E. (2021). PPIT: an R package for inferring microbial taxonomy from *nifH* sequences. *Bioinformatics*, 37(16), 2289–2298.
- Karl, D. M., & Lukas, R. (1996). The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep Sea Research Part II: Topical Studies in Oceanography*, 43(2–3), 129–156.

- Karlusich, J. J. P., Ibarbalz, F. M., & Bowler, C. (2020). Phytoplankton in the Tara Ocean. *Annual Review of Marine Sciences*, *12*, 233–265.
- Karlusich, J. J. P., Pelletier, E., Lombard, F., Carsique, M., Dvorak, E., Colin, S., Picheral, M., Cornejo-Castillo, F. M., Acinas, S. G., Pepperkok, R., Karsenti, E., de Vargas, C., Wincker, P., Bowler, C., & Foster, R. A. (2021). Global distribution patterns of marine nitrogen-fixers by imaging and molecular methods. *Nature Communications*, *12*(1), 1–18.
- Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., Ding, H., Martinen, P., Malmstrom, R. R., Stocker, R., Follows, M. J., Stepanauskas, R., & Chisholm, S. W. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science*, *344*(6182), 416–420.
- Keeling, P. J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *The Annual Review of Plant Biology*, *64*, 583–607.
- Kellogg, C. T. E., McClelland, J. W., Dunton, K. H., & Crump, B. C. (2019). Strong seasonality in Arctic estuarine microbial food webs. *Frontiers in Microbiology*, *10*(2628), 1–22.
- Kepkay, P. E., Niven, S. E. H., & Jellett, J. F. (1997). Colloidal organic carbon and phytoplankton speciation during a coastal bloom. *Journal of Plankton Research*, *19*(3), 369–389.
- Kerrigan, E. A., Kienast, M., Thomas, H., & Wallace, D. W. R. (2017). Using oxygen isotopes to establish freshwater sources in Bedford Basin, Nova Scotia, a Northwestern Atlantic fjord. *Estuarine, Coastal and Shelf Science*, *199*, 96–104.
- Kim, J., & Rees, D. C. (1994). Perspectives in biochemistry: nitrogenase and biological nitrogen fixation. *Biochemistry*, *33*(2), 389–397.
- Kling, J. D., Lee, M. D., Fu, F., Phan, M. D., Wang, X., Qu, P., & Hutchins, D. A. (2020). Transient exposure to novel high temperatures reshapes coastal phytoplankton communities. *The ISME Journal*, *14*(2), 413–424.
- Koelling, J., Atamanchuk, D., Karstensen, J., Handmann, P., & Wallace, D. W. R. (2022). Oxygen export to the deep ocean following Labrador Sea Water formation. *Biogeosciences*, *19*(2), 437–454.

- Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., Rooney, A. P., Brambilla, E., Connor, N., Ratcliff, R. M., Nevo, E., & Cohan, F. M. (2008). Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2504–2509.
- König, S., Gros, O., Heiden, S. E., Hinzke, T., Thürmer, A., Poehlein, A., Meyer, S., Vatin, M., Mbéguié-A-Mbéguié, D., Toczny, J., Ponnudurai, R., Daniel, R., Becher, D., Schweder, T., & Markert, S. (2016). Nitrogen fixation in a chemoautotrophic lucinid symbiosis. *Nature Microbiology*, *2*(1), 1–10.
- Kooistra, W. H. C. F., Sarno, D., Balzano, S., Gu, H., Andersen, R. A., & Zingone, A. (2008). Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). *Protist*, *159*(2), 177–193.
- Kraemer, S., Ramachandran, A., Colatrisano, D., Lovejoy, C., & Walsh, D. A. (2019). Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *The ISME Journal*, *14*(1), 79–90.
- Kranck, K., & Milligan, T. G. (1988). Macroflocs from diatoms: in situ photography of particles in Bedford Basin, Nova Scotia. *Marine Ecology Progress Series*, *44*, 183–189.
- Krupke, A., Mohr, W., Laroche, J., Fuchs, B. M., Amann, R. I., & Kuypers, M. M. M. (2015). The effect of nutrients on carbon and nitrogen fixation by the UCYN-A-haptophyte symbiosis. *ISME Journal*, *9*(7), 1635–1647.
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874.
- Lacour, L., Claustre, H., Prieur, L., & D’Ortenzio, F. (2015). Phytoplankton biomass cycles in the North Atlantic subpolar gyre: A similar mechanism for two different blooms in the Labrador Sea. *Geophysical Research Letters*, *42*(13), 5403–5410.
- Lahti, L., Shetty, S., et al. (2019). *microbiome R package*. <http://microbiome.github.io>
- Lalucat, J., Bennisar, A., Bosch, R., García-Valdés, E., & Palleroni, N. J. (2006). Biology of *Pseudomonas stutzeri*. *Microbiology and Molecular Biology Reviews*, *70*(2), 510–547.

- Lalucat, J., Gomila, M., Mulet, M., Zaruma, A., & García-Valdés, E. (2022). Past, present and future of the boundaries of the *Pseudomonas* genus: Proposal of *Stutzerimonas* gen. nov. *Systematic and Applied Microbiology*, *45*(1), 126289.
- Landa, M., Turk-Kubo, K. A., Cornejo-Castillo, F. M., Henke, B. A., & Zehr, J. P. (2021). Critical role of light in the growth and activity of the marine N₂-fixing UCYN-A symbiosis. *Frontiers in Microbiology*, *12*(666739), 1–13.
- Langlois, R., Großkopf, T., Mills, M., Takeda, S., & LaRoche, J. (2015). Widespread distribution and expression of Gamma A (UMB), an uncultured, diazotrophic, γ -proteobacterial *nifH* phylotype. *PLOS ONE*, *10*(6), 1–17.
- Langlois, R. J., Hümmer, D., & LaRoche, J. (2008). Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean. *Applied and Environmental Microbiology*, *74*(6), 1922–1931.
- Langlois, R. J., Mills, M. M., Ridame, C., Croot, P., & LaRoche, J. (2012). Diazotrophic bacteria respond to Saharan dust additions. *Marine Ecology Progress Series*, *470*, 1–14.
- LaRoche, J., & Breitbarth, E. (2005). Importance of the diazotrophs as a source of new nitrogen in the ocean. *Journal of Sea Research*, *53*(1–2), 67–91.
- Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu De Madron, X., Bowler, C., Malviya, S., Thyssen, M., Grégori, G., Rembauville, M., Grosso, O., Poulain, J., De Vargas, C., Pujo-Pay, M., & Conan, P. (2018). Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nature Communications*, *9*(1), 1–12.
- Lehman, P. W. (1981). Comparison of chlorophyll *a* and carotenoid pigments as predictors of phytoplankton biomass. *Marine Biology*, *65*(3), 237–244.
- Lehmann, N., Kienast, M., Granger, J., Bourbonnais, A., Altabet, M. A., & Tremblay, J. (2019). Remote western Arctic nutrients fuel remineralization in Deep Baffin Bay. *Global Biogeochemical Cycles*, *33*(6), 649–667.
- Lehmann, N., Kienast, M., Granger, J., & Tremblay, J. (2022). Physical and biogeochemical influences on nutrients through the Canadian Arctic Archipelago: Insights from nitrate isotope ratios. *Journal of Geophysical Research: Oceans*, *127*(3), 1–24.

- Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116.
- Letunic, I., & Bork, P. (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296.
- Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., & Ettema, T. J. G. (2020). Innovations to culturing the uncultured microbial majority. *Nature Reviews Microbiology*, *19*(4), 225–240.
- Li, W.K.W. (1998). Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnology and Oceanography*, *43*(7), 1746–1753.
- Li, W., Dickie, P., & Spry, J. (1998). Plankton monitoring programme in the Bedford Basin, 1991–1997. In *Canadian Data Report of Fisheries and Aquatic Sciences 1036*. Ocean Sciences Division, Maritimes Region, Fisheries and Oceans Canada.
- Li, W. K. W., & Dickie, P. (2001). Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry*, *44*, 236–246.
- Li, W. K.W., Harrison, W. G., & Head, E. J. H. (2006). Coherent assembly of phytoplankton communities in diverse temperate ocean ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1596), 1953–1960.
- Li, X., Fonseca-Batista, D., Roevros, N., Dehairs, F., & Chou, L. (2018). Environmental and nutrient controls of marine nitrogen fixation. *Progress in Oceanography*, *167*, 125–137.
- Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F., & De Vargas, C. (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 12803–12808.
- Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5970–5975.

- Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmiento, H., Hingamp, P., Ogata, H., de Vargas, C., Lima-Mendez, G., Raes, J., Poulain, J., Jaillon, O., Wincker, P., Kandels-Lewis, S., Karsenti, E., Bork, P., & Acinas, S. G. (2014). Metagenomic *16S* rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology*, *16*(9), 2659–2671.
- Longhurst, A. (1995). Seasonal cycles of pelagic production and consumption. *Progress in Oceanography*, *36*(2), 77–167).
- López-Sandoval, D. C., Rodríguez-Ramos, T., Cermeño, P., & Marañón, E. (2013). Exudation of organic carbon by marine phytoplankton: dependence on taxon and cell size. *Marine Ecology Progress Series*, *477*, 53–60.
- Luca, D. De, Kooistra, W., Sarno, D., Gaonkar, C. C., & Piredda, R. (2019). Global distribution and diversity of *Chaetoceros* (Bacillariophyta, Mediophyceae): integration of classical and novel strategies. *PeerJ*, *7*(e7410), 1–23.
- Lundholm, N., & Hasle, G. R. (2010). *Fragilariopsis* (Bacillariophyceae) of the Northern Hemisphere – morphology, taxonomy, phylogeny and distribution, with a description of *F. pacifica* sp. nov. *Phycologia*, *49*(5), 438–460.
- Luo, Y. W., Doney, S. C., Anderson, L. A., et al. (2012). Database of diazotrophs in global ocean: Abundance, biomass and nitrogen fixation rates. *Earth System Science Data*, *4*(1), 47–73.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12.
- Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B. M., Lavik, G., Schmitz, R. A., LaRoche, J., & Kuypers, M. M. M. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, *1*(11), 1–7.
- Masuda, Y., Yamanaka, Y., Hirata, T., & Nakano, H. (2017). Competition and community assemblage dynamics within a phytoplankton functional group: Simulation using an eddy-resolving model to disentangle deterministic and random effects. *Ecological Modelling*, *343*, 1–14.
- Mayzaud, P., & Taguchi, S. (1979). Spectral and Biochemical Characteristics of the Particulate Matter in Bedford Basin. *Journal of the Fisheries Board of Canada*, *36*(2), 211–218.

- McLachlan, J. L., Seguel, M. R., & Fritz, L. (1994). *Tetretreptia pomquetensis* gen. et sp. nov. (Euglenophyceae): a quadriflagellate, phototrophic marine euglenoid. *Journal of Phycology*, 30(3), 538–544.
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE*, 8(4), 1–11.
- Medlin, L. K., Elwood, H. J., Stickel, S., & Sogin, M. L. (1991). Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyta): evidence for a new species, *Skeletonema pseudocostatum*. *Journal of Phycology*, 27(4), 514–524.
- Meiler, S., Britten, G. L., Dutkiewicz, S., Gradoville, M. R., Moisander, P. H., Jahn, O., & Follows, M. J. (2022). Constraining uncertainties of diazotroph biogeography from *nifH* gene abundance. *Limnology and Oceanography*, 67(4), 816–829.
- Messer, L. F., Doubell, M., Jeffries, T. C., Brown, M. V., & Seymour, J. R. (2015). Prokaryotic and diazotrophic population dynamics within a large oligotrophic inverse estuary. *Aquatic Microbial Ecology*, 74(1), 1–15.
- Meunier, V., Bonnet, S., Pernice, M., Benavides, M., Lorrain, A., Grosso, O., Lambert, C., & Houlbrèque, F. (2019). Bleaching forces coral's heterotrophy on diazotrophs and *Synechococcus*. *ISME Journal*, 13(11), 2882–2886.
- Mills, M. M., Turk-Kubo, K. A., van Dijken, G. L., Henke, B. A., Harding, K., Wilson, S. T., Arrigo, K. R., & Zehr, J. P. (2020). Unusual marine cyanobacteria/haptophyte symbiosis relies on N₂ fixation even in N-rich environments. *The ISME Journal*, 14(10), 2395–2406.
- Mohr, W., Intermaggio, M. P., & LaRoche, J. (2010). Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium *Crocospaera watsonii* WH8501. *Environmental Microbiology*, 12(2), 412–421.
- Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P., & Zehr, J. P. (2010). Unicellular cyanobacterial distributions broaden the oceanic N₂ fixation domain. *Science*, 327(5972), 1512–1514.

- Moisander, P. H., Benavides, M., Bonnet, S., Berman-Frank, I., White, A. E., & Riemann, L. (2017). Chasing after non-cyanobacterial nitrogen fixation in marine pelagic environments. *Frontiers in Microbiology*, 8(1736), 1–8.
- Monier, A., Worden, A. Z., & Richards, T. A. (2016). Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. *Environmental Microbiology Reports*, 8(4), 461–469.
- Montoya, J. P., Holl, C. M., Zehr, J. P., Hansen, A., Villareal, T. A., & Capone, D. G. (2004). High rates of N₂ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature*, 430(7003), 1027–1031.
- Moore, R. M., Grefe, I., Zorz, J., Shan, S., Thompson, K., Ratten, J., & LaRoche, J. (2018). On the relationship between hydrogen saturation in the tropical Atlantic Ocean and nitrogen fixation by the symbiotic diazotroph UCYN-A. *Journal of Geophysical Research: Oceans*, 123(4), 2353–2362.
- Moreira-Coello, V., Mouriño-Carballido, B., Marañón, E., Fernández-Carrera, A., Bode, A., Sintés, E., Zehr, J. P., Turk-Kubo, K., & Varela, M. M. (2019). Temporal variability of diazotroph community composition in the upwelling region off NW Iberia. *Scientific Reports*, 9(1), 1–13.
- Morel, A., Yu-Hwan Ahn, Partensky, F., Vaultot, D., & Claustre, H. (1993). *Prochlorococcus* and *Synechococcus* - a comparative-study of their optical-properties in relation to their size and pigmentation. *Journal of Marine Research*, 51(3), 617–649.
- Moreno-Hagelsieb, G., & Latimer, K. (2008). Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, 24(3), 319–324.
- Morgan, M. (2022). BiocManager: Access the Bioconductor Project Package Repository (1.30.18). R package.
- Mulholland, M. R., Bernhardt, P. W., Blanco-Garcia, J. L., Mannino, A., Hyde, K., Mondragon, E., Turk, K., Moisander, P. H., & Zehr, J. P. (2012). Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank. *Limnology and Oceanography*, 57(4), 1067–1083.
- Müller, K., & Wickham, H. (2022). tibble: Simple Data Frames (R package version 3.1.8). <https://cran.r-project.org/package=tibble>

- Muñoz-Marín, M. D. C., Shilova, I. N., Shi, T., Farnelid, H., Cabello, A. M., & Zehr, J. P. (2019). The transcriptional cycle is suited to daytime N₂ fixation in the unicellular cyanobacterium “*Candidatus Atelocyanobacterium thalassa*” (UCYN-A). *MBio*, *10*(1), 1–17.
- Myung, G. P., Kim, S., Hyung, S. K., Myung, G., Yi, G. K., & Yih, W. (2006). First successful culture of the marine dinoflagellate *Dinophysis acuminata*. *Aquatic Microbial Ecology*, *45*(2), 101–106.
- NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13.
- Nearing, J. T., Douglas, G. M., Comeau, A. M., & Langille, M. G. I. (2018). Denoising the Denoisers: An independent evaluation of microbiome sequence error-correction approaches. *PeerJ*, *2018*(8), 1–22.
- Needham, D. M., Fichot, E. B., Wang, E., Berdjeb, L., Cram, J. A., Fichot, C. G., & Fuhrman, J. A. (2018). Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. *The ISME Journal*, *12*(10), 2417–2432.
- Needham, D. M., Sachdeva, R., & Fuhrman, J. A. (2017). Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *The ISME Journal*, *11*(7), 1614–1629.
- NOAA National Geophysical Data Center. (2009). *ETOPO1 1 Arc-Minute Global Relief Model*. NOAA National Centers for Environmental Information.
- Nogales, B., Lanfranconi, M. P., Piña-Villalonga, J. M., & Bosch, R. (2011). Anthropogenic perturbations in marine microbial communities. *FEMS Microbiology Reviews*, *35*(2), 275–298.
- Oksanen, J., Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., & et al. (2019). *vegan: community ecology package* (R package version 2.5-6). <https://cran.r-project.org/package=vegan>
- Oksanen, J., Simpson, G. L., Blanchet, F. G., et al. (2022). *vegan: Community Ecology Package* (R package version 2.6-2). <https://cran.r-project.org/package=vegan>

- Olson, R. J., Chisholm, S. W., Zettler, E. R., & Armbrust, E. V. (1990). Pigments, size, and distributions of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnology and Oceanography*, *35*(1), 45–58.
- Orcutt, B. N., Sylvan, J. B., Knab, N. J., & Edwards, K. J. (2011). Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiology and Molecular Biology Reviews*, *75*(2), 361–422.
- Ortega-Retuerta, E., Joux, F., Jeffrey, W. H., & Ghiglione, J. F. (2013). Spatial variability of particle-attached and free-living bacterial diversity in surface waters from the Mackenzie River to the Beaufort Sea (Canadian Arctic). *Biogeosciences*, *10*(4), 2747–2759.
- Pajares, S., & Ramos, R. (2019). Processes and microorganisms involved in the marine nitrogen cycle: knowledge and gaps. *Frontiers in Marine Science*, *6*(739), 1–33.
- Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Pedrós-Alió, C. (2006). Marine microbial diversity: can it be determined? *Trends in Microbiology*, *14*(6), 257–263.
- Pedrós-Alió, C., & Manrubia, S. (2016). The vast unknown microbial biosphere. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), 6585–6587.
- Pedrós, C., & Pedrós-Alí, P. (2012). The rare bacterial biosphere. *Annual Review of Marine Science*, *4*, 449–466.
- Percopo, I., Siano, R., Cerino, F., Sarno, D., & Zingone, A. (2011). Phytoplankton diversity during the spring bloom in the northwestern Mediterranean Sea. *Botanica Marina*, *54*(3), 243–267.

- Pommier, T., Canbäck, B., Riemann, L., Boström, K. H., Simu, K., Lundberg, P., Tunlid, A., & Hagström, Å. (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, *16*(4), 867–880.
- Prosser, J. I. (2020). Putting science back into microbial ecology: a question of approach. *Philosophical Transactions of the Royal Society B*, *375*(1798), 1–9.
- 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*(D1), D590–D596.
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Raes, E. J., Tolman, J., Desai, D., Ratten, J. M., Zorz, J., Robicheau, B. M., Haider, D., & LaRoche, J. (2022). Seasonal bacterial niche structures and chemolithoautotrophic ecotypes in a North Atlantic fjord. *Scientific Reports*, *12*(1), 1–13.
- Raes, E. J., van de Kamp, J., Bodrossy, L., Fong, A. A., Riekenberg, J., Holmes, B. H., Erler, D. V., Eyre, B. D., Weil, S. S., & Waite, A. M. (2020). N₂ fixation and new insights into nitrification from the ice-edge to the equator in the South Pacific Ocean. *Frontiers in Marine Science*, *7*(389), 1–20.
- Ramakers, C., Ruijter, J. M., Lekanne Deprez, R. H., & Moorman, A. F. M. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters*, *339*(1), 62–66.
- Randelhoff, A., Lacour, L., Marec, C., Leymarie, E., Lagunas, J., Xing, X., Darnis, G., Penker, C., Sampei, M., Fortier, L., D’Ortenzio, F., Claustre, H., & Babin, M. (2020). Arctic mid-winter phytoplankton growth revealed by autonomous profilers. *Science Advances*, *6*(39), 1–9.
- Ratten, J-M. (2017). The diversity, distribution and potential metabolism of non-cyanobacterial diazotrophs in the North Atlantic Ocean. *Dalhousie University Thesis*. Dalhousie University, Halifax, Nova Scotia, Canada.
- Ratten, J. M., LaRoche, J., Desai, D. K., Shelley, R. U., Landing, W. M., Boyle, E., Cutter, G. A., & Langlois, R. J. (2015). Sources of iron and phosphate affect the distribution of diazotrophs in the North Atlantic. *Deep Sea Research Part II: Topical Studies in Oceanography*, *116*, 332–341.

- Reitmeier, S., Hitch, T. C. A., Treichel, N., Fikas, N., Hausmann, B., Ramer-Tait, A. E., Neuhaus, K., Berry, D., Haller, D., Lagkouvelos, I., & Clavel, T. (2021). Handling of spurious sequences affects the outcome of high-throughput *16S* rRNA gene amplicon profiling. *ISME Communications*, *1*(1), 1–12.
- Resovsky, J. S., Ritzwoller, M. H. ; R. D., Widiyantoro, S., Engdahl, E. R., Grand, D. P., Van Der Hilst, R. D., Gilbert, F., Laske, G., & Masters, G. (1999). Phytoplankton death in the sea. *Nature*, *398*(6725), 293–294.
- Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., & Maury, O. (2013). Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles*, *27*(4), 1046–1058.
- Ridame, C., Dinasquet, J., Hallstrøm, S., Bigeard, E., Riemann, L., Van Wambeke, F., Bressac, M., Pulido-Villena, E., Taillandier, V., Gazeau, F., Tovar-Sanchez, A., Baudoux, A. C., & Guieu, C. (2022). N₂ fixation in the Mediterranean Sea related to the composition of the diazotrophic community and impact of dust under present and future environmental conditions. *Biogeosciences*, *19*(2), 415–435.
- Riemann, L., Farnelid, H., & Steward, G. (2010). Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters. *Aquatic Microbial Ecology*, *61*(3), 235–247.
- Righetti, D., Vogt, M., Gruber, N., Psomas, A., & Zimmermann, N. E. (2019). Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science Advances*, *5*(5), 1–11.
- Rigosi, A., Fleenor, W., & Rueda, F. (2010). State-of-the-art and recent progress in phytoplankton succession modelling. *Environmental Reviews*, *18*(1), 423–440.
- Robicheau, B. M., Tolman, J., Bertrand, E. M., & LaRoche, J. (2022). Highly-resolved interannual phytoplankton community dynamics of the coastal Northwest Atlantic. *ISME Communications*, *2*(1), 1–12.
- Robidart, J. C., Church, M. J., Ryan, J. P., Ascani, F., Wilson, S. T., Bombar, D., Marin, R., Richards, K. J., Karl, D. M., Scholin, C. A., & Zehr, J. P. (2014). Ecogenomic sensor reveals controls on N₂-fixing microorganisms in the North Pacific Ocean. *The ISME Journal*, *8*(6), 1175–1185.
- RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio, Inc.

- RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. (2021.9.1.372). RStudio, PBC. <http://www.rstudio.com/>
- Rubio, L. M., & Ludden, P. W. (2005). Maturation of nitrogenase: A biochemical puzzle. *Journal of Bacteriology*, *187*(2), 405–414.
- Ruijter, J. M., Ramakers, C., Hoogaars, W. M. H., Karlen, Y., Bakker, O., van den hoff, M. J. B., & Moorman, A. F. M. (2009). Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research*, *37*(6), e45–e45.
- Salazar, G., Paoli, L., Alberti, A., et al. (2019). Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell*, *179*(5), 1068–1083.
- Sarno, D., Kooistra, W. H. C. F., Medlin, L. K., Percopo, I., & Zingone, A. (2005). Diversity in the genus *Skeletonema* (bacillariophyceae). ii. an assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *Journal of Phycology*, *41*(1), 151–176.
- Saulia, E., Benavides, M., Henke, B., Turk-Kubo, K., Cooperguard, H., Grosso, O., Desnues, A., Rodier, M., Dupouy, C., Riemann, L., & Bonnet, S. (2020). Seasonal shifts in diazotrophs players: Patterns observed over a two-year time series in the New Caledonian Lagoon (Western Tropical South Pacific Ocean). *Frontiers in Marine Science*, *7*(581755), 1–11.
- Sayers, E. W., Bolton, E. E., Brister, J. R., et al. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26.
- Scavotto, R. E., Dziallas, C., Bentzon-Tilia, M., Riemann, L., & Moisander, P. H. (2015). Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. *Environmental Microbiology*, *17*(10), 3754–3765.
- Schlitzer, R. (2002). Interactive analysis and visualization of geoscience data with Ocean Data View. *Computers & Geosciences*, *28*(10), 1211–1218.
- Schlitzer, R. (2021). *Ocean Data View*. <https://odv.awi.de>
- Schnepf, E., & Elbrächter, M. (1999). Dinophyte chloroplasts and phylogeny - A review. *Grana*, *38*, 81–97.

- Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V., & Lancelot, C. (2005). *Phaeocystis* blooms in the global ocean and their controlling mechanisms: a review. *J. Sea Res.*, *53*, 43–66.
- Schuback, N., Hoppe, C. J. M., Tremblay, J. É., Maldonado, M. T., & Tortell, P. D. (2017). Primary productivity and the coupling of photosynthetic electron transport and carbon fixation in the Arctic Ocean. *Limnology and Oceanography*, *62*(3), 898–921.
- Schvarcz, C. R., Wilson, S. T., Caffin, M., Stancheva, R., Li, Q., Turk-Kubo, K. A., White, A. E., Karl, D. M., Zehr, J. P., & Steward, G. F. (2022). Overlooked and widespread pennate diatom-diazotroph symbioses in the sea. *Nature Communications*, *13*(1), 1–9.
- Selden, C. R., Einarsson, S. V., Lowry, K. E., Crider, K. E., Pickart, R. S., Lin, P., Ashjian, C. J., & Chappell, P. D. (2022). Coastal upwelling enhances abundance of a symbiotic diazotroph (UCYN-A) and its haptophyte host in the Arctic Ocean. *Frontiers in Marine Science*, *9*(877562), 1–8.
- Shan, S., & Sheng, J. (2012). Examination of circulation, flushing time and dispersion in Halifax Harbour of Nova Scotia. *Water Quality Research Journal*, *47*(3–4), 353–374.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.
- Sherwood, O. A., Davin, S. H., Lehmann, N., Buchwald, C., Edinger, E. N., Lehmann, M. F., & Kienast, M. (2021). Stable isotope ratios in seawater nitrate reflect the influence of Pacific water along the northwest Atlantic margin. *Biogeosciences*, *18*(15), 4491–4510.
- Shi, Q., & Wallace, D. (2018). A 3-year time series of volatile organic iodocarbons in Bedford Basin, Nova Scotia: A northwestern Atlantic fjord. *Ocean Science*, *14*(6), 1385–1403.
- Shi, T., & Falkowski, P. G. (2008). Genome evolution in cyanobacteria: The stable core and the variable shell. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2510–2515.

- Shiozaki, T., Nagata, T., Ijichi, M., & Furuya, K. (2015). Nitrogen fixation and the diazotroph community in the temperate coastal region of the northwestern North Pacific. *Biogeosciences*, *12*, 4751–4764.
- Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., Ehama, M., Hamasaki, K., & Furuya, K. (2017). Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. *Global Biogeochemical Cycles*, *31*(6), 996–1009.
- Shiozaki, T., Fujiwara, A., Ijichi, M., Harada, N., Nishino, S., Nishi, S., Nagata, T., & Hamasaki, K. (2018). Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnology and Oceanography*, *63*(5), 2191–2205.
- Siddig, A. A. H., Ellison, A. M., Ochs, A., Villar-Leeman, C., & Lau, M. K. (2016). How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in Ecological Indicators. *Ecological Indicators*, *60*, 223–230.
- Simons, R.A. (2019). *ERDDAP*. NOAA/NMFS/SWFSC/ERD.
<https://coastwatch.pfeg.noaa.gov/erddap>
- Sipler, R. E., Gong, D., Baer, S. E., Sanderson, M. P., Roberts, Q. N., Mulholland, M. R., & Bronk, D. A. (2017). Preliminary estimates of the contribution of Arctic nitrogen fixation to the global nitrogen budget. *Limnology and Oceanography Letters*, *2*(5), 159–166.
- Slowikowski, K. (2020). ggrepel: Automatically Position Non-Overlapping Text Labels with ‘ggplot2’ (R package version 0.8.2).
- Slowikowski, K. (2021). ggrepel: Automatically Position Non-Overlapping Text Labels with “ggplot2” (version 0.9.1). R package.
- Smith, D., Scott, J., Steele, A., Cody, G., Ohara, S., & Fogel, M. (2014). Effects of metabolism and physiology on the production of okenone and bacteriochlorophyll *a* in purple sulfur bacteria. *Geomicrobiology Journal*, *31*(2), 128–137.
- Sohm, J. A., Ahlgren, N. A., Thomson, Z. J., Williams, C., Moffett, J. W., Saito, M. A., Webb, E. A., & Rocap, G. (2016). Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *The ISME Journal*, *10*(2), 333–345.

- Sohm, J. A., Webb, E. A., & Capone, D. G. (2011). Emerging patterns of marine nitrogen fixation. *Nature Reviews Microbiology*, *9*(7), 499–508.
- South, A. (2017). rnatuarearth: World Map Data from Natural Earth (R package version 0.1.0).
- Stal, L. J. (2009). Is the distribution of nitrogen-fixing cyanobacteria in the oceans related to temperature? *Environmental Microbiology*, *11*(7), 1632–1645.
- Stal, L. J. (2017). The effect of oxygen concentration and temperature on nitrogenase activity in the heterocystous cyanobacterium *Fischerella* sp. *Scientific Reports*, *7*(1), 1–10.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313.
- Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., & Knap, A. H. (2001). Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(8–9), 1405–1447.
- Stenegren, M., Caputo, A., Berg, C., Bonnet, S., & Foster, R. A. (2018). Distribution and drivers of symbiotic and free-living diazotrophic cyanobacteria in the western tropical South Pacific. *Biogeosciences*, *15*(5), 1559–1578.
- Strohl, W. R., Schmidt, T. M., & Lawry, N. H. (1986). Characterization of *Vitreoscilla beggiatoides* and *Vitreoscilla filiformis* sp. nov., nom. rev., and comparison with *Vitreoscilla stercoraria* and *Beggiatoa alba*. *International Journal of Systematic Bacteriology*, *36*(2), 302–313.
- Sun, J., & Liu, D. (2003). Geometric models for calculating cell biovolume and surface area for phytoplankton. *Journal of Plankton Research*, *25*(11), 1331–1346.
- Sunagawa, S., Coelho, L. P., Chaffron, S., et al. (2015). Structure and function of the global ocean microbiome. *Science*, *348*(6237), 1–9.
- Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*, W609–W612.

- Suzuki, S., Kawachi, M., Tsukakoshi, C., Nakamura, A., Hagino, K., Inouye, I., & Ishida, K. I. (2021). Unstable relationship between *Braarudosphaera bigelowii* (= *Chrysochromulina parkeae*) and its nitrogen-fixing endosymbiont. *Frontiers in Plant Science*, *12*, 749895–749895.
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027.
- Tang, W., & Cassar, N. (2019). Data-driven modeling of the distribution of diazotrophs in the global ocean. *Geophysical Research Letters*, *46*(21), 12258–12269.
- Tang, W., Cerdán-García, E., Berthelot, H., Polyviou, D., Wang, S., Baylay, A., Whitby, H., Planquette, H., Mowlem, M., Robidart, J., & Cassar, N. (2020). New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods. *The ISME Journal*, *14*(10), 2514–2526.
- Tang, W., Li, Z., & Cassar, N. (2019). Machine learning estimates of global marine nitrogen fixation. *Journal of Geophysical Research: Biogeosciences*, *124*(3), 717–730.
- Tang, W., Wang, S., Fonseca-Batista, D., Dehairs, F., Gifford, S., Gonzalez, A. G., Gallinari, M., Planquette, H., Sarthou, G., & Cassar, N. (2019). Revisiting the distribution of oceanic N₂ fixation and estimating diazotrophic contribution to marine production. *Nature Communications*, *10*(1), 1–10.
- Tao, J., Wang, S., Liao, T., & Luo, H. (2021). Evolutionary origin and ecological implication of a unique *nif* island in free-living *Bradyrhizobium* lineages. *The ISME Journal*, *15*(11), 3195–3206.
- Thamdrup, B., & Dalsgaard, T. (2002). Production of N₂ through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Applied and Environmental Microbiology*, *68*(3), 1312–1318.
- Thompson, A., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F., & Zehr, J. P. (2014). Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environmental Microbiology*, *16*(10), 3238–3249.
- Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vault, D., Kuypers, M. M. M., & Zehr, J. P. (2012). Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science*, *337*(6101), 1546–1550.

- Thompson, A. W., Kouba, K., & Ahlgren, N. A. (2021). Niche partitioning of low-light adapted *Prochlorococcus* subecotypes across oceanographic gradients of the North Pacific Subtropical Front. *Limnology and Oceanography*, *66*(4), 1548–1562.
- Thyng, K. M., Greene, C. A., Hetland, R. D., Zimmerle, H. M., & DiMarco, S. F. (2016). True colors of oceanography: Guidelines for effective and accurate colormap selection. *Oceanography*, *29*(3), 9–13.
- Tomas, C. (1997). *Identifying marine phytoplankton*. Academic Press, Elsevier. London, UK.
- Tomas, C. (1993). *Marine phytoplankton: a guide to naked flagellates and coccolithophorids*. Academic Press, Inc. California, USA.
- Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L., Dugdale, R., Finkel, Z., Iudicone, D., Jahn, O., Guidi, L., Lasbleiz, M., Leblanc, K., Levy, M., & Pondaven, P. (2018). Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience*, *11*(1), 27–37.
- Tripp, H. J., Bench, S. R., Turk, K. A., Foster, R. A., Desany, B. A., Niazi, F., Affourtit, J. P., & Zehr, J. P. (2010). Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature*, *464*(7285), 90–94.
- Tsuchiya, M., Chikaraishi, Y., Nomaki, H., Sasaki, Y., Tame, A., Uematsu, K., & Ohkouchi, N. (2018). Compound-specific isotope analysis of benthic foraminifer amino acids suggests microhabitat variability in rocky-shore environments. *Ecology and Evolution*, *8*(16), 8380–8395.
- Tsuchiya, M., Toyofuku, T., Uematsu, K., Brüchert, V., Collen, J., Yamamoto, H., & Kitazato, H. (2015). Cytologic and genetic characteristics of endobiotic bacteria and kleptoplasts of *Virgulinema fragilis* (Foraminifera). *Journal of Eukaryotic Microbiology*, *62*(4), 454–469.
- Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B., & Zehr, J. P. (2017). Distinct ecological niches of marine symbiotic N₂-fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa* sublineages. *Journal of Phycology*, *53*(2), 451–461.
- Turk-Kubo, K. A., Gradoville, M. R., Cheung, S., Cornejo-Castillo, F. M., Harding, K. J., Morando, M., Mills, M., & Zehr, J. P. (2022). Non-cyanobacterial diazotrophs: global diversity, distribution, ecophysiology, and activity in marine waters. *FEMS Microbiology Reviews*, *fuac046*, 1–25.

- Turk-Kubo, K. A., Karamchandani, M., Capone, D. G., & Zehr, J. P. (2014). The paradox of marine heterotrophic nitrogen fixation: Abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. *Environmental Microbiology*, *16*(10), 3095–3114.
- Turk-Kubo, K. A., Mills, M. M., Arrigo, K. R., van Dijken, G., Henke, B. A., Stewart, B., Wilson, S. T., & Zehr, J. P. (2021). UCYN-A/haptophyte symbioses dominate N₂ fixation in the Southern California Current System. *ISME Communications*, *1*(1), 1–13.
- Vezzulli, L., Brettar, I., Pezzati, E., Reid, P. C., Colwell, R. R., Höfle, M. G., & Pruzzo, C. (2012). Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. *The ISME Journal*, *6*(1), 21–30.
- Vihtakari, M. (2022). ggOceanMaps: Plot Data on Oceanographic Maps using “ggplot2” (R package version 1.3.4). <https://cran.r-project.org/package=ggOceanMaps>
- von Friesen, L. W., & Riemann, L. (2020). Nitrogen fixation in a changing Arctic Ocean: An overlooked source of nitrogen? *Frontiers in Microbiology*, *11*(596426), 1–9.
- Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., & Knight, R. (2016). Improved bacterial *16S* rRNA gene (V4 and V4-5) and fungal Internal Transcribed Spacer marker gene primers for microbial community surveys. *MSystems*, *1*(1), 1–10.
- Wang, D. (2017). ggords: Ordination Visualization (R package version 1.0). <https://github.com/wdy91617/ggords>
- Wang, S., Tang, W., Delage, E., Gifford, S., Whitby, H., González, A. G., Eveillard, D., Planquette, H., & Cassar, N. (2021). Investigating the microbial ecology of coastal hotspots of marine nitrogen fixation in the western North Atlantic. *Scientific Reports*, *11*(5508), 1–14.
- Ward, B. B. (1996). Nitrification and denitrification: probing the nitrogen cycle in aquatic environments. *Microbial Ecology*, *32*, 247–261.
- Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer-Verlag.

- Wickham, H., & Seidel, D. (2020). scales: Scale Functions for Visualization (R package version 1.1.1).
- Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1–20.
- Wickham, H., Averick, M., Bryan, J., et al. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Wickham, H., Hester, J., & Bryan, J. (2022). readr: Read Rectangular Text Data (R package version 2.1.2).
- Wickham, H., & Seidel, D. (2022). scales: Scale Functions for Visualization (1.2.0). R package.
- Wilhelm, S. W., & Suttle, C. A. (1999). Viruses and nutrient cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. *BioScience*, 49(10), 781–788.
- Wilke, C. (2019). cowplot: Streamlined Plot Theme and Plot Annotations for ‘ggplot2’ (R package version 1.0.0). <https://cran.r-project.org/package=cowplot>
- Wilke, C. O. (2020). cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2” (1.1.1). R package. <https://cran.r-project.org/package=cowplot>
- Willis, C., Desai, D., & Laroche, J. (2019). Influence of *16S* rRNA variable region on perceived diversity of marine microbial communities of the Northern North Atlantic. *FEMS Microbiology Letters*, 366(13), 152.
- Wolf, Y. I., & Koonin, E. V. (2012). A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. *Genome Biology and Evolution*, 4(12), 1286–1294.
- Worden, A. Z., Janouskovec, J., McRose, D., Engman, A., Welsh, R. M., Malfatti, S., Tringe, S. G., & Keeling, P. J. (2012). Global distribution of a wild alga revealed by targeted metagenomics. *Current Biology*, 22(17), R675–R677.
- Wu, C., Sun, J., Liu, H., Xu, W., Zhang, G., Lu, H., & Guo, Y. (2021). Evidence of the significant contribution of heterotrophic diazotrophs to nitrogen fixation in the Eastern Indian Ocean during pre-southwest monsoon period. *Ecosystems*, 25, 1066–1083

- Wu, Y., Tang, C., & Hannah, C. (2012). The circulation of eastern Canadian seas. *Progress in Oceanography*, 106, 28–48.
- Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., Van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., Herndl, G. J., Middelburg, J. J., Schouten, S., & Damsté, J. S. S. (2006). Archaeal nitrification in the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, 103(33), 12317–12322.
- Yang, Z., & Bielawski, J. R. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, 15(12), 496–503.
- Yoo, Y. Du, Seong, K. A., Kim, H. S., Jeong, H. J., Yoon, E. Y., Park, J., Kim, J. I., Shin, W., & Palenik, B. (2018). Feeding and grazing impact by the bloom-forming euglenophyte *Eutreptiella eupharyngea* on marine eubacteria and cyanobacteria. *Harmful Algae*, 73, 98–109.
- Yool, A., Martin, A. P., Fernández, C., & Clark, D. R. (2007). The significance of nitrification for oceanic new production. *Nature* 2007, 447(7147), 999–1002.
- Zablen, L. B., Kissil, M. S., Woese, C. R., & Buetow, D. E. (1975). Phylogenetic origin of the chloroplast and prokaryotic nature of its ribosomal RNA. *Proceedings of the National Academy of Sciences*, 72(6), 2418–2422.
- Zani, S., Mellon, M. T., Collier, J. L., & Zehr, J. P. (2000). Expression of *nifH* genes in natural microbial assemblages in Lake George, New York, detected by reverse transcriptase PCR. *Applied and Environmental Microbiology*, 66(7), 3119–3124.
- Zehr, J. P., & McReynolds, L. A. (1989). Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Applied and Environmental Microbiology*, 55(10), 2522–2526.
- Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. *Trends in Microbiology*, 19(4), 162–173.
- Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., Tripp, H. J., & Affourtit, J. P. (2008). Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. *Science*, 322(5904), 1110–1112.
- Zehr, J. P., & Capone, D. G. (2020). Changing perspectives in marine nitrogen fixation. *Science*, 368, 1–9.

- Zehr, J. P., & Capone, D. G. (2021a). Biogeography of N₂ fixation in the surface ocean. In *Marine Nitrogen Fixation*. Springer Nature, Switzerland.
- Zehr, J. P., & Capone, D. G. (2021b). Factors Controlling N₂ Fixation. In *Marine Nitrogen Fixation*. Springer Nature, Switzerland.
- Zehr, J. P., Jenkins, B. D., Short, S. M., & Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology*, 5(7), 539–554.
- Zehr, J. P., Shilova, I. N., Farnelid, H. M., Muñoz-Maríncarmen, M. D. C., & Turk-Kubo, K. A. (2016). Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A. *Nature Microbiology*, 2(16214), 1–11.
- Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., Hansen, A., & Karl, D. M. (2001). Unicellular cyanobacteria fix N₂ in the subtropical North Pacific Ocean. *Nature*, 412(6847), 635–638.
- Zehr, J. P., & Kudela, R. M. (2010). Nitrogen cycle of the open ocean: from genes to ecosystems. *Annual Review of Marine Science*, 3, 197–225.
- Zhang, H., Zhan, Y., Yan, Y., Liu, Y., Hu, G., Wang, S., Yang, H., Qiu, X., Liu, Y., Li, J., Lu, W., Elmerich, C., & Lin, M. (2019). The *Pseudomonas stutzeri*-specific regulatory noncoding RNA Nfis targets *katB* mRNA encoding a catalase essential for optimal oxidative resistance and nitrogenase activity. *Journal of Bacteriology*, 201(19), 1–16.
- Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30(5), 614–620.
- Zhang, W., Cao, S., Ding, W., Wang, M., Fan, S., Yang, B., McMinn, A., Wang, M., Xie, B. Bin, Qin, Q. L., Chen, X. L., He, J., & Zhang, Y. Z. (2020). Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. *Microbiome*, 8(1), 1–12.
- Zorz, J., Willis, C., Comeau, A. M., Langille, M. G. I., Johnson, C. L., Li, W. K. W., & LaRoche, J. (2019). Drivers of regional bacterial community structure and diversity in the northwest Atlantic Ocean. *Frontiers in Microbiology*, 10(281), 1–24.

APPENDIX A

Chapter 2 Copyright Permission

August 08, 2022

ISME Communications
ISME Office
Netherlands Institute of Ecology
Droevendaalsesteeg 10, 6708 PB Wageningen, The Netherlands

I am preparing my Doctor of Philosophy (Biology) thesis for submission to the Faculty of Graduate Studies at Dalhousie University, Halifax, Nova Scotia, Canada. I am seeking your permission to include a manuscript version of the following paper(s) as a chapter in the thesis:

Highly-resolved interannual phytoplankton community dynamics of the coastal Northwest Atlantic, authored by Brent M. Robicheau, Jennifer Tolman, Erin M. Bertrand, and Julie LaRoche in *ISME Communications*, volume 2, article number 38, pages 1–12, 2022.

Canadian graduate theses are collected and stored online by the Library and Archives of Canada. I am also seeking your permission for the material described above to be stored online with the LAC. Further details about the LAC thesis program are available on the LAC website (www.bac-lac.gc.ca).

Full publication details and a copy of this permission letter will be included in the thesis.


Yours sincerely,

Brent M. Robicheau

Permission is granted for

- a) the inclusion of the material described above in your thesis
- b) for the material described above to be included in the copy of your thesis that is sent to the Library and Archives of Canada (formerly National Library of Canada) for online storage.

Name REBECCA SHREEVE Title Executive Publisher, Springer Nature
Signature  Date 17/11/2022

Note from the Publisher - Please note that permission is not required to re-use this content because it is covered by a CC-BY license, and copyright is therefore retained by the author 

APPENDIX B

Supplemental Datasets

Supplemental Data A1. Number of reads filtered during the sequence processing pipeline of Chapter 2; data for Bedford Basin V4-V5 *16S* rRNA are shown. Values highlighted red lacked chloroplast and cyanobacterial reads after data processing.

Pipeline Steps >>>>>	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
BB14.03A	51782	48065	47774	47774	19526	18449	295
BB14.03B	5910	5279	5186	5186	2179	2086	77
BB14.03C	13109	11863	11755	11755	5051	4897	75
BB14.03D	30351	27879	27421	27421	11421	11148	67
BB14.04A	43057	39850	33398	33398	12511	12044	713
BB14.04B	31214	28761	25773	25773	9966	9291	977
BB14.04C	37598	34865	34410	34410	12932	12474	688
BB14.04D	58114	53832	53439	53439	21643	20856	403
BB14.05A	29729	27290	27077	27077	10909	10507	823
BB14.05B	23608	21654	21472	21472	8809	8517	685
BB14.05C	20928	19086	18894	18894	7467	7267	766
BB14.05D	26108	23859	22021	22021	8366	8056	53
BB14.06A	30852	28480	28091	28091	10949	10775	1326
BB14.06B	26677	24423	24206	24205	9763	9550	1084
BB14.06C	53201	49031	47770	47770	18708	18383	1135
BB14.06D	30546	27774	26962	26962	9323	9140	64
BB14.07A	29497	27223	27051	27051	10800	10567	532
BB14.07B	18810	17357	17259	17259	7344	7218	454
BB14.07C	21936	20116	19894	19894	8031	7859	251
BB14.07D	24049	22117	21780	21780	8954	8760	30
BB14.08A	53335	49519	49288	49288	14786	14362	1347
BB14.08B	50655	47041	46816	46816	14433	14062	771
BB14.08C	124782	116497	116025	116023	37696	37005	1406
BB14.08D	45528	42266	42028	42028	14240	14000	38
BB14.09A	64552	59790	59532	59532	17497	17102	2398
BB14.09B	64855	60163	59796	59796	20390	20070	1004
BB14.09C	62025	57057	56713	56713	17811	17382	817
BB14.09D	51668	47479	47095	47095	15543	14883	57
BB14.10A	65493	60741	60427	60427	18965	18472	4671
BB14.10B	56492	52037	51707	51707	17215	16874	2459
BB14.10C	49731	46070	45663	45663	14849	14487	586
BB14.10D	49366	45490	45204	45204	14681	14235	106
BB14.11A	10871	9754	9601	9601	4048	3965	391
BB14.11B	13300	11858	11758	11758	5135	5050	273
BB14.11C	16283	14584	14463	14463	6097	5952	286
BB14.11D	37267	33933	33597	33597	12689	11429	414
BB14.12A	15311	13693	13614	13614	5183	4737	101
BB14.12B	8059	7009	6948	6948	2732	2473	49
BB14.12C	41493	38119	37961	37959	13968	13944	34
BB14.12D	37147	34052	33924	33924	12568	12515	28
BB14.13A	27728	25698	25588	25587	11049	10997	65
BB14.13B	29164	27017	26883	26882	11719	11641	111
BB14.13C	37515	34492	34287	34287	14549	14411	121
BB14.13D	33719	31041	30873	30873	12629	12543	66
BB14.14A	33047	30495	30348	30348	12532	12404	378
BB14.14B	31698	29285	29105	29105	11907	11779	578
BB14.14C	16527	15302	15179	15179	6605	6535	614
BB14.14D	35684	33005	32862	32862	14812	14659	106

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		Chloroplast AND Cyanobacterial Reads
BB14.15A	20979	19266	19095	19095	8256	8157	651
BB14.15B	27690	25463	25218	25218	11134	10966	1792
BB14.15C	38838	35798	35458	35457	15213	14952	1063
BB14.15D	39118	36036	35807	35807	14990	14802	58
BB14.16A	64627	60276	60080	60079	15060	14627	0
BB14.16B	79112	73305	73038	73038	20191	19739	0
BB14.16C	53691	50013	49802	49802	12578	11990	11
BB14.16D	90604	84128	83832	83832	23492	23080	0
BB14.17A	37741	35159	34997	34997	10522	10252	2663
BB14.17B	63560	59450	59245	59245	24533	24363	19881
BB14.17C	64511	59906	59672	59671	21736	21415	14073
BB14.17D	47784	44215	43936	43936	13066	12592	67
BB14.18A	46694	43570	43401	43401	12721	12326	654
BB14.18B	45432	42285	42100	42100	13023	12779	326
BB14.18C	38416	35849	35556	35556	11607	11375	488
BB14.18D	49952	46484	46263	46263	15283	15059	113
BB14.19A	31183	28888	28702	28702	9013	8818	424
BB14.19B	48956	45447	45151	45151	14506	14259	2857
BB14.19C	52701	48737	48268	48268	15153	14814	1197
BB14.19D	39520	36430	36147	36147	11520	11270	160
BB14.20A	53586	49933	49657	49657	16830	16602	3639
BB14.20B	36077	33400	33181	33181	10984	10767	3056
BB14.20C	52935	49172	48127	48126	15778	15525	3498
BB14.20D	40489	37545	37334	37334	11085	10864	39
BB14.21A	47912	44513	44322	44322	10920	10449	65
BB14.21B	27833	25665	25532	25532	7889	7799	29
BB14.21C	52151	48066	47812	47812	13930	13629	342
BB14.21D	37302	34226	34038	34038	11034	10810	7
BB14.22A	61508	57157	56929	56929	15198	14631	230
BB14.22B	48507	44825	44597	44597	12631	12345	223
BB14.22C	51201	47648	47450	47448	13082	12709	285
BB14.22D	32031	29630	29466	29466	9078	8934	23
BB14.23A	29663	27621	27471	27471	7134	6877	126
BB14.23B	30968	28800	28663	28663	8465	8286	76
BB14.23C	36796	33957	33735	33735	9966	9735	319
BB14.23D	21954	20064	19914	19914	7180	7095	17
BB14.24A	22850	21160	21016	21015	6637	6466	311
BB14.24B	15264	14019	13920	13920	4111	4003	31
BB14.24C	34489	32078	31927	31927	8547	8270	51
BB14.24D	64279	59628	59316	59314	19995	19689	97
BB14.25A	41029	38286	38182	38182	10260	9819	53
BB14.25B	25645	23924	23842	23842	6455	6152	131
BB14.25C	31580	29243	29121	29121	7373	7059	66
BB14.25D	27147	25147	24999	24999	8453	8240	81
BB14.26A	30322	28277	28147	28147	7888	7621	463
BB14.26B	36573	33989	33852	33852	8664	8256	63
BB14.26C	31451	29445	29336	29336	8244	7946	63
BB14.26D	29544	27525	27356	27355	9892	9750	104
BB14.27A	40449	37622	37409	37409	12293	12172	677
BB14.27B	23814	22197	22105	22105	5895	5700	172
BB14.27C	34522	32039	31910	31910	8441	8177	181
BB14.27D	40117	37340	37140	37139	10086	9735	27
BB14.28A	7814	7154	7093	7093	2820	2788	186
BB14.28B	48145	44431	44244	44244	10555	10008	126
BB14.28C	37220	34681	34537	34537	8814	8581	211
BB14.28D	48328	45020	44514	44514	11035	10373	2
BB14.29A	283	163	140	140	6	4	0
BB14.29B	540	418	397	397	100	96	0
BB14.29C	9609	8814	8765	8765	3475	3452	1450
BB14.29D	44860	41722	41522	41522	11865	11629	122
BB14.30A	22585	20772	20677	20677	7753	7681	106
BB14.30B	22328	20310	20217	20217	6952	6850	73

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB14.30C	15888	14868	14812	14812	4174	4093	14
BB14.30D	67086	62234	61867	61865	12452	11141	37
BB14.31A	17311	16117	16046	16046	5459	5398	420
BB14.31B	37144	34794	34692	34692	10095	9880	171
BB14.31C	17036	15845	15802	15802	4822	4707	205
BB14.31D	39541	36885	36690	36690	9968	9721	15
BB14.32A	14064	12985	12933	12933	4273	4195	532
BB14.32B	18167	16557	16495	16495	5149	4979	172
BB14.32C	12926	12115	12073	12073	4408	4361	416
BB14.32D	44952	41926	41765	41765	10393	9961	43
BB14.33A	66893	62431	62178	62178	20243	19974	624
BB14.33B	65769	61413	61155	61153	19381	19012	480
BB14.33C	16227	15038	14962	14961	4370	4246	121
BB14.33D	36848	34257	34033	34033	10082	9829	33
BB14.34A	17167	15771	15660	15660	5310	5214	562
BB14.34B	19810	18057	17963	17963	5681	5553	698
BB14.34C	14691	13741	13667	13667	4339	4260	161
BB14.34D	44528	41325	41091	41091	12449	12107	33
BB14.35A	39868	37348	37229	37229	12778	12597	3327
BB14.35B	8858	8241	8196	8196	3249	3215	955
BB14.35C	21908	20261	20181	20181	6486	6366	717
BB14.35D	40295	37468	37254	37254	12299	12064	19
BB14.36A	32812	30616	30513	30513	9711	9462	3320
BB14.36B	3160	2681	2662	2662	966	945	187
BB14.36C	20085	18708	18650	18650	5766	5595	379
BB14.36D	41798	39036	38890	38890	10914	10470	14
BB14.37A	12228	11383	11318	11318	2835	2736	309
BB14.37B	6946	6443	6406	6406	2118	2076	272
BB14.37C	17663	16431	16362	16362	4310	4132	574
BB14.37D	24365	22598	22363	22363	7004	6847	9
BB14.38A	5912	5448	5426	5426	1854	1816	315
BB14.38B	18809	17297	17225	17225	5155	5018	1050
BB14.38C	16598	15466	15415	15415	4792	4712	1056
BB14.38D	19673	18191	17978	17978	6095	5992	14
BB14.39A	15111	13983	13909	13909	4628	4512	190
BB14.39B	12333	11356	11303	11303	4060	3943	316
BB14.39C	17928	16582	16418	16418	5701	5552	2263
BB14.39D	32489	30024	29720	29719	9270	9061	56
BB14.40A	16444	15079	14840	14840	5229	5044	300
BB14.40B	17391	15608	15191	15191	4784	4631	141
BB14.40C	12787	11757	11687	11687	3862	3761	105
BB14.40D	29732	27550	27318	27318	8494	8268	32
BB14.41A	35378	32946	32809	32809	10094	9929	63
BB14.41B	23885	22149	22047	22047	7454	7351	105
BB14.41C	32680	30364	30218	30218	9355	9116	254
BB14.41D	20311	18858	18720	18719	5430	5258	15
BB14.42A	28603	26571	26450	26450	8264	8114	58
BB14.42B	28001	25707	25587	25587	7964	7780	55
BB14.42C	8560	7977	7939	7938	3120	3098	54
BB14.42D	23364	21645	21530	21530	6696	6502	14
BB14.43A	26246	24535	24440	24440	7601	7431	549
BB14.43B	14562	13563	13502	13502	4230	4116	167
BB14.43C	12798	11886	11833	11832	3795	3674	48
BB14.43D	29735	27766	27586	27586	8010	7843	9
BB14.44A	12507	11662	11616	11615	3579	3423	247
BB14.44B	8816	7970	7929	7929	2823	2747	171
BB14.44C	5674	5292	5268	5268	2029	2001	53
BB14.44D	11357	10538	10374	10374	3184	3089	0
BB14.45A	24346	22746	22614	22614	7180	6967	1097
BB14.45B	9294	8730	8707	8707	2887	2813	338
BB14.45C	3390	3167	3152	3152	1190	1167	63
BB14.45D	34417	32089	31810	31810	9669	9506	11

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB14.46A	10888	10181	10154	10154	2834	2664	66
BB14.46B	25584	23829	23648	23648	6853	6534	740
BB14.46C	12024	11214	11156	11156	3484	3376	299
BB14.46D	35678	33340	32872	32872	9501	9263	15
BB14.47A	110	88	81	81	0	0	0
BB14.47B	73	48	45	45	0	0	0
BB14.47C	24086	22601	22533	22532	6558	6279	411
BB14.47D	35809	33540	33355	33355	9783	9478	16
BB14.48A	34548	32252	32074	32074	10402	10085	818
BB14.48B	57283	53786	53537	53536	15002	14247	763
BB14.48C	14572	13649	13565	13565	4519	4414	158
BB14.48D	15653	14580	14145	14145	4577	4376	17
BB14.49A	29088	27283	27151	27151	6672	6321	757
BB14.49B	11092	10414	10368	10368	2648	2531	288
BB14.49C	32687	30632	30493	30493	7712	7260	665
BB14.49D	80204	74967	74583	74583	22001	21367	49
BB14.50A	37368	34942	34827	34827	9208	8668	805
BB14.50B	27812	26002	25898	25898	6979	6565	422
BB14.50C	29937	27995	27892	27892	6884	6478	250
BB14.50D	13797	12807	12573	12573	4669	4564	8
BB14.51A	1411	1321	1306	1306	376	362	2
BB14.51B	9968	9229	9177	9177	2992	2930	85
BB14.51C	10137	9381	9281	9281	3312	3251	43
BB14.51D	13002	12092	11878	11878	4142	4065	11
BB15.02A	106849	102158	101696	101675	43767	42441	734
BB15.02B	79350	75879	75548	75530	34996	33992	721
BB15.02C	92035	88105	87791	87772	40080	38855	1258
BB15.02D	87279	83426	82681	82662	38895	37855	68
BB15.03A	79546	76354	76075	76055	31979	30791	2901
BB15.03B	50134	48036	47815	47807	22029	21597	1199
BB15.03C	81424	78105	77323	77299	39130	38025	202
BB15.03D	63277	60468	60024	60008	26566	26126	2029
BB15.04A	84965	79835	79440	79424	35485	34320	2629
BB15.04B	58575	55128	54863	54853	26104	25184	2000
BB15.04C	60763	57342	57185	57168	26922	26053	2649
BB15.04D	70726	66601	66057	66038	31797	30833	500
BB15.05A	86246	81586	81340	81322	37836	36787	4284
BB15.05B	41051	38669	38466	38464	18364	17659	1044
BB15.05C	49967	47278	46334	46322	21969	21351	1776
BB15.05D	59749	56129	55685	55667	26012	25510	383
BB15.06A	32659	31154	30991	30984	13769	13524	1294
BB15.06B	53103	50555	50376	50368	23694	23302	2271
BB15.06C	66267	63411	63101	63080	27761	27169	2982
BB15.06D	70747	67390	66846	66833	30570	30120	314
BB15.07A	52767	50503	50344	50326	23645	22948	859
BB15.07B	31655	30238	30109	30105	13381	12571	311
BB15.07C	72806	69845	69646	69626	30652	30068	588
BB15.07D	48338	46104	45865	45857	20696	20367	61
BB15.08A	39189	36794	36440	36430	18174	17978	5316
BB15.08B	85300	80849	79817	79796	40242	39805	12334
BB15.08C	70300	66166	65779	65775	28166	27844	8191
BB15.08D	60528	57129	56739	56729	26447	25816	163
BB15.09A	42062	39401	38957	38956	11079	10932	2353
BB15.09B	60557	57534	57040	57026	26597	26166	7549
BB15.09C	69232	66062	65671	65661	33478	33085	8569
BB15.09D	45262	42739	42406	42400	22526	22195	250
BB15.10A	1684	1496	1469	1469	661	644	199
BB15.10B	40440	38356	37935	37925	19366	19133	7518
BB15.10C	64008	60823	58782	58771	28028	27575	11255
BB15.10D	59597	56680	56372	56359	26447	25847	1953
BB15.11A	51622	49038	48126	48116	22935	22597	6390
BB15.11B	36243	34529	34257	34247	16771	16560	5136

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB15.11C	39119	37275	36521	36514	19352	19161	6607
BB15.11D	35630	33659	33352	33341	17459	17308	1371
BB15.12A	29922	28523	28419	28413	13612	13458	2439
BB15.12B	26828	25435	25345	25338	11668	11459	1976
BB15.12C	32281	30810	30747	30740	14230	13954	1722
BB15.12D	48328	45890	45419	45409	18858	18364	697
BB15.13A	35566	34111	34055	34050	15761	15472	1188
BB15.13B	24270	23190	23128	23124	10612	10474	2527
BB15.13C	11713	11261	11244	11241	5817	5776	1814
BB15.13D	28034	26633	26315	26305	11254	11089	807
BB15.14A	1261	876	764	764	28	28	0
BB15.14B	17165	16147	16022	16017	7711	7598	344
BB15.14C	16196	15248	15211	15207	7212	7064	35
BB15.14D	47174	44300	44062	44052	21833	21515	280
BB15.15A	44285	41690	41236	41221	17323	16709	2627
BB15.15B	18335	17311	17202	17202	8954	8853	688
BB15.15C	26418	25070	25002	24997	11606	11452	3344
BB15.15D	27272	25523	25239	25233	12871	12744	419
BB15.16A	595	491	449	449	24	21	0
BB15.16B	12661	11914	11862	11860	5728	5627	38
BB15.16C	27207	25489	25248	25245	12916	12694	162
BB15.16D	23034	21574	21454	21450	10395	10234	55
BB15.17A	386	354	342	342	99	85	0
BB15.17B	5185	4898	4867	4867	2239	2215	4
BB15.17C	11591	11024	10992	10991	5355	5274	77
BB15.17D	15109	14163	14107	14106	7076	7008	254
BB15.18A	11703	10998	10949	10946	5420	5368	45
BB15.18B	14150	13628	13578	13575	6853	6784	22
BB15.18C	28537	27465	27185	27176	12321	12158	13
BB15.18D	27465	26160	25920	25912	11915	11721	66
BB15.19A	4914	4708	4689	4687	2401	2375	27
BB15.19B	564	538	526	526	233	228	7
BB15.19C	17954	17362	17322	17318	8303	8218	70
BB15.19D	17186	16468	16379	16375	8937	8887	261
BB15.20A	24187	22368	22259	22257	10601	10397	396
BB15.20B	40362	38363	38033	38026	17106	16857	143
BB15.20C	23556	22449	22304	22294	10895	10757	33
BB15.20D	27590	26320	26250	26246	10985	10764	4
BB15.21A	31641	30064	29916	29911	13734	13513	5372
BB15.21B	24101	23039	22932	22923	11083	10998	3238
BB15.21C	30649	29277	28416	28409	14393	14255	1800
BB15.21D	14367	13693	13593	13592	7156	7087	35
BB15.22A	26055	24596	24494	24487	9931	9733	39
BB15.22B	31379	29855	29755	29751	12303	12080	132
BB15.22C	37839	36008	35784	35773	16052	15783	1765
BB15.22D	19620	18552	18266	18261	9272	9085	57
BB15.23A	22865	21723	21664	21660	9588	9288	74
BB15.23B	27879	26568	26511	26505	11007	10781	97
BB15.23C	30495	29086	28957	28949	13629	13439	303
BB15.23D	21138	20010	19924	19918	10257	10183	56
BB15.24A	47178	44733	44594	44583	20353	20180	20
BB15.24B	32953	30977	30817	30812	12426	12157	95
BB15.24C	32765	30992	30862	30855	12800	12497	192
BB15.24D	22189	20943	20789	20782	10006	9809	195
BB15.25A	22926	21748	21682	21672	10260	10056	642
BB15.25B	22088	21045	20982	20978	8651	8496	50
BB15.25C	20319	19356	19253	19247	10375	10341	17
BB15.25D	107860	103332	102772	102771	21801	18713	26
BB15.26A	33065	31484	31182	31171	15657	15322	2420
BB15.26B	21833	20847	20667	20661	11706	11625	4678
BB15.26C	19218	18443	18382	18379	9695	9618	1470
BB15.26D	41645	39496	38436	38429	18962	18667	119

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB15.27A	45304	43322	43184	43170	21720	21491	1563
BB15.27B	886	833	798	798	148	145	12
BB15.27C	48214	46073	45876	45872	22914	22678	2829
BB15.27D	29772	28296	27526	27517	14366	14184	54
BB15.28A	69740	66517	66276	66269	33660	33321	1430
BB15.28B	49855	47661	47504	47491	22431	22170	679
BB15.28C	29702	28487	28397	28388	14210	14014	403
BB15.28D	44080	41900	41103	41090	20838	20525	81
BB15.29A	48967	46708	46537	46527	20367	19383	448
BB15.29B	62837	60194	59842	59828	28582	28344	1326
BB15.29C	42740	40980	40412	40401	19890	19760	303
BB15.29D	31477	29710	29486	29482	13878	13730	41
BB15.30A	53466	51048	50591	50573	22833	22514	1276
BB15.30B	68776	65664	65240	65227	30676	30359	1132
BB15.30C	57345	54619	54043	54034	26326	26029	288
BB15.30D	39108	37084	36359	36351	19290	18947	39
BB15.31A	46659	44572	43932	43923	24120	23886	2752
BB15.31B	593	558	529	529	160	145	2
BB15.31C	58425	55817	55243	55234	28979	28759	738
BB15.31D	36083	34249	34020	34012	17893	17662	11
BB15.32A	39032	36952	36591	36585	17617	17325	2892
BB15.32B	15881	15063	14977	14974	7899	7795	1164
BB15.32C	29097	27305	26703	26692	13825	13624	432
BB15.32D	36861	34625	34365	34350	16107	15575	20
BB15.33A	31520	29625	27253	27248	6787	3074	977
BB15.33B	46130	43704	43353	43344	20759	20486	1575
BB15.33C	53830	51124	50931	50920	20371	19653	392
BB15.33D	36546	34294	33777	33773	18621	18305	36
BB15.34A	61296	58254	57741	57732	24374	23918	1940
BB15.34B	47819	45432	45029	45014	18871	18439	678
BB15.34C	38953	36963	36211	36204	16079	15708	1255
BB15.34D	43721	41429	39611	39603	18617	18232	77
BB15.35A	45698	43279	43064	43054	19373	19023	1573
BB15.35B	37251	35549	35369	35359	15766	15474	809
BB15.35C	40821	38947	38375	38364	17278	16958	1704
BB15.35D	25658	24191	23844	23842	12241	12098	68
BB15.36A	36570	34896	34600	34592	15533	15237	3272
BB15.36B	18666	17795	17647	17646	8213	8088	1825
BB15.36C	62675	59712	58900	58878	24520	23827	3186
BB15.36D	47501	45008	44590	44581	21836	21502	23
BB15.37A	47817	45827	45647	45636	23208	22968	5194
BB15.37B	33183	31819	31679	31673	14019	13752	2360
BB15.37C	48915	46883	46397	46388	20286	19908	1092
BB15.37D	25876	24596	24403	24395	13123	13019	16
BB15.38A	44653	42577	41830	41825	21825	19514	1049
BB15.38B	13015	12334	11024	11023	5387	5253	403
BB15.38C	68516	65257	63168	63147	31570	31019	3419
BB15.38D	71581	67750	67200	67189	33916	33272	61
BB15.39A	70435	67121	66064	66049	31392	30801	1762
BB15.39B	60195	57316	56421	56412	24486	23914	1324
BB15.39C	60266	57630	56939	56933	26316	25750	978
BB15.39D	30422	28722	28346	28340	14713	14562	19
BB15.40A	34356	32807	32582	32576	13781	13465	701
BB15.40B	33270	31803	31667	31660	14615	14404	606
BB15.40C	26794	25638	25527	25519	12396	12222	518
BB15.40D	41014	38829	35865	35864	19005	18670	79
BB15.41A	27859	26595	26428	26420	11551	11131	599
BB15.41B	48774	46532	45088	45079	21062	20694	1134
BB15.41C	47260	45113	44072	44059	20820	20445	707
BB15.41D	44103	41739	41397	41388	21004	20731	46
BB15.42A	23942	22651	22505	22496	8933	8598	282
BB15.42B	27986	26397	26185	26180	11797	11573	361

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB15.42C	40988	38713	37576	37563	17213	16860	463
BB15.42D	28478	26760	26410	26406	12885	12678	18
BB15.43A	58789	55544	53160	53150	24197	23590	1352
BB15.43B	29537	27869	25296	25290	10819	10510	626
BB15.43C	35757	33810	32661	32656	15371	15062	647
BB15.43D	42011	39414	39220	39215	18967	18658	25
BB15.44A	51651	48523	47313	47297	19930	19260	3990
BB15.44B	38076	35748	35266	35260	14962	14425	2069
BB15.44C	29062	27444	27277	27270	12379	11962	1661
BB15.44D	39581	37119	36513	36509	18061	17657	69
BB15.45A	41741	40902	40806	40784	20439	19932	1845
BB15.45B	31570	30956	30717	30699	16797	16545	1446
BB15.45C	39874	39103	38882	38864	22397	21715	3648
BB15.45D	22077	20407	20331	20316	9025	8885	38
BB15.46A	39434	37856	37541	37532	15536	15190	6088
BB15.46B	32628	31173	30892	30887	13590	13404	7852
BB15.46C	29026	27639	27227	27218	12283	12064	5624
BB15.46D	44913	42897	42525	42519	19305	18953	84
BB15.47A	55430	53118	52459	52447	24984	24604	12171
BB15.47B	50468	48256	47611	47604	22926	22663	13551
BB15.47C	51199	49182	48837	48824	23065	22756	13080
BB15.47D	44079	41897	41564	41556	21038	20702	101
BB15.48A	21633	20633	20476	20471	8759	8487	765
BB15.48B	31677	30203	30036	30031	13757	13392	2069
BB15.48C	55283	52544	52288	52276	25220	24827	8685
BB15.48D	90888	86378	85841	85815	44577	43505	232
BB15.49A	47973	45552	45293	45285	20498	20045	1980
BB15.49B	34928	33514	33426	33420	13648	13079	1573
BB15.49C	50032	47735	47533	47517	21271	20684	3450
BB15.49D	141093	138886	138430	138430	62185	60675	331
BB15.50A	64134	60205	59691	59680	32013	31574	9262
BB15.50B	67336	63634	63204	63194	33252	32816	6017
BB15.50C	51205	48166	47437	47426	26146	25884	6298
BB15.50D	43220	40603	40304	40293	21546	21153	78
BB15.51A	85347	84245	83780	83780	45155	44222	14898
BB15.51B	115846	114279	113787	113787	58907	57933	23369
BB15.51C	77705	76474	76168	76168	39131	38537	14039
BB15.51D	149576	147596	146976	146976	71701	70251	90
BB16.01A	51208	48668	48446	48433	22806	22512	11637
BB16.01B	34453	32809	32587	32583	17138	16914	5788
BB16.01C	42962	40762	40421	40414	22412	22161	10221
BB16.01D	34480	32950	32714	32706	16777	16368	61
BB16.02A	38411	36589	36421	36408	17693	17167	3055
BB16.02B	53246	50879	50609	50597	26160	25455	8100
BB16.02C	32183	30701	30548	30544	15784	15545	5028
BB16.02D	51718	49085	48764	48760	21234	19546	68
BB16.03A	53491	50333	50140	50127	26221	25830	15167
BB16.03B	21644	20385	20280	20271	11706	11618	4931
BB16.03C	31844	29867	29709	29697	15773	15535	7036
BB16.03D	47858	45293	44985	44972	23027	22507	331
BB16.04A	75963	71347	70991	70975	34303	33733	7866
BB16.04B	54596	51635	51347	51340	25990	25611	7066
BB16.04C	48877	46148	45926	45917	23534	23252	6681
BB16.04D	59223	55634	55302	55294	25162	24562	678
BB16.05A	49630	47213	46941	46924	20842	20503	3857
BB16.05B	39960	38008	37727	37717	18817	18550	3113
BB16.05C	59113	56098	55556	55550	25090	24630	4020
BB16.05D	81326	77816	77266	77249	36904	36333	325
BB16.06A	46234	44006	43828	43811	20977	20512	2055
BB16.06B	27655	26311	26180	26173	12010	11549	1341
BB16.06C	15149	14438	14386	14385	7235	6976	558
BB16.06D	62708	59705	59327	59313	29631	29203	215

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB16.07A	26512	25121	25023	25018	13796	13677	1438
BB16.07B	43609	41249	41012	41001	22120	21850	2367
BB16.07C	69843	66314	66091	66072	31906	31513	3039
BB16.07D	86700	82562	82247	82227	43922	43285	170
BB16.08A	35385	33408	33309	33301	16709	16358	1532
BB16.08B	35254	33528	33324	33318	18256	17920	2239
BB16.08C	23071	21878	21799	21793	12037	11882	911
BB16.08D	46296	43712	43440	43423	23651	23393	128
BB16.09A	34411	32606	32393	32390	15763	15496	2688
BB16.09B	40478	38353	37991	37984	19581	19212	3638
BB16.09C	25553	24070	23909	23906	12078	11863	1376
BB16.09D	61369	58551	58274	58258	30606	30176	167
BB16.10A	14724	14073	14047	14042	7631	7469	2588
BB16.10B	40044	38168	38030	38022	21750	21398	10301
BB16.10C	26425	25137	24918	24911	13878	13711	1744
BB16.10D	44700	42317	42085	42071	23184	22827	681
BB16.11A	22890	21694	21547	21545	11541	11303	3621
BB16.11B	14560	13916	13793	13791	7555	7455	3096
BB16.11C	28188	26698	26199	26193	13934	13790	3123
BB16.11D	77503	74271	73924	73905	38036	37493	270
BB16.13A	21959	20811	20763	20760	11802	11563	803
BB16.13B	8303	7944	7919	7914	4384	4331	307
BB16.13C	296	275	265	265	83	76	1
BB16.13D	24037	22875	22788	22782	11524	11347	303
BB16.14A	43767	41159	41028	41021	20075	19778	798
BB16.14B	32138	30090	29805	29797	16597	16398	2908
BB16.14C	16789	15793	15704	15699	8826	8754	2131
BB16.14D	42578	40294	39860	39846	21616	21286	643
BB16.15A	53359	50106	49886	49875	22168	21527	5149
BB16.15B	24313	22917	22851	22846	11496	11315	2870
BB16.15C	38351	36220	36114	36103	20924	20807	11623
BB16.15D	85129	79952	78760	78746	31803	28255	260
BB16.16A	34176	32078	31956	31948	14856	14530	1680
BB16.16B	39677	37156	36957	36946	18532	18355	1660
BB16.16C	38407	35927	35757	35747	19695	19557	9471
BB16.16D	50621	47583	47124	47116	23484	23183	582
BB16.17A	100124	93809	93473	93442	46680	46196	1351
BB16.17B	53967	50811	50611	50599	26006	25747	531
BB16.17C	25989	24318	24075	24072	13217	13114	1396
BB16.17D	44954	42195	42017	42008	21040	20785	176
BB16.18A	24432	23413	23228	23223	12832	12767	3364
BB16.18B	40593	38977	38750	38742	21923	21810	4609
BB16.18C	59598	57271	57039	57026	30178	29976	3384
BB16.18D	45469	43759	43547	43536	22540	22321	138
BB16.19A	64867	62218	61616	61604	33175	32971	13911
BB16.19B	42307	40678	40426	40417	21799	21626	6611
BB16.19C	32767	31429	31217	31209	17370	17259	5206
BB16.19D	13559	12997	12906	12904	7046	6960	45
BB16.20A	25777	24559	24063	24058	11436	11362	5552
BB16.20B	35385	33691	32852	32848	16777	16701	9287
BB16.20C	44470	42260	41862	41847	23129	22934	3212
BB16.20D	48315	46336	46072	46061	23374	23093	171
BB16.21A	73753	70741	70393	70374	30567	30154	1414
BB16.21B	81528	77906	77539	77519	36417	35993	8753
BB16.21C	38477	36868	36650	36642	18959	18795	2105
BB16.21D	61997	59297	59077	59061	18576	15065	120
BB16.22A	36708	34838	34615	34611	16389	14799	2508
BB16.22B	32737	31028	30692	30688	15007	14817	3018
BB16.22C	40115	38052	37871	37858	17941	17653	1528
BB16.22D	43227	41224	41009	41002	19914	19702	147
BB16.23A	40184	38019	37865	37862	20404	20339	155
BB16.23B	27141	25840	25601	25595	13396	13307	192

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB16.23C	9823	9309	9199	9197	4873	4831	166
BB16.23D	21302	20028	19618	19613	10055	9964	119
BB16.24A	44810	42515	42230	42217	19128	18821	577
BB16.24B	48237	45757	45491	45485	20722	20489	776
BB16.24C	53072	50330	49900	49891	22778	22471	731
BB16.24D	60267	57536	57264	57255	26784	26360	292
BB16.25A	841	798	776	776	310	307	45
BB16.25B	47919	45664	45361	45348	22431	22171	3811
BB16.25C	31981	30412	30245	30236	15427	15305	1090
BB16.25D	137703	135301	134768	134768	60072	58984	597
BB16.26A	64099	61027	60760	60748	27767	27437	4423
BB16.26B	79086	75577	75254	75240	33332	32978	2915
BB16.26C	89894	85773	85409	85382	37494	36908	1981
BB16.26D	77543	74122	73764	73747	37274	36889	1147
BB16.27A	125657	119515	118050	118021	61644	60845	24968
BB16.27B	116294	111113	109968	109939	54831	54216	13954
BB16.27C	47610	45370	44881	44872	21769	21496	862
BB16.27D	126776	120658	119903	119880	56527	55670	1099
BB16.28A	44103	42139	41980	41964	19451	19219	3798
BB16.28B	80117	76062	75514	75488	33786	33446	4364
BB16.28C	41106	39172	38947	38932	17426	17031	2376
BB16.28D	101064	97170	96753	96733	44748	44035	875
BB16.29A	52473	49808	49464	49452	22868	22459	3374
BB16.29B	34666	33091	32818	32813	14150	13787	2006
BB16.29C	44044	42145	41622	41613	19953	19666	1001
BB16.29D	185505	176608	175016	174967	80840	79305	351
BB16.30A	65051	61794	61389	61379	31900	31457	5972
BB16.30B	38039	36220	35983	35973	16090	15740	1827
BB16.30C	29074	27735	27578	27572	11631	11303	909
BB16.30D	88119	84591	84162	84146	37130	36491	1223
BB16.31A	28854	27452	27362	27355	13949	13760	2394
BB16.31B	71302	68237	67960	67949	32018	31541	3773
BB16.31C	55168	52866	52738	52729	25342	25059	1851
BB16.31D	67655	64018	63207	63199	28461	27987	1216
BB16.32A	38003	35874	35763	35753	17548	17375	2594
BB16.32B	73969	69651	69180	69163	32227	31794	4735
BB16.32C	34772	32670	32514	32507	14810	14428	1136
BB16.32D	115188	108708	107912	107883	50792	49254	223
BB16.33A	76817	72337	72016	71997	34782	34209	5863
BB16.33B	93102	88068	87572	87550	40612	39748	5046
BB16.33C	14553	13841	13800	13797	7531	7460	1079
BB16.33D	127158	119964	119597	119565	53583	52416	64
BB16.34A	33417	31737	31439	31431	13953	13707	2522
BB16.34B	69890	66307	65450	65433	29979	29553	5747
BB16.34C	49262	46779	46281	46264	23162	22937	4736
BB16.34D	65538	62375	61937	61919	29238	27172	59
BB16.35A	121210	115092	114636	114597	49405	48446	13326
BB16.35B	104245	99446	98809	98786	45711	44904	12124
BB16.35C	68873	65581	65391	65373	33940	33638	10343
BB16.35D	63237	59854	59259	59234	30398	29704	77
BB16.36A	56507	53973	53454	53435	25412	24995	5855
BB16.36B	11490	10919	10760	10758	4294	4131	1419
BB16.36C	77388	73746	73259	73246	31874	31085	8336
BB16.36D	120548	115428	114935	114900	51388	50552	225
BB16.37A	59755	56878	56454	56440	27789	27404	4306
BB16.37B	70396	67295	66881	66870	31970	31411	4981
BB16.37C	30139	28835	28749	28743	14155	13984	2595
BB16.37D	92273	87668	87232	87217	41147	40532	335
BB16.38A	142187	134819	134041	134008	68533	67741	3729
BB16.38B	93764	88969	87987	87970	47178	46626	27175
BB16.38C	69349	65854	65215	65197	32124	31480	12320
BB16.38D	56410	53973	53488	53476	26619	26244	4810

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB16.39A	41358	39304	38892	38888	18150	17787	4621
BB16.39B	57765	55138	54081	54063	25886	25389	6327
BB16.39C	31201	29830	29715	29708	12861	12386	2989
BB16.39D	63285	60046	59589	59569	30199	29560	813
BB16.40A	101593	96497	93588	93569	37492	36420	11010
BB16.40B	49850	47386	46358	46350	20261	19824	6657
BB16.40C	93070	88281	85141	85122	38474	37766	12683
BB16.40D	73546	70447	70159	70145	34616	34202	195
BB16.41A	55970	53159	50074	50068	23744	23237	8394
BB16.41B	70178	66922	60398	60393	29669	29242	12477
BB16.41C	62478	59619	59052	59037	29646	29135	13580
BB16.41D	151602	143928	143298	143263	79034	77676	411
BB16.42A	53171	49924	46908	46900	21588	21181	4048
BB16.42B	47652	44914	41937	41924	19635	19360	4243
BB16.42C	39913	37521	32521	32515	14832	14547	3754
BB16.42D	66388	62914	62528	62509	31218	30806	157
BB16.43A	39379	36990	35808	35802	17239	16710	1980
BB16.43B	65831	62117	59916	59899	29609	29205	1718
BB16.43C	33530	31610	30602	30597	14742	14516	848
BB16.43D	56573	53273	52840	52829	26122	25710	67
BB16.44A	67175	63370	62955	62939	29149	28058	3425
BB16.44B	77221	72639	71367	71349	35277	34678	5212
BB16.44C	112827	106128	104945	104925	53168	52235	3685
BB16.44D	113535	107574	106849	106816	53453	52491	405
BB16.45A	148809	140139	139238	139196	70350	69302	4071
BB16.45B	109615	103200	100051	100023	51562	50797	3862
BB16.45C	175123	165052	161326	161285	93209	91329	9913
BB16.45D	119256	112180	110995	110964	53240	52212	323
BB16.46A	101988	97897	97302	97277	42319	41497	13181
BB16.46B	86297	82611	81870	81847	40516	40163	4606
BB16.46C	96466	92118	91034	91014	44823	44354	4907
BB16.46D	59879	57342	56804	56787	29886	29551	356
BB16.47A	81195	77437	76820	76800	37814	37070	3830
BB16.47B	75795	72735	71920	71907	34427	33686	4804
BB16.47C	122108	116765	114411	114387	53906	52958	4626
BB16.47D	273974	261789	259829	259785	129650	127460	669
BB16.48A	57431	54551	53000	52978	26766	26364	3972
BB16.48B	107142	102076	100561	100539	51686	50876	8268
BB16.48C	94594	89889	88140	88118	45060	44390	5975
BB16.48D	57716	54995	54499	54491	28453	28005	91
BB16.49A	59928	56861	56440	56429	30833	30189	4700
BB16.49B	135314	128939	127768	127739	60588	59035	10851
BB16.49C	22577	21458	21222	21217	11419	11241	1296
BB16.49D	112794	111341	110993	110993	57750	57066	253
BB16.50A	135321	133407	132860	132860	58628	57034	9568
BB16.50B	81154	79934	79528	79528	36606	35919	3425
BB16.50C	125135	123105	122600	122600	56431	55258	5112
BB16.50D	115405	113715	113250	113250	60395	59449	352
BB16.51A	118225	116502	116197	116197	56047	54248	3664
BB16.51B	109598	107921	107582	107582	54117	52943	2944
BB16.51C	90454	89105	88920	88920	43136	42309	2120
BB16.51D	110045	108156	107455	107455	54522	53224	185
BB17.01A	35839	33417	33337	33320	13402	13098	2378
BB17.01B	33646	32935	32850	32831	13673	13224	1951
BB17.01C	34757	33187	33069	33043	14555	14265	1986
BB17.01D	41334	39606	39485	39466	20051	19783	40
BB17.02A	33250	32610	32543	32532	14553	14250	1925
BB17.02B	46562	45607	45478	45449	19770	19398	2527
BB17.02C	34816	34121	34007	33993	15609	15362	1812
BB17.02D	36840	34515	34417	34393	17075	16004	38
BB17.03A	46171	43333	43265	43251	18872	18550	1649
BB17.03B	51801	50777	50678	50651	23647	23193	2063

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB17.03C	42354	40646	40483	40459	19106	18799	1229
BB17.03D	43476	41813	41691	41678	21945	21654	48
BB17.04A	35326	34647	34536	34525	16909	16612	3109
BB17.04B	35808	35028	34905	34890	16891	16596	2272
BB17.04C	34271	33577	33302	33279	16904	16697	2191
BB17.04D	39161	36467	36268	36255	7543	6329	4
BB17.05A	42433	39990	39907	39888	17965	17688	5777
BB17.05B	35232	34491	34368	34345	17109	16785	2473
BB17.05C	38791	37235	37018	37000	18647	18412	1910
BB17.05D	49427	47706	47536	47508	24863	24484	66
BB17.06A	38644	37776	37604	37589	19815	19559	5254
BB17.06B	42403	41597	41441	41418	21967	21658	5718
BB17.06C	35903	35259	35125	35114	18816	18613	4546
BB17.06D	37437	35317	35161	35147	19482	19070	222
BB17.07A	33458	31364	31290	31276	15506	15313	2908
BB17.07B	50430	49377	49251	49225	26069	25750	4940
BB17.07C	32234	30931	30826	30812	16707	16477	3192
BB17.07D	42852	41342	41140	41116	21065	20762	500
BB17.08A	46834	46086	46006	45993	19692	19309	3397
BB17.08B	40005	39300	38838	38819	17530	17286	3386
BB17.08C	35673	35037	34953	34937	15507	15307	3388
BB17.08D	31015	30383	30240	30217	13772	13501	823
BB17.09A	32089	30298	30238	30225	12678	12436	1496
BB17.09B	34251	33694	33620	33601	14916	14682	1551
BB17.09C	31523	30996	30894	30875	13044	12752	1404
BB17.09D	29723	29167	28996	28979	14235	14072	472
BB17.10A	57286	56227	55972	55944	24442	24015	4757
BB17.10B	25993	25534	25429	25416	11427	11251	2283
BB17.10C	27117	26665	26576	26557	11489	11282	2039
BB17.10D	39804	39010	38926	38899	17582	17232	925
BB17.11A	35685	35036	34896	34883	15979	15727	2519
BB17.11B	43510	42701	42558	42537	19469	19211	3542
BB17.11C	29396	28790	28632	28615	11619	11390	1882
BB17.11D	27930	27436	27343	27332	11987	11785	496
BB17.12A	45932	45158	45040	45021	18569	18260	3045
BB17.12B	26266	25819	25741	25731	10680	10461	1845
BB17.12C	26938	26437	26260	26242	10962	10743	2018
BB17.12D	31854	31316	31226	31205	13926	13672	974
BB17.13A	48855	48081	48028	48011	18932	18685	9942
BB17.13B	32297	31683	31627	31613	14726	14558	1508
BB17.13C	31615	31070	30965	30954	14131	13928	1708
BB17.13D	44596	43815	43688	43664	20462	20093	621
BB17.14A	41214	40524	40455	40435	17429	17130	1838
BB17.14B	65859	64549	64331	64301	34255	33924	4566
BB17.14C	76728	75528	75380	75349	33512	33095	8413
BB17.14D	36479	35772	35658	35639	16244	15935	268
BB17.15A	15855	15560	15530	15525	7420	7324	798
BB17.15B	29497	28915	28599	28583	16001	15714	3138
BB17.15C	23266	22842	22747	22740	11689	11529	2375
BB17.15D	32095	31487	31384	31369	14386	14194	803
BB17.16A	21560	21236	21163	21151	9508	9335	1218
BB17.16B	48984	48174	47957	47928	23672	23405	1138
BB17.16C	49524	48657	48529	48502	22427	22045	2640
BB17.16D	30838	30255	30115	30104	14045	13833	649
BB17.17A	13879	13645	13580	13575	5703	5623	440
BB17.17B	74960	73751	73588	73553	35829	35554	1514
BB17.17C	44139	43466	43399	43381	18744	18387	1828
BB17.17D	32963	32398	32157	32141	14343	14100	235
BB17.18A	25050	24647	24594	24579	10119	9933	706
BB17.18B	48117	47278	47160	47141	27936	27783	15695
BB17.18C	25936	25472	25430	25423	11010	10801	1151
BB17.18D	31654	31120	30959	30945	12717	12453	68

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB17.19A	40893	40136	40055	40034	16214	15807	1349
BB17.19B	59665	58556	58450	58421	36888	36711	25992
BB17.19C	36630	36028	35949	35932	16705	16497	2262
BB17.19D	22802	22365	22098	22082	10893	10825	140
BB17.20A	29581	29112	29014	29005	13053	12879	2310
BB17.20B	39457	38678	38319	38305	13255	11172	7621
BB17.20C	36917	36234	36045	36029	17739	17519	5367
BB17.20D	38646	37810	37537	37519	17182	16929	143
BB17.21A	50271	49404	49256	49237	22143	21860	5431
BB17.21B	39072	38388	38252	38233	19406	19268	5223
BB17.21C	55340	54415	54286	54257	30874	30671	17806
BB17.21D	32596	31933	31537	31524	16209	16064	70
BB17.22A	75644	74537	74432	74403	34933	34575	4232
BB17.22B	39592	37855	37753	37735	20913	20798	2483
BB17.22C	32569	32079	32040	32028	15261	15008	6850
BB17.22D	33304	32507	32235	32218	16924	16741	185
BB17.23A	77848	76515	76361	76323	33516	33060	4707
BB17.23B	47316	46388	46120	46100	21956	21763	3992
BB17.23C	25387	24908	24807	24798	11389	11240	2321
BB17.23D	32011	31197	31076	31057	15321	15155	93
BB17.24A	35374	34848	34803	34785	15118	14902	698
BB17.24B	38354	37718	37550	37527	17190	17000	1059
BB17.24C	20049	19709	19668	19655	7806	7645	1399
BB17.24D	18391	17984	17923	17916	8936	8845	41
BB17.25A	104446	102771	102542	102493	50563	50146	13228
BB17.25B	39624	38955	38789	38773	18291	18090	3308
BB17.25C	21643	21303	21258	21251	9066	8925	1462
BB17.25D	21292	20812	20751	20742	9168	8979	57
BB17.26A	53498	52480	52323	52302	25512	25321	4010
BB17.26B	41884	41131	40810	40793	19754	19546	2805
BB17.26C	28107	27520	27086	27067	12441	12275	2627
BB17.26D	35690	34936	34680	34663	18296	18106	138
BB17.27A	42599	41855	41754	41734	19950	19810	4450
BB17.27B	21276	20818	20744	20731	9816	9732	2420
BB17.27C	32867	32213	31991	31978	16698	16623	8097
BB17.27D	31170	30507	30147	30138	15442	15340	6707
BB17.28A	32911	32291	32015	31994	15626	15482	2818
BB17.28C	37239	36555	36394	36382	17773	17596	3724
BB17.28D	18995	18575	18485	18477	10005	9932	444
BB17.29A	49488	48532	48204	48176	27181	26883	5058
BB17.29B	50085	49255	49010	48986	28640	28403	4180
BB17.29C	21353	20956	20888	20878	10617	10515	2457
BB17.29D	37878	37185	36826	36816	19082	18812	1179
BB17.30A	28855	28358	28247	28228	15616	15451	3231
BB17.30B	27153	26708	26658	26643	16175	16076	3972
BB17.30C	24986	24537	24479	24465	11342	11132	909
BB17.30D	27331	26749	26560	26545	13823	13626	915
BB17.31A	8841	8472	4755	4755	1923	1883	499
BB17.31B	49131	48368	48203	48178	27562	27028	8800
BB17.31C	7315	7168	7158	7155	3537	3475	1082
BB17.31D	24077	23546	23324	23313	11935	11757	312
BB17.32A	13624	13426	13405	13400	6964	6885	1778
BB17.32B	15773	14383	14351	14343	6966	6890	1676
BB17.32C	12608	12386	12374	12367	6381	6335	1238
BB17.32D	31785	31219	31012	30995	14787	14584	363
BB17.33A	17525	17260	17216	17207	9675	9552	2736
BB17.33B	26151	25695	25545	25540	13229	13078	3475
BB17.33C	22355	21948	21866	21858	11751	11624	875
BB17.33D	16249	15929	15779	15769	7975	7879	94
BB17.34A	24034	23511	23452	23442	11248	11137	1744
BB17.34B	20412	20114	20049	20041	10345	10230	1022
BB17.34C	17748	17409	17319	17313	9179	9082	950

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB17.34D	29455	28908	28747	28729	13805	13506	143
BB17.35A	33093	32504	32435	32425	15891	15751	2658
BB17.35B	45476	44846	44773	44752	22759	22582	4629
BB17.35C	24410	23943	23881	23871	10054	9913	1696
BB17.35D	26519	25997	25854	25839	12662	12541	138
BB17.36A	19843	19205	19080	19071	7443	7362	1645
BB17.36B	47910	47204	47137	47112	21880	21684	6719
BB17.36C	20046	19692	19627	19619	9291	9153	1834
BB17.36D	22172	21753	21637	21630	10551	10412	101
BB17.37A	47937	47198	46955	46931	22368	22130	8025
BB17.37B	31806	31174	30958	30948	16199	16045	5675
BB17.37C	30513	30031	29909	29890	13583	13307	4184
BB17.37D	27391	26914	26818	26806	13564	13390	85
BB17.38A	36577	36037	35983	35973	16064	15764	3354
BB17.38B	31198	30605	30428	30417	14716	14502	3242
BB17.38C	33567	32973	32652	32635	14317	14042	2698
BB17.38D	37774	37054	36942	36921	17861	17587	134
BB17.39A	22132	21765	21717	21707	11162	11041	2580
BB17.39B	31561	27025	26870	26864	14589	14441	3509
BB17.39C	14616	14332	14243	14236	6941	6853	1090
BB17.39D	36811	36098	35881	35859	17305	17079	184
BB17.40A	21544	21210	21058	21041	11848	11743	2148
BB17.40B	32101	31524	31357	31345	16501	16349	2416
BB17.40C	38934	38238	38026	38010	18515	18257	2215
BB17.40D	26842	26267	25975	25961	12661	12486	306
BB17.41A	23802	23352	22942	22929	12291	12183	1823
BB17.41B	40104	39395	39170	39149	20401	20213	2096
BB17.41C	29928	29429	29385	29369	13265	13018	1377
BB17.41D	32359	31780	31637	31622	14313	14090	228
BB17.42A	36282	35469	35226	35209	21274	21024	5485
BB17.42B	15339	14973	14835	14833	7573	7491	2019
BB17.42C	40422	39689	39524	39505	18344	18024	4222
BB17.42D	40450	39690	39438	39415	18351	18064	268
BB17.43A	52086	51153	50723	50693	26663	26424	11063
BB17.43B	28756	28244	28171	28162	14042	13876	5553
BB17.43C	35242	34551	34498	34480	16582	16316	5447
BB17.43D	32637	31989	31787	31770	16259	16085	200
BB17.44A	32969	32411	32312	32297	15839	15589	3448
BB17.44B	19744	19343	19290	19278	8661	8532	1888
BB17.44C	21284	20881	20775	20766	8682	8495	1947
BB17.44D	35969	35439	35360	35340	16413	16102	164
BB17.45A	54830	53938	53640	53619	27032	26755	5893
BB17.45B	22102	21766	21738	21726	10232	10060	2143
BB17.45C	31886	31334	31265	31255	14203	13993	3945
BB17.45D	28417	27951	27852	27835	13584	13374	151
BB17.46A	38194	37575	37365	37347	17968	17735	1500
BB17.46B	37154	36553	36450	36431	18230	18042	1648
BB17.46C	39461	38749	38633	38610	17292	16994	1649
BB17.46D	33785	33120	32942	32928	16386	16113	183
BB17.47A	24941	24522	24312	24300	10188	9992	1412
BB17.47B	27344	25277	24969	24958	12030	11887	1604
BB17.47C	19813	19416	19372	19366	8433	8280	889
BB17.47D	35619	34854	34726	34712	16770	16462	149
BB17.48A	41702	41004	40818	40797	17588	17323	2216
BB17.48B	47139	44873	44475	44451	20624	20332	2689
BB17.48C	31658	31101	30974	30957	13651	13367	1258
BB17.48D	46692	45759	45356	45330	22958	22525	61
BB17.49A	83283	81727	80739	80694	42880	42230	6569
BB17.49B	35478	34819	33851	33834	15922	15687	2971
BB17.49C	36516	35888	35577	35560	16631	16352	3583
BB17.49D	43350	42487	42176	42160	21277	21034	162
BB17.50A	29340	28764	28656	28637	13511	13186	1449

Pipeline Steps >>>>>						FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur		
BB17.50B	12280	12087	12073	12069	6278	6161	566
BB17.50C	47044	46231	45938	45912	20404	20040	866
BB17.50D	34714	34117	33984	33968	17562	17277	48
BB17.51A	35876	27408	27278	27268	13727	13475	1745
BB17.51B	34006	33446	33232	33217	14453	14137	1992
BB17.51C	30259	29647	29534	29517	12886	12620	633
BB17.51D	19848	19474	19440	19428	8408	8152	97

Supplemental Data A2. Number of reads filtered during the sequence processing pipeline of Chapter 2; data for Bedford Basin V6-V8 *16S* rRNA are shown. Values highlighted red lacked chloroplast and cyanobacterial reads after data processing.

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB14.03A	82798	82619	82235	82168	40180	37494	202
BB14.03B	108497	108278	107842	107764	60684	58686	738
BB14.03C	110650	110377	109865	109776	58916	57267	423
BB14.03D	48069	47933	47701	47660	25693	24914	60
BB14.04A	149094	148779	148235	148120	82877	79627	1980
BB14.04B	54499	54307	53983	53936	29530	27510	924
BB14.04C	162604	162242	161582	161445	86633	83265	1803
BB14.04D	108912	108662	108247	108152	59014	56165	302
BB14.05A	106462	106208	105771	105677	59258	56953	1655
BB14.05B	83555	83360	82991	82922	45347	43366	1335
BB14.05C	115995	115767	115354	115249	59917	58165	2255
BB14.05D	129628	129336	128809	128726	66120	63914	424
BB14.06A	119912	119663	119193	119122	64998	63905	3807
BB14.06B	110306	110083	109589	109516	59813	58618	2614
BB14.06C	109562	109277	108766	108693	56140	55180	1325
BB14.06D	93488	93282	92913	92856	47457	46069	262
BB14.07A	123685	123403	122937	122850	67162	65570	1273
BB14.07B	65406	65206	64829	64784	35786	34963	759
BB14.07C	60045	59858	59533	59494	33082	32433	438
BB14.07D	113887	113621	113200	113110	60775	59147	127
BB14.08A	95617	95392	95014	94942	41555	39090	1288
BB14.08B	35450	35366	35186	35165	15965	15054	336
BB14.08C	115233	115035	114661	114577	53161	51167	837
BB14.08D	103450	103256	102899	102830	48419	46741	89
BB14.09A	73678	73523	73242	73192	32748	31149	1919
BB14.09B	82696	82505	82152	82091	40716	39751	923
BB14.09C	80888	80715	80360	80310	36671	35457	723
BB14.09D	68567	68393	68106	68069	31223	29361	85
BB14.10A	111302	111046	110629	110545	51630	49942	5228
BB14.10B	130350	130030	129481	129386	62542	60946	4179
BB14.10C	144069	143768	143179	143105	69310	67442	1334
BB14.10D	89431	89226	88861	88794	40775	38765	144
BB14.11A	77276	77144	76913	76843	42443	41431	2297
BB14.11B	57194	57055	56812	56779	33771	33372	762
BB14.11C	69068	68915	68685	68644	39138	38433	563
BB14.11D	78905	78739	78359	78313	41560	40862	677
BB14.12A	784	769	739	739	266	245	3
BB14.12Ab	13673	13644	13449	13448	6020	5502	35
BB14.12B	1731	1725	1683	1681	861	820	2
BB14.12Bb	19643	19605	19288	19286	7845	7394	71
BB14.12C	104314	104086	103729	103642	53311	53205	15
BB14.12D	107483	107254	106868	106781	52164	51970	5
BB14.13A	87397	87209	86920	86865	51118	50750	87
BB14.13B	81801	81593	81324	81266	47142	46731	163
BB14.13C	49413	49283	49093	49064	29394	29141	76
BB14.13D	90793	90596	90254	90183	52176	51095	147
BB14.14A	61330	61093	60469	60462	24926	24553	273
BB14.14B	59575	59316	58746	58743	24411	23903	426
BB14.14C	46602	46428	45979	45970	20803	20473	674
BB14.14D	35629	35489	35095	35089	15416	15263	55
BB14.15A	35671	35572	35208	35203	15510	15140	354
BB14.15B	28885	28779	28494	28491	11953	11747	579
BB14.15C	1695	1671	1540	1540	391	379	9
BB14.15Cb	130924	130648	128813	128802	55433	54248	1800
BB14.15D	16906	16815	16586	16585	6641	6588	8

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB14.16A	56081	55896	55464	55459	15105	14677	3
BB14.16B	56845	56651	56219	56215	15857	15506	0
BB14.16C	67595	67399	66818	66811	16939	16398	0
BB14.16D	51276	51076	50547	50541	12326	11590	3
BB14.17A	71516	71284	70394	70387	19549	18988	1665
BB14.17B	58973	58723	57805	57799	14956	14601	4664
BB14.17C	68166	67891	67052	67049	17100	16701	3605
BB14.17D	42419	42268	41654	41643	12968	12697	56
BB14.18A	63934	63745	63051	63042	18765	18243	418
BB14.18B	62713	62440	61813	61805	17431	17016	179
BB14.18C	72175	71910	70930	70921	21868	21432	302
BB14.18D	55628	55364	54634	54629	14505	14097	39
BB14.19A	70730	70409	69741	69733	20774	20320	354
BB14.19B	56442	56073	55314	55312	13947	13674	797
BB14.19C	69425	69219	68453	68447	20959	20444	619
BB14.19D	53359	53134	52568	52561	13900	13339	61
BB14.20A	75762	75527	74959	74950	21787	21239	2118
BB14.20B	53941	53689	53280	53273	15381	14895	1888
BB14.20C	59775	59576	59205	59196	17575	17111	1683
BB14.20D	30776	30538	30051	30047	6129	6036	3
BB14.21A	67867	67614	67078	67069	16750	16207	57
BB14.21B	50143	49934	49580	49568	13306	12897	16
BB14.21C	73382	73125	72570	72564	21600	21070	273
BB14.21D	51584	51360	50880	50874	13875	13390	4
BB14.22A	53955	53753	53375	53368	14123	13528	123
BB14.22B	58309	58097	57749	57737	15757	15193	146
BB14.22C	66244	66032	65617	65605	17598	16884	200
BB14.22D	48812	48620	48205	48203	11967	11431	11
BB14.23A	63716	63479	62827	62819	16756	16221	103
BB14.23B	60233	59953	59433	59425	16878	16416	63
BB14.23C	80987	80643	79956	79942	21917	21390	256
BB14.23D	46799	46610	46110	46108	14196	13963	34
BB14.24A	84199	83933	83311	83300	24001	23289	316
BB14.24B	42489	42297	41894	41886	11029	10828	33
BB14.24C	74572	74297	73504	73496	19515	18830	83
BB14.24D	56047	55772	55135	55128	14142	13592	28
BB14.25A	56052	55820	55353	55346	14290	13739	41
BB14.25B	60556	60343	59788	59776	16821	16265	190
BB14.25C	48418	48223	47760	47752	12313	11820	68
BB14.25D	49441	49214	48606	48604	13267	12826	62
BB14.26A	53312	53092	52611	52605	13078	12556	254
BB14.26B	43110	42879	42540	42531	10314	9826	27
BB14.26C	63924	63606	63094	63084	15873	15197	77
BB14.26D	9504	8912	8433	8433	431	413	2
BB14.26Db	92561	92370	90927	90915	28950	26369	183
BB14.27A	65249	64995	64480	64472	17415	17039	562
BB14.27B	49588	49356	48970	48964	13239	12925	173
BB14.27C	67563	67255	66680	66671	17522	16965	190
BB14.27D	54825	54540	54062	54056	13377	12770	42
BB14.28A	24328	24222	23968	23966	7659	7540	322
BB14.28B	62008	61766	61269	61263	13964	13180	85
BB14.28C	78250	77930	77337	77325	20280	19605	299
BB14.28D	39087	38913	38580	38578	9746	9410	0
BB14.29A	27192	27060	26686	26684	9500	9428	361
BB14.29B	33438	33264	32872	32869	13471	13344	436
BB14.29C	9811	9750	9615	9613	3443	3413	753
BB14.29Cb	26951	26868	26467	26464	9611	9191	2148
BB14.29D	35920	35808	35436	35435	10222	10006	38
BB14.30A	65938	65723	65157	65148	19902	19391	100
BB14.30B	33172	33027	32680	32679	10357	10283	41
BB14.30C	66110	65860	65153	65144	18477	18048	37
BB14.30D	57021	56695	56044	56043	13838	13354	7

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB14.31A	52230	52034	51537	51534	16793	16561	465
BB14.31B	43751	43594	43187	43183	12138	11835	78
BB14.31C	20070	19977	19779	19777	6611	6555	174
BB14.31D	42345	42191	41790	41789	10400	10025	8
BB14.32A	43717	43561	43038	43034	11845	11506	215
BB14.32B	31946	31800	31449	31443	9496	9317	77
BB14.32C	12978	12903	12749	12747	4004	3953	168
BB14.32Cb	30369	30302	29864	29862	10394	9632	417
BB14.32D	44412	44243	43848	43840	12727	12411	29
BB14.33A	61073	60880	60290	60283	17308	16941	176
BB14.33B	51672	51457	50911	50908	15236	14939	130
BB14.33C	65486	65243	64511	64505	17605	17104	193
BB14.33D	59999	59709	59143	59133	16602	16112	41
BB14.34A	49335	49161	48704	48696	13679	13325	470
BB14.34B	49545	49372	48820	48817	13255	12940	557
BB14.34C	46174	46021	45550	45547	12023	11692	205
BB14.34D	50800	50595	50077	50070	13682	13273	21
BB14.35A	64857	64624	63886	63877	18067	17678	1717
BB14.35B	43571	43372	42860	42854	13966	13784	1666
BB14.35C	74162	73845	73007	72997	18643	18154	927
BB14.35D	39062	38924	38494	38489	11150	10924	8
BB14.36A	38617	38473	37964	37964	11408	11259	1229
BB14.36B	32107	31965	31591	31584	9734	9612	710
BB14.36C	65488	65264	64429	64424	17244	16799	485
BB14.36D	48736	48477	47878	47872	12486	12168	10
BB14.37A	30516	30409	30063	30062	8699	8558	255
BB14.37B	27271	27117	26669	26668	6460	6391	230
BB14.37C	54097	53917	53350	53346	13023	12534	580
BB14.37D	71030	70776	69919	69909	18630	17964	5
BB14.38A	25083	25002	24744	24741	7743	7540	445
BB14.38B	31115	31015	30612	30610	9012	8783	396
BB14.38C	29159	29059	28690	28684	8568	8361	498
BB14.38D	37664	37557	37037	37026	10091	9731	4
BB14.39A	25155	25061	24725	24723	8253	8027	183
BB14.39B	44019	43869	43354	43349	13320	12870	525
BB14.39C	50193	50019	49436	49427	12946	12393	1762
BB14.39D	48927	48716	48022	48018	13145	12793	43
BB14.40A	26578	26478	26121	26118	10108	9937	171
BB14.40B	54089	53897	53173	53171	16860	16244	135
BB14.40D	32167	32058	31552	31550	8036	7788	19
BB14.41A	49838	49679	49131	49129	14505	14000	54
BB14.41B	39089	38978	38457	38454	11997	11634	60
BB14.41C	47011	46821	46327	46319	13936	13479	117
BB14.41D	44743	44603	43994	43987	12441	11948	24
BB14.42A	43432	43250	42756	42753	12382	11956	32
BB14.42B	51188	50974	50303	50298	14795	14367	50
BB14.42C	48195	48030	47393	47393	14282	13843	110
BB14.42D	53060	52903	52375	52367	14142	13459	23
BB14.43A	27062	26967	26740	26734	8212	7998	176
BB14.43B	21951	21887	21694	21691	6429	6202	122
BB14.43C	20619	20540	20360	20359	5656	5397	14
BB14.43D	39796	39651	39253	39248	9574	9111	0
BB14.44A	23436	23339	23128	23125	6431	6131	151
BB14.44B	31348	31252	31002	30996	8940	8588	191
BB14.44C	38610	38427	38085	38075	10353	9938	110
BB14.44D	42203	41987	41555	41550	10226	9767	4
BB14.45A	38480	38341	37968	37963	11805	11320	885
BB14.45B	48101	47913	47372	47359	14952	14473	1112
BB14.45C	51111	50933	50419	50415	14664	14244	709
BB14.45D	46964	46763	46209	46203	11926	11527	19
BB14.46A	54074	53881	53364	53359	15780	15267	263
BB14.46B	29798	29684	29274	29272	6992	6594	348

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB14.46C	31757	31623	31235	31230	7115	6728	344
BB14.46D	45715	45553	45048	45043	12179	11710	38
BB14.47A	40957	40762	40282	40281	10930	10390	341
BB14.47B	53016	52778	52074	52063	13943	13353	355
BB14.47C	57643	57467	56930	56919	14045	13283	396
BB14.47D	47182	47036	46622	46611	11932	11344	21
BB14.48A	41719	41516	41124	41116	11406	10987	428
BB14.48B	48475	48296	47886	47876	11590	10827	257
BB14.48C	38305	38151	37701	37696	8566	8149	108
BB14.48D	34232	34068	33731	33731	8069	7712	9
BB14.49A	36701	36562	36221	36217	9989	9587	453
BB14.49B	41103	40914	40538	40532	10294	9812	483
BB14.49C	44900	44634	44070	44064	10049	9650	309
BB14.49D	65206	64990	64206	64197	16933	16248	39
BB14.50A	42276	42136	41719	41712	10878	10487	338
BB14.50B	28275	28181	27842	27838	7182	6912	147
BB14.50C	59187	59008	58357	58349	15670	15143	232
BB14.50D	37124	37022	36555	36552	9929	9612	12
BB14.51A	23076	23007	22749	22748	6407	6227	56
BB14.51B	42162	41999	41571	41569	13958	13663	140
BB14.51C	47328	47133	46620	46616	14223	13892	75
BB14.51D	49267	49075	48412	48409	13285	12832	46
BB15.02A	71688	71437	70721	70716	21164	20495	148
BB15.02B	34746	34622	34246	34246	9325	9135	66
BB15.02C	39351	39189	38687	38687	9672	9410	110
BB15.02D	63314	63087	62458	62451	17411	16856	13
BB15.03A	44596	44445	43947	43942	11656	11214	333
BB15.03B	44046	43868	43414	43405	11661	11378	313
BB15.03C	22945	22844	22610	22604	6692	6517	12
BB15.03D	54167	53970	53414	53407	15795	15335	374
BB15.04A	52879	52664	52066	52062	14755	14342	372
BB15.04B	73665	73428	72621	72607	20558	19881	562
BB15.04C	31470	31139	30358	30358	4951	4893	154
BB15.04D	55578	55374	54722	54716	14494	14119	76
BB15.05A	62570	62387	61585	61578	17329	17052	698
BB15.05B	54492	54309	53728	53725	18595	18288	273
BB15.05C	58749	58499	57878	57868	16122	15676	445
BB15.05D	48149	47991	47432	47425	12503	12091	99
BB15.06A	61177	60931	60214	60206	19074	18669	409
BB15.06B	43692	43571	43057	43051	12354	11983	360
BB15.06C	72125	71908	71294	71284	19403	18701	626
BB15.06D	41720	41547	41159	41158	9333	8899	29
BB15.07A	62505	62256	61404	61397	17670	16856	114
BB15.07B	75434	75184	74320	74316	22621	22044	141
BB15.07C	95195	94839	93874	93861	28695	27923	363
BB15.07D	54587	54432	53809	53803	14243	13730	22
BB15.08A	52741	52531	51902	51898	14262	13856	1834
BB15.08B	43781	43583	42890	42889	10097	9861	1351
BB15.08C	54763	54581	53994	53988	14169	13621	1694
BB15.08D	48962	48753	48248	48239	11471	10859	58
BB15.09A	52810	52629	52080	52075	14168	13669	1364
BB15.09B	32391	32232	31828	31827	6723	6417	789
BB15.09C	43495	43361	42841	42837	11942	11655	1358
BB15.09D	45793	45623	45070	45065	12031	11661	94
BB15.10A	40517	40382	39958	39954	10608	10271	1204
BB15.10B	38721	38591	38175	38167	10811	10536	1962
BB15.10C	42841	42682	42217	42214	12072	11719	2379
BB15.10D	46900	46747	46259	46252	12221	11798	359
BB15.11A	49508	49341	48824	48823	14018	13758	1746
BB15.11B	52322	52132	51546	51543	14562	14264	1854
BB15.11C	53471	53199	52491	52488	12843	12596	2067
BB15.11D	72842	72595	71685	71681	19320	18906	614

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB15.12A	54842	54643	54091	54083	15850	15559	902
BB15.12B	49739	49561	49061	49055	12141	11717	736
BB15.12C	63763	63538	62941	62933	16772	16359	742
BB15.12D	75986	75730	74952	74944	19603	18990	307
BB15.13A	51526	51312	50703	50701	12171	11919	391
BB15.13B	42375	42209	41795	41792	12712	12536	1225
BB15.13C	26254	26148	25867	25860	7907	7766	1126
BB15.13D	46152	45958	45413	45406	11639	11320	359
BB15.14A	14355	14310	14301	14300	6912	6813	4
BB15.14B	312	310	310	310	207	207	0
BB15.14C	11626	11594	11584	11582	6118	6076	0
BB15.14D	25811	25741	25725	25722	11110	10901	43
BB15.15A	79	79	79	79	40	40	0
BB15.15B	26	25	25	25	0	0	0
BB15.15C	80	78	78	78	37	37	0
BB15.15D	28529	28456	28441	28437	13191	12932	114
BB15.16A	13	13	12	12	0	0	0
BB15.16B	15666	15651	15643	15641	7643	7518	0
BB15.16C	19177	19130	19115	19113	9820	9656	24
BB15.16D	26199	26140	26121	26118	11689	11480	16
BB15.17A	17	17	16	16	0	0	0
BB15.17B	8637	8614	8610	8608	4031	3964	0
BB15.17C	19981	19946	19937	19934	9007	8828	34
BB15.17D	22779	22734	22725	22724	10682	10466	113
BB15.18A	12035	12004	11999	11997	5285	5231	3
BB15.18B	13117	13086	13083	13082	5789	5690	0
BB15.18C	28045	27984	27971	27969	13149	12921	5
BB15.18D	23765	23677	23663	23662	9982	9760	21
BB15.19A	16244	16206	16201	16197	6983	6855	9
BB15.19B	17388	17346	17334	17327	6989	6841	48
BB15.19C	32256	32197	32181	32173	14060	13834	28
BB15.19D	32069	31957	31939	31934	16704	16560	209
BB15.20A	35302	35249	35227	35223	15228	14986	127
BB15.20B	35334	35274	35258	35256	14953	14660	22
BB15.20C	34023	33974	33954	33947	17798	17661	11
BB15.20D	40356	40276	40246	40238	18986	18686	0
BB15.21A	40583	40500	40477	40473	18775	18482	3222
BB15.21B	35978	35903	35884	35879	16861	16626	2460
BB15.21C	40277	40214	40186	40180	18520	18279	636
BB15.21D	41448	41361	41335	41327	22795	22530	78
BB15.22A	34914	34840	34822	34813	14888	14407	16
BB15.22B	29414	29372	29362	29349	13533	13203	33
BB15.22C	41406	41321	41284	41278	18733	18338	712
BB15.22D	31011	30952	30930	30926	15816	15556	39
BB15.23A	35573	35508	35496	35485	17466	17031	46
BB15.23B	48607	48506	48485	48479	21580	20960	57
BB15.23C	47389	47313	47293	47285	19731	19149	133
BB15.23D	50298	50198	50166	50153	24596	24201	26
BB15.24A	48611	48441	48414	48406	20445	20139	0
BB15.24B	30103	30027	30016	30013	12644	12412	14
BB15.24C	35304	35215	35196	35190	15849	15594	53
BB15.24D	28508	28433	28419	28415	13188	13019	66
BB15.25A	42053	41951	41930	41925	19238	18867	136
BB15.25B	38991	38883	38867	38860	16627	16222	15
BB15.25C	54339	54221	54185	54170	24062	23632	4
BB15.25D	45167	45056	45029	45026	25834	25520	53
BB15.26A	26798	26751	26737	26734	11432	11166	209
BB15.26B	10530	10511	10504	10501	6091	6044	86
BB15.26C	3157	3150	3146	3145	1937	1928	2
BB15.26D	24381	24324	24311	24305	10235	10007	16
BB15.27A	19523	19484	19471	19469	9528	9384	45
BB15.27B	47	45	44	44	5	5	0

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB15.27C	3643	3632	3625	3625	2297	2290	0
BB15.27D	42902	42836	42804	42797	20155	19770	30
BB15.28A	40341	40240	40219	40213	20962	20656	92
BB15.28B	30654	30587	30568	30562	14873	14546	26
BB15.28C	20917	20868	20850	20847	9992	9819	27
BB15.28D	27036	26957	26940	26938	11384	11051	7
BB15.29A	30890	30807	30781	30776	17784	17571	162
BB15.29B	31705	31634	31616	31612	14303	13890	80
BB15.29C	36950	36852	36837	36831	16785	16492	62
BB15.29D	40063	39966	39941	39932	18664	18292	14
BB15.30A	14589	14556	14543	14538	7145	7041	16
BB15.30B	16630	16594	16590	16584	7501	7348	11
BB15.30C	31997	31915	31896	31891	15520	15214	31
BB15.30D	22222	22157	22136	22134	10345	10183	4
BB15.31A	20408	20368	20360	20359	9678	9497	209
BB15.31B	22771	22709	22699	22692	10370	10081	131
BB15.31C	36798	36708	36688	36680	17952	17608	116
BB15.31D	38790	38690	38664	38657	18516	18213	4
BB15.32A	3538	3525	3522	3522	2405	2396	14
BB15.32B	18017	17987	17972	17971	8769	8544	187
BB15.32C	30204	30138	30125	30119	14362	14065	94
BB15.32D	31141	31048	31034	31029	13035	12653	0
BB15.33A	31351	31273	31256	31252	16061	15721	384
BB15.33B	37288	37192	37179	37172	17407	16899	398
BB15.33C	46438	46328	46308	46304	18024	17338	121
BB15.33D	50483	50363	50340	50335	25067	24521	13
BB15.34A	31192	31134	31112	31101	13079	12660	238
BB15.34B	23642	23604	23592	23588	10116	9800	78
BB15.34C	11380	11355	11343	11340	4767	4599	62
BB15.34D	20672	20632	20617	20612	8676	8450	13
BB15.35A	23979	23929	23914	23911	10803	10464	267
BB15.35B	25275	25219	25206	25205	11072	10655	144
BB15.35C	20758	20701	20679	20677	7990	7721	259
BB15.35D	37668	37599	37581	37578	16243	15858	33
BB15.36A	22709	22643	22635	22634	11125	10864	359
BB15.36B	31268	31187	31161	31157	14366	13958	693
BB15.36C	19424	19394	19384	19382	8305	7991	249
BB15.36D	31564	31484	31468	31463	13090	12694	4
BB15.37A	19244	19199	19188	19186	9304	9056	457
BB15.37B	21667	21615	21602	21598	9168	8803	322
BB15.37C	37301	37198	37180	37175	15374	14791	343
BB15.37D	32351	32252	32232	32224	12830	12309	7
BB15.38A	32190	32103	32076	32072	18761	18555	549
BB15.38B	30850	30764	30736	30727	19503	19276	754
BB15.38C	31561	31465	31437	31430	15374	15008	620
BB15.38D	40904	40804	40767	40764	17265	16829	14
BB15.39A	36772	36667	36647	36641	18642	18176	224
BB15.39B	25820	25766	25751	25749	11352	10983	105
BB15.39C	31651	31590	31578	31570	12767	12242	117
BB15.39D	39386	39318	39285	39280	16463	16023	18
BB15.40A	20302	20261	20253	20250	8144	7776	77
BB15.40B	18273	18252	18237	18235	7386	7042	57
BB15.40C	9263	9188	9175	9175	3786	3683	23
BB15.40D	14003	13879	13858	13858	5984	5817	8
BB15.41A	9149	9068	9057	9057	3538	3443	30
BB15.41B	8604	8532	8508	8508	3393	3332	41
BB15.41C	13175	13063	13046	13046	5396	5286	44
BB15.41D	10817	10723	10712	10712	4517	4462	2
BB15.42A	3332	3292	3284	3284	1366	1323	0
BB15.42B	6580	6514	6502	6502	2458	2388	8
BB15.42C	7265	7209	7195	7195	2713	2647	3
BB15.42D	17457	17309	17296	17296	7432	7283	2

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB15.43A	10135	10050	10041	10041	3908	3840	75
BB15.43B	8426	8356	8344	8344	3341	3278	54
BB15.43C	10876	10755	10732	10732	4422	4349	74
BB15.43D	9177	9096	9079	9079	3884	3809	0
BB15.44A	9668	9569	9552	9552	3991	3898	238
BB15.44B	8412	8337	8325	8325	4009	3947	128
BB15.44C	8258	8178	8164	8164	3045	2994	91
BB15.44D	24	23	21	21	0	0	0
BB15.45A	30400	30330	30309	30300	13233	12608	412
BB15.45B	25839	25783	25761	25753	10952	10471	344
BB15.45C	24590	24529	24520	24516	10646	10069	785
BB15.45D	33260	33197	33143	33141	14144	13587	5
BB15.46A	13012	12867	12857	12857	6012	5942	624
BB15.46B	8814	8692	8677	8677	3440	3337	972
BB15.46C	9667	9508	9495	9495	3639	3539	1075
BB15.46D	9213	9103	9074	9074	3935	3823	0
BB15.47A	9743	9630	9621	9621	3986	3882	1080
BB15.47B	7178	7091	7088	7088	3543	3505	793
BB15.47C	11263	11163	11144	11144	4635	4540	1379
BB15.47D	7176	7103	7061	7061	3262	3211	0
BB15.48A	24	23	21	21	0	0	0
BB15.48B	9910	9823	9804	9804	3731	3647	163
BB15.48C	11801	11716	11706	11706	4897	4784	841
BB15.48D	12779	12659	12638	12638	4801	4645	13
BB15.49A	12211	12083	12069	12069	4774	4648	129
BB15.49B	7436	7364	7357	7357	3496	3439	63
BB15.49C	9534	9474	9457	9457	3691	3623	180
BB15.49D	4335	4300	4255	4255	2052	2029	0
BB15.50A	14201	14076	14065	14065	6227	6160	474
BB15.50B	17725	17556	17521	17521	6608	6425	538
BB15.50C	18756	18584	18566	18566	7321	7142	936
BB15.50D	14225	14095	14072	14072	5755	5619	11
BB15.51A	12522	12364	12349	12349	5260	5107	749
BB15.51B	12582	12474	12460	12460	6403	6320	636
BB15.51C	13996	13889	13863	13863	6011	5919	918
BB15.51D	13429	13295	13276	13276	4454	4270	0
BB16.01A	29380	29297	29161	29161	11051	10943	3738
BB16.01B	30674	30489	30397	30397	11906	11546	2710
BB16.01C	31450	31263	31145	31145	13258	12977	4485
BB16.01D	32631	32403	32325	32325	13663	13280	41
BB16.02A	33202	33107	32920	32919	10917	10775	792
BB16.02B	18555	18432	18374	18374	6929	6772	1078
BB16.02C	25989	25820	25737	25737	10281	9949	1689
BB16.02D	31689	31505	31418	31417	12705	12222	18
BB16.03A	36188	36064	35884	35884	13997	13876	5994
BB16.03B	28372	28196	28100	28099	11849	11567	3689
BB16.03C	29276	29099	28999	28996	12534	12306	3664
BB16.03D	28295	28076	27997	27997	9896	9570	88
BB16.04A	37793	37678	37478	37478	15184	15085	1602
BB16.04B	31355	31154	31053	31053	11914	11671	1363
BB16.04C	39655	39361	39272	39272	15830	15505	1806
BB16.04D	26590	26326	26270	26269	9121	8828	122
BB16.05A	31911	31812	31658	31658	10697	10582	799
BB16.05B	31268	31049	30953	30952	11651	11336	695
BB16.05C	41967	41668	41528	41528	15894	15475	788
BB16.05D	35356	35087	34992	34992	12961	12572	38
BB16.06A	27306	27222	27068	27068	9182	9106	320
BB16.06B	25301	25109	25014	25014	8694	8410	213
BB16.06C	28894	28673	28598	28597	10998	10640	281
BB16.06D	26121	25915	25854	25854	10050	9816	24
BB16.07A	26821	26742	26606	26606	10436	10361	378
BB16.07B	30357	30178	30110	30108	11970	11775	368

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB16.07C	30637	30392	30326	30326	10923	10683	256
BB16.07D	32490	32270	32199	32199	12853	12591	21
BB16.08A	26123	26012	25873	25873	8988	8924	283
BB16.08B	29759	29537	29445	29445	11536	11243	415
BB16.08C	29137	28916	28843	28843	11204	10977	245
BB16.08D	28725	28491	28418	28418	11228	11000	32
BB16.09A	27305	27227	27111	27111	10114	10030	620
BB16.09B	25685	25488	25422	25422	8725	8496	514
BB16.09C	31792	31575	31509	31509	11209	10894	369
BB16.09D	34543	34225	34159	34158	14010	13725	14
BB16.10A	27263	27205	27061	27061	9742	9635	1852
BB16.10B	16672	16526	16467	16467	5695	5577	939
BB16.10C	33005	32762	32695	32695	12464	12154	576
BB16.10D	32965	32680	32606	32605	12174	11795	131
BB16.11A	29601	29530	29370	29370	11536	11390	1484
BB16.11B	16787	16662	16612	16611	5932	5763	879
BB16.11C	28154	27862	27789	27789	13121	12950	1819
BB16.11D	23759	23603	23543	23543	8558	8338	28
BB16.13A	29486	29394	29244	29244	14267	14101	405
BB16.13B	25307	25157	25074	25074	9075	8841	162
BB16.13C	21977	21853	21801	21801	8164	7962	90
BB16.13D	22459	22298	22241	22241	8371	8183	127
BB16.14A	34461	34339	34169	34169	12341	12244	124
BB16.14B	29009	28820	28761	28760	11264	11066	613
BB16.14C	22616	22466	22394	22393	8667	8494	588
BB16.14D	32153	31962	31882	31882	13491	13302	150
BB16.15A	29689	29605	29462	29462	9005	8875	991
BB16.15B	28812	28593	28512	28512	10018	9799	1007
BB16.15C	17276	17149	17097	17097	5460	5273	1429
BB16.15D	29919	29714	29631	29630	10062	9775	116
BB16.16A	34568	34449	34297	34297	11825	11712	350
BB16.16B	35448	35223	35146	35146	14166	13932	343
BB16.16C	27338	27140	27045	27045	9624	9464	1875
BB16.16D	29734	29535	29472	29472	10715	10400	103
BB16.17A	31521	31422	31241	31241	11266	11181	72
BB16.17B	32034	31819	31719	31718	11975	11685	63
BB16.17C	33621	33409	33331	33331	12678	12346	288
BB16.17D	31593	31430	31321	31320	11217	10937	31
BB16.18A	33396	33318	33150	33150	13597	13514	1085
BB16.18B	37304	37090	36969	36968	17071	16882	1113
BB16.18C	37425	37217	37142	37142	14228	13970	332
BB16.18D	30890	30656	30565	30565	10931	10631	17
BB16.19A	38620	38504	38275	38274	14159	14058	3099
BB16.19B	43225	42968	42851	42849	17644	17293	2306
BB16.19C	37068	36845	36716	36715	15082	14909	1416
BB16.19D	34739	34533	34455	34455	13765	13511	31
BB16.20A	34399	34273	34015	34015	12655	12566	3211
BB16.20B	29626	29459	29322	29322	10953	10792	2514
BB16.20C	33065	32843	32738	32738	13371	13183	613
BB16.20D	28278	28088	28025	28025	10339	10028	14
BB16.21A	50120	49934	49699	49699	17323	17136	252
BB16.21B	31435	31217	31140	31140	11589	11323	795
BB16.21C	38933	38626	38519	38519	14735	14373	395
BB16.21D	42605	42277	42184	42184	13694	12913	64
BB16.22A	41115	41016	40805	40805	14101	13957	852
BB16.22B	38695	38440	38315	38315	13423	13065	630
BB16.22C	45188	44908	44835	44834	17452	17142	401
BB16.22D	37721	37451	37368	37368	12724	12409	15
BB16.23A	33783	33710	33590	33589	12067	11985	29
BB16.23B	28963	28810	28754	28754	10665	10449	50
BB16.23C	29993	29768	29690	29690	12069	11873	136
BB16.23D	24422	24233	24189	24188	8889	8676	30

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB16.24A	34519	34426	34259	34259	11385	11265	114
BB16.24B	31937	31695	31612	31612	10550	10220	78
BB16.24C	41124	40783	40710	40710	14130	13693	144
BB16.24D	33915	33671	33599	33599	11942	11611	67
BB16.25A	31608	31518	31329	31328	10466	10354	868
BB16.25B	34080	33837	33776	33775	12117	11774	717
BB16.25C	38801	38522	38449	38449	13255	12952	303
BB16.25D	34095	33833	33731	33728	11783	11435	66
BB16.26A	37941	37834	37650	37650	14623	14445	433
BB16.26B	28192	27884	27781	27781	10752	10568	190
BB16.26C	34575	34305	34258	34257	12632	12329	147
BB16.26D	33878	33559	33492	33492	13125	12780	156
BB16.27A	28866	28799	28655	28655	11481	11341	2193
BB16.27B	38423	38089	38029	38029	15450	15057	1129
BB16.27C	38042	37772	37695	37695	12852	12450	151
BB16.27D	37613	37337	37266	37266	13535	13225	98
BB16.28A	32300	32232	32079	32079	11656	11538	792
BB16.28B	29020	28741	28664	28664	9462	9290	271
BB16.28C	27446	27278	27198	27198	10233	9913	447
BB16.28D	35048	34854	34771	34771	12977	12550	93
BB16.29A	30930	30838	30629	30629	11526	11397	340
BB16.29B	22986	22849	22810	22810	9259	9015	163
BB16.29C	23169	23001	22941	22941	9050	8756	66
BB16.29D	31785	31546	31473	31473	11200	10660	26
BB16.30A	26711	26625	26481	26481	11458	11345	541
BB16.30B	17282	17187	17136	17136	7339	7135	69
BB16.30C	27454	27271	27218	27218	10729	10340	141
BB16.30D	30843	30614	30551	30551	11672	11411	98
BB16.31A	23623	23558	23415	23415	9153	9008	307
BB16.31B	15157	15068	15019	15018	6153	6005	104
BB16.31C	21951	21808	21743	21743	8813	8536	109
BB16.31D	25935	25742	25682	25680	9852	9603	77
BB16.32A	16819	16768	16679	16679	7075	7021	116
BB16.32B	22427	22265	22204	22204	8977	8687	179
BB16.32C	22916	22723	22665	22665	8987	8629	140
BB16.32D	22702	22534	22472	22472	8851	8644	6
BB16.33A	30689	30580	30395	30395	12517	12404	503
BB16.33B	30402	30208	30132	30131	12364	11947	292
BB16.33C	22786	22604	22572	22572	9932	9731	185
BB16.33D	27568	27356	27307	27307	10903	10615	10
BB16.34A	34658	34544	34370	34370	13441	13273	562
BB16.34B	14715	14602	14586	14586	6442	6301	161
BB16.34C	20319	20097	20054	20054	8468	8274	379
BB16.34D	24649	24428	24393	24393	9958	9686	0
BB16.35A	36468	36352	36173	36173	14231	14019	1654
BB16.35B	18028	17915	17857	17857	7308	7160	598
BB16.35C	21337	21175	21140	21140	8385	8149	1160
BB16.35D	25708	25492	25456	25456	10876	10688	9
BB16.36A	35185	35072	34899	34899	12744	12534	1638
BB16.36B	19867	19709	19670	19670	7377	7129	692
BB16.36C	18420	18279	18244	18244	6968	6730	411
BB16.36D	16290	16133	16113	16113	6616	6457	7
BB16.37A	35351	35258	35085	35085	14178	14024	551
BB16.37B	11703	11602	11581	11581	5457	5328	106
BB16.37C	12736	12626	12610	12610	5691	5540	178
BB16.37D	21658	21494	21464	21464	9353	9148	25
BB16.38A	42575	42474	42223	42223	16374	16222	564
BB16.38B	15816	15688	15659	15659	6273	6045	1535
BB16.38C	12583	12466	12451	12451	4873	4730	371
BB16.38D	16415	16233	16204	16204	6407	6231	142
BB16.39A	27801	27703	27564	27564	9259	9086	879
BB16.39B	5738	5685	5672	5672	2489	2455	127

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB16.39C	6021	5976	5968	5968	2373	2336	132
BB16.39D	16198	15961	15938	15938	6939	6768	38
BB16.40A	27667	27598	27447	27447	8784	8613	939
BB16.40B	11541	11405	11387	11387	4204	4003	326
BB16.40C	9296	9196	9183	9183	3568	3494	314
BB16.40D	13147	12978	12966	12966	5657	5566	11
BB16.41A	31341	31236	31093	31093	10041	9773	1793
BB16.41B	6369	6284	6271	6271	2391	2355	352
BB16.41C	5272	5227	5198	5198	2250	2216	174
BB16.41D	15230	15103	15078	15078	6569	6421	6
BB16.42A	42426	42310	42083	42083	14722	14522	1158
BB16.42B	5353	5259	5248	5248	1674	1643	75
BB16.42C	8148	8059	8048	8048	3312	3198	170
BB16.42D	7163	6356	6288	6288	439	427	2
BB16.43A	34031	33947	33770	33770	12167	11927	657
BB16.43B	25334	25192	25151	25151	10525	10332	156
BB16.43C	21761	21604	21577	21577	8560	8344	121
BB16.43D	17217	17085	17053	17053	5667	5480	12
BB16.44A	42939	42829	42578	42578	15136	14923	742
BB16.44B	24455	24309	24264	24264	9239	8957	486
BB16.44C	23400	23278	23252	23252	9626	9354	265
BB16.44D	18038	17889	17866	17866	6634	6506	22
BB16.45A	33858	33789	33641	33640	12413	12297	370
BB16.45B	23212	23078	23058	23058	9342	9037	181
BB16.45C	27099	26833	26802	26802	9749	9420	426
BB16.45D	26347	26206	26165	26165	9616	9303	24
BB16.46A	39469	39333	39095	39095	12866	12747	1541
BB16.46B	18827	18716	18687	18687	6786	6585	232
BB16.46C	22353	22186	22149	22149	7547	7386	247
BB16.46D	17106	16930	16898	16898	5998	5880	35
BB16.47A	34176	34071	33908	33908	13550	13426	553
BB16.47B	15151	15014	15002	15002	5722	5584	233
BB16.47C	14980	14865	14852	14852	5357	5235	85
BB16.47D	18408	18193	18164	18164	6960	6784	23
BB16.48A	30588	30494	30343	30343	11098	11000	602
BB16.48B	24368	24129	24107	24107	9712	9412	586
BB16.48C	21771	21594	21568	21568	8148	8015	359
BB16.48D	18927	18729	18711	18711	7204	7023	13
BB16.49A	28867	28783	28637	28637	11520	11370	847
BB16.49B	13670	13555	13527	13527	5770	5684	380
BB16.49C	10971	10893	10857	10857	4370	4285	155
BB16.49D	14536	14428	14393	14393	5901	5742	28
BB16.50A	30294	30206	30057	30057	10943	10845	687
BB16.50B	11760	11679	11666	11666	5019	4921	110
BB16.50C	12842	12730	12716	12716	5353	5216	123
BB16.50D	18573	18439	18406	18406	7170	6913	32
BB16.51A	28689	28611	28477	28477	9853	9735	289
BB16.51B	17043	16932	16906	16906	6469	6310	88
BB16.51C	17717	17613	17591	17591	7275	7158	102
BB16.51D	11816	11721	11684	11684	5093	5016	15
BB17.01A	46868	46771	46738	46732	16500	15369	698
BB17.01B	57537	57423	57377	57370	21804	20665	656
BB17.01C	45070	44989	44965	44958	18002	17234	586
BB17.01D	47596	47497	47427	47419	20723	20143	21
BB17.02A	33957	33896	33874	33870	13327	12827	297
BB17.02B	30744	30711	30694	30691	12857	12436	285
BB17.02C	25822	25779	25772	25767	10842	10520	210
BB17.02D	30491	30407	30387	30378	13649	13156	10
BB17.03A	48542	48446	48423	48409	18032	17096	191
BB17.03B	40407	40338	40312	40307	15240	14698	169
BB17.03C	31052	31006	30985	30982	12782	12484	157
BB17.03D	17374	17316	17276	17276	5806	5717	2

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB17.04A	27714	27656	27626	27621	10971	10610	438
BB17.04B	33876	33827	33807	33803	13769	13383	347
BB17.04C	28681	28627	28618	28614	11423	11040	269
BB17.04D	29928	29852	29817	29814	12320	11906	5
BB17.05A	31190	31119	31086	31082	12049	11470	809
BB17.05B	31712	31647	31624	31620	13737	13375	411
BB17.05C	35048	34987	34959	34954	13444	13078	385
BB17.05D	31778	31714	31676	31672	13066	12649	8
BB17.06A	32016	31964	31947	31941	14470	14127	1018
BB17.06B	31691	31652	31632	31625	13601	13151	1199
BB17.06C	30339	30286	30267	30265	13398	12946	909
BB17.06D	28796	28716	28695	28692	12255	11768	45
BB17.07A	42398	42292	42275	42269	19770	19213	705
BB17.07B	37833	37759	37734	37727	17915	17577	792
BB17.07C	38429	38364	38340	38334	19522	19131	1048
BB17.07D	31469	31421	31390	31389	13691	13412	106
BB17.08A	309863	302226	300807	300716	134101	124214	6260
BB17.08B	17792	17176	17108	17099	8460	8035	509
BB17.08C	201247	195031	194428	194372	91919	87703	6406
BB17.08D	16742	16273	16210	16204	8069	7760	189
BB17.09A	232170	228136	227545	227453	105531	98436	3221
BB17.09B	16629	16041	15983	15979	7695	7172	281
BB17.09C	183030	178768	177839	177778	74551	67770	2124
BB17.09D	18590	18083	18026	18022	9041	8659	119
BB17.10A	124796	122262	121983	121945	60698	57829	3433
BB17.10B	28019	27213	27140	27129	12710	11911	852
BB17.10C	131038	128416	128062	128024	57031	52369	2990
BB17.10D	22697	21851	21782	21775	10143	9417	156
BB17.11A	165702	162689	162215	162160	72143	67006	3014
BB17.11B	262874	256703	255959	255862	123192	118129	7274
BB17.11C	132737	130201	129863	129831	58107	54356	2457
BB17.11D	240760	235789	235103	235015	113232	108852	1405
BB17.12A	299578	294104	293462	293369	134157	125100	7236
BB17.12B	122287	118804	118373	118341	57170	54865	2958
BB17.12C	213381	209074	208540	208466	86846	78654	4622
BB17.12D	341860	336419	335305	335188	157307	151315	3453
BB17.13A	339845	333635	332995	332874	146812	136396	34587
BB17.13B	21324	20723	20658	20655	10698	10240	262
BB17.13C	390927	383203	382413	382299	169592	160141	4786
BB17.13D	29930	29399	29310	29296	13624	12797	106
BB17.14A	297326	290907	290236	290147	143512	138903	3682
BB17.14B	10201	9636	9619	9615	5347	5165	148
BB17.14C	400671	392822	391984	391829	170249	158699	12013
BB17.14D	14254	13999	13954	13950	6941	6745	34
BB17.15A	212933	208201	207366	207297	101823	97978	2135
BB17.15B	7318	6752	6726	6724	3522	3361	163
BB17.15C	233367	227844	227306	227241	109746	105992	5148
BB17.15D	20935	20528	20458	20450	9153	8636	173
BB17.16A	201041	197000	196432	196367	93888	90182	2523
BB17.16B	12679	12173	12143	12141	6069	5818	71
BB17.16C	290473	286202	285389	285287	138344	134864	3733
BB17.16D	12687	12369	12322	12319	6192	5894	126
BB17.17A	36934	36140	36021	36007	17359	16865	302
BB17.17B	16316	15628	15577	15575	7436	7103	73
BB17.17C	51034	50294	50194	50183	24424	23481	517
BB17.17D	17671	17259	17202	17201	8048	7576	68
BB17.18A	10950	10672	10648	10644	5562	5335	59
BB17.18B	20768	19982	19911	19903	8909	8487	1323
BB17.18C	14927	14694	14652	14651	7555	7373	87
BB17.18D	30514	29724	29649	29640	13135	12270	21
BB17.19A	19076	18558	18513	18508	8462	7857	153
BB17.19B	140033	136188	135719	135678	61069	59277	19877

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB17.19C	13910	13646	13592	13587	6439	6137	161
BB17.19D	152579	149964	149546	149489	79726	78453	359
BB17.20A	20410	19823	19757	19749	9928	9430	399
BB17.20B	173377	168806	168205	168150	70651	67600	18447
BB17.20C	11723	11464	11433	11432	5884	5662	336
BB17.20D	358293	353566	352469	352346	164971	160115	502
BB17.21A	34188	33307	33223	33209	16755	16164	926
BB17.21B	14009	13442	13404	13398	6683	6393	344
BB17.21C	18063	17716	17651	17648	8286	7972	1469
BB17.21D	22228	21819	21756	21745	10948	10621	23
BB17.22A	115747	113272	113088	113048	53636	51919	715
BB17.22B	7626	7374	7348	7346	4193	4136	93
BB17.22C	20145	19771	19716	19710	8559	8091	809
BB17.22D	13806	13535	13500	13497	6920	6793	15
BB17.23A	258834	253275	252698	252606	119294	115486	2250
BB17.23B	12597	12166	12137	12134	5583	5159	174
BB17.23C	111707	109854	109575	109542	58430	57549	3096
BB17.23D	21127	20738	20678	20666	10031	9748	9
BB17.24A	187611	183808	183261	183200	81914	76766	504
BB17.24B	14792	14252	14199	14191	6914	6623	134
BB17.24C	259755	256178	255472	255378	122716	119317	7286
BB17.24D	17074	16736	16667	16664	8678	8486	23
BB17.25A	42305	41394	41299	41288	19869	18702	1231
BB17.25B	15991	15271	15228	15222	7495	7122	378
BB17.25C	31791	31256	31165	31151	15723	15224	679
BB17.25D	69624	68667	68503	68484	34489	33536	92
BB17.26A	13988	13564	13533	13533	7168	6935	230
BB17.26B	24723	23784	23738	23735	11563	10882	549
BB17.26C	13912	13661	13633	13626	7813	7675	538
BB17.26D	17543	17159	17103	17097	8807	8565	30
BB17.27A	13092	12788	12741	12739	5885	5547	202
BB17.27B	107630	104433	104130	104099	55858	54649	3925
BB17.27C	17474	17154	17080	17076	8318	8002	1797
BB17.27D	145627	143377	142961	142908	79437	78402	15156
BB17.28A	16467	15905	15850	15843	8270	7867	256
BB17.28C	25812	25375	25315	25309	13069	12599	956
BB17.28D	299645	294876	294121	294021	153187	150329	2579
BB17.29A	15444	14814	14750	14747	8381	8164	358
BB17.29B	133861	129755	129272	129234	70513	68749	1965
BB17.29C	15702	15486	15445	15442	8918	8714	687
BB17.29D	48360	47584	47493	47465	23371	22435	454
BB17.30A	14670	13953	13915	13910	7711	7480	321
BB17.30B	14267	13776	13724	13716	7045	6772	304
BB17.30C	17995	17594	17538	17530	8741	8285	206
BB17.30D	25943	25550	25500	25494	13526	13187	344
BB17.31A	102195	99303	98985	98955	54817	49177	3390
BB17.31B	3486	3280	3262	3260	1944	1854	118
BB17.31C	70681	69311	69098	69077	41592	40855	4312
BB17.31D	16538	15951	15905	15896	8063	7723	92
BB17.32A	74485	72000	71759	71731	41684	40846	2117
BB17.32B	6859	6574	6538	6534	3345	3142	166
BB17.32C	120388	118936	118700	118665	63070	60710	3088
BB17.32D	18691	18379	18331	18328	9539	9222	106
BB17.33A	14483	14130	14081	14079	7018	6605	541
BB17.33B	14484	13845	13794	13790	6556	6129	608
BB17.33C	24456	24100	24059	24050	13844	13466	340
BB17.33D	26479	26045	25986	25981	14851	14525	57
BB17.34A	6576	6068	6053	6053	3311	3212	110
BB17.34B	8768	8322	8276	8274	4301	4120	150
BB17.34C	7598	7449	7433	7432	4263	4188	203
BB17.34D	20607	20099	20054	20049	9992	9414	55
BB17.35A	14539	14166	14140	14137	7074	6610	199

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB17.35B	28132	27340	27264	27255	13056	12131	531
BB17.35C	17450	17217	17193	17185	8372	7789	281
BB17.35D	153944	151048	150686	150629	77761	75640	399
BB17.36A	20353	19699	19664	19657	9824	9178	528
BB17.36B	158042	154097	153572	153523	79167	76046	3885
BB17.36C	12729	12543	12515	12512	7219	6996	349
BB17.36D	357499	351650	350695	350582	177855	171622	743
BB17.37A	19861	19156	19110	19104	9506	8943	1241
BB17.37B	139231	136040	135593	135548	67766	64707	7354
BB17.37C	19879	19587	19556	19546	9576	8823	766
BB17.37D	142013	140134	139877	139836	70823	67763	185
BB17.38A	27400	26405	26351	26344	13028	12100	840
BB17.38B	29908	29039	28962	28956	14373	13364	827
BB17.38C	21393	21065	21041	21032	10116	9164	453
BB17.38D	24993	24591	24544	24530	12823	12303	36
BB17.39A	121471	117336	117037	116996	67061	65667	5476
BB17.39B	7851	7608	7584	7582	4651	4561	322
BB17.39C	69599	67930	67811	67790	38550	37581	2744
BB17.39D	24499	24054	23987	23983	11411	10727	64
BB17.40A	60252	58241	57974	57960	30252	29334	1710
BB17.40B	11398	11029	10994	10986	5425	5073	220
BB17.40C	284729	279680	278545	278466	131286	127333	4971
BB17.40D	29964	29358	29285	29278	16560	16017	252
BB17.41A	5782	5563	5525	5523	2711	2528	124
BB17.41B	13641	13177	13143	13141	6177	5723	249
BB17.41C	33965	33424	33314	33302	15464	14546	470
BB17.41D	26794	26301	26222	26217	12970	12175	150
BB17.42A	786	677	668	668	272	253	23
BB17.42B	11260	10736	10703	10701	5557	5315	695
BB17.42C	13638	13319	13283	13277	6660	6368	495
BB17.42D	41192	40355	40265	40249	19254	18111	150
BB17.43A	9127	8802	8757	8754	4160	3860	1037
BB17.43B	21697	21014	20979	20973	10494	9908	2172
BB17.43C	12241	11965	11927	11925	5366	4867	766
BB17.43D	180685	177618	177185	177126	94803	91584	683
BB17.44A	7528	7131	7105	7105	3698	3510	365
BB17.44B	137662	134005	133640	133587	64403	62270	5545
BB17.44C	15444	15087	15026	15021	7315	6998	647
BB17.44D	344014	338075	336849	336765	163620	156136	876
BB17.45A	15868	14962	14908	14903	7378	6993	556
BB17.45B	124417	121371	120873	120831	51583	47505	3292
BB17.45C	15952	15578	15520	15514	7312	6831	680
BB17.45D	329033	324622	323891	323781	167383	160509	1204
BB17.46A	17719	17048	16996	16985	7767	7338	192
BB17.46B	101747	99541	99280	99240	45219	43241	1278
BB17.46C	18798	18317	18258	18251	7654	6956	217
BB17.46D	230317	227447	226991	226905	129742	126506	931
BB17.47A	257613	251234	250449	250382	100473	93383	4449
BB17.47B	26805	26265	26196	26191	10635	9470	323
BB17.47C	157984	155165	154687	154638	67679	64949	2130
BB17.47D	254363	250676	250092	250005	126280	120591	638
BB17.48A	199761	195333	194664	194597	76622	69214	2344
BB17.48B	32548	31655	31546	31540	12921	11279	455
BB17.48C	399003	392430	391129	390997	161873	154040	5385
BB17.48D	297242	292984	292444	292346	165381	159626	278
BB17.49A	17033	16525	16455	16450	8175	7705	420
BB17.49B	15531	14771	14692	14689	7031	6482	485
BB17.49C	31289	30664	30547	30536	14204	13259	881
BB17.49D	244985	241467	240985	240889	130444	125610	745
BB17.50A	6146	5916	5893	5893	3342	3173	78
BB17.50B	6234	5954	5911	5911	3078	2881	81
BB17.50C	13024	12734	12696	12695	6285	5869	53

Pipeline Steps >>>>>						FINAL All Reads	FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
BB17.50D	218847	214911	214504	214442	111788	106400	249
BB17.51A	5506	5261	5227	5226	2616	2409	105
BB17.51B	155121	151063	150738	150681	70002	65330	2524
BB17.51C	15493	15130	15071	15061	6375	5716	103
BB17.51D	215583	211816	211065	210961	97746	91478	474

Supplemental Data A3. Number of reads filtered during the sequence processing pipeline of Chapter 2; data for AZMP (Atlantic Zone Monitoring Program) samples V6-V8 *16S* rRNA are shown. Values highlighted red lacked chloroplast and cyanobacterial reads after data processing.

Pipeline Steps >>>>> Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
16SV6-AF14-HL1-1m-L	69496	69223	69048	69044	33388	32366	5626
16SV6-AF14-HL1-1m-Se	15894	15847	15737	15720	6235	5896	357
16SV6-AF14-HL1-1m-Sf	62603	62456	62104	62056	24626	23383	1621
16SV6-AF14-HL1-20m-L	41247	41117	41013	41009	21144	20509	4041
16SV6-AF14-HL1-20m-Sd	47698	47589	47324	47286	19544	18826	997
16SV6-AF14-HL1-20m-Sf	47137	46992	46656	46626	18478	17704	1154
16SV6-AF14-HL1-40m-L	64725	64501	64350	64347	33653	31757	4021
16SV6-AF14-HL1-40m-Sd	56653	56533	56266	56234	21712	20483	282
16SV6-AF14-HL1-40m-Sf	54774	54652	54371	54338	23314	22457	359
16SV6-AF14-HL1-60m-L	66757	66513	66352	66350	36065	34604	2659
16SV6-AF14-HL1-60m-Se	66372	66193	65865	65832	22977	21674	195
16SV6-AF14-HL1-60m-Sf	45979	45869	45657	45622	16726	15415	133
16SV6-AF14-HL11-1m-L	70398	70135	69967	69965	33733	31871	2759
16SV6-AF14-HL11-1m-Sd	57637	57517	57337	57293	14491	12089	1706
16SV6-AF14-HL11-1m-Se	16464	16426	16353	16344	4063	3220	394
16SV6-AF14-HL11-20m-L	67792	67565	67332	67330	26212	23569	3659
16SV6-AF14-HL11-20m-Sd	48742	48639	48483	48444	9836	7604	1810
16SV6-AF14-HL11-20m-Sf	36732	36617	36507	36485	7660	5754	1335
16SV6-AF14-HL11-250m-L	109925	109429	109134	109130	53551	51082	489
16SV6-AF14-HL11-250m-Se	39375	39306	39182	39143	12848	10593	3
16SV6-AF14-HL11-250m-Sf	49000	48889	48721	48682	15681	13572	3
16SV6-AF14-HL11-80m-L	70840	70544	70393	70392	30889	27890	3486
16SV6-AF14-HL11-80m-Sd	57199	57069	56918	56867	15985	12740	362
16SV6-AF14-HL11-80m-Se	32346	32276	32163	32145	11062	9717	238
16SV6-AF14-HL2-1m-L	74006	73723	73563	73563	36380	35029	7302
16SV6-AF14-HL2-1m-Sd	44330	44246	44075	44035	16156	15124	887
16SV6-AF14-HL2-1m-Sf	43156	43070	42873	42852	18584	18040	1285
16SV6-AF14-HL2-20m-L	55111	54888	54785	54784	24556	23714	3607
16SV6-AF14-HL2-20m-Se	41425	41349	41211	41172	16652	15930	986
16SV6-AF14-HL2-20m-Sf	20978	20903	20788	20765	7768	7403	547
16SV6-AF14-HL2-40m-L	70764	70518	70365	70361	32913	31149	3310
16SV6-AF14-HL2-40m-Se	42591	42497	42304	42278	15978	14907	223
16SV6-AF14-HL2-40m-Sf	29795	29710	29568	29549	11340	10681	109
16SV6-AF14-HL2-80m-L	45734	45639	45535	45534	21847	19746	723
16SV6-AF14-HL2-80m-Sd	30324	30262	30140	30120	10356	9386	14
16SV6-AF14-HL2-80m-Se	30879	30814	30701	30680	10582	9489	25
16SV6-AF14-HL4-1m-L	24833	24735	24672	24671	11689	11234	2308
16SV6-AF14-HL4-1m-Sd	30764	30714	30589	30564	11038	10305	318
16SV6-AF14-HL4-1m-Sf	23761	23697	23608	23588	8310	7507	388
16SV6-AF14-HL4-20m-L	35520	35425	35296	35296	17423	16775	4288
16SV6-AF14-HL4-20m-Se	55133	55015	54700	54655	21337	20593	912
16SV6-AF14-HL4-20m-Sf	44505	44408	44171	44142	16521	15592	590
16SV6-AF14-HL4-40m-L	11204	11144	11105	11104	4147	3793	1021
16SV6-AF14-HL4-40m-Sd	41827	41728	41539	41508	14341	13473	369
16SV6-AF14-HL4-40m-Se	47271	47172	46924	46896	18458	17863	747
16SV6-AF14-HL4-60m-L	64281	64048	63917	63912	28244	26615	2933
16SV6-AF14-HL4-60m-Se	65955	65843	65559	65516	22673	20709	190
16SV6-AF14-HL4-60m-Sf	38217	38118	37917	37888	13070	12289	135
16SV6-AF14-HL6-1m-L	34836	34638	34549	34546	16883	15994	4314
16SV6-AF14-HL6-1m-Se	47707	47614	47323	47298	11960	10356	1756
16SV6-AF14-HL6-1m-Sf	32438	32358	32185	32169	8133	6991	990
16SV6-AF14-HL6-20m-L	33063	32952	32857	32856	15138	14341	4041
16SV6-AF14-HL6-20m-Sd	34631	34559	34403	34388	8411	7091	1491
16SV6-AF14-HL6-20m-Se	34879	34811	34680	34661	8415	7092	1333

Pipeline Steps >>>>>							FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	FINAL All Reads Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
16SV6-AF14-HL6-250m-L	79106	78762	78524	78521	38615	33275	1023
16SV6-AF14-HL6-250m-Se	55716	55590	55353	55321	14547	11993	46
16SV6-AF14-HL6-250m-Sf	63332	63172	62962	62915	17755	14848	74
16SV6-AF14-HL6-80m-L	64777	64464	64318	64313	28327	26564	2895
16SV6-AF14-HL6-80m-Sd	45468	45378	45156	45117	15026	13564	86
16SV6-AF14-HL6-80m-Se	34843	34764	34589	34572	11361	9867	54
16SV6-AF14-HL8-100m-L	78198	77974	77796	77795	37389	35087	1997
16SV6-AF14-HL8-100m-Sd	67391	67256	66933	66883	19794	17015	27
16SV6-AF14-HL8-100m-Sf	63062	62866	62601	62541	17688	15117	28
16SV6-AF14-HL8-1m-L	99422	99006	98759	98753	47705	45989	5371
16SV6-AF14-HL8-1m-Sd	64176	64042	63839	63788	16379	14092	1082
16SV6-AF14-HL8-1m-Se	38478	38381	38216	38193	11161	10422	1101
16SV6-AF14-HL8-20m-L	72342	72035	71875	71872	28451	26486	5834
16SV6-AF14-HL8-20m-Sd	52669	52556	52386	52353	11510	9485	1892
16SV6-AF14-HL8-20m-Sf	52829	52690	52514	52482	12725	11144	2979
16SV6-AF14-HL8-250m-L	78417	78188	78025	78021	38941	36092	263
16SV6-AF14-HL8-250m-Se	58008	57872	57656	57599	17345	15353	2
16SV6-AF14-HL8-250m-Sf	55932	55809	55584	55537	16567	14751	8
16SV6-AF16-HL1-1m-L	21614	21515	21177	21177	6845	6591	2127
16SV6-AF16-HL1-1m-S	20900	20831	20484	20484	5769	5612	253
16SV6-AF16-HL1-20m-L	24824	24707	24375	24375	7933	7685	3111
16SV6-AF16-HL1-20m-S	24403	24291	23913	23913	6716	6536	373
16SV6-AF16-HL1-40m-L	38267	38093	37526	37526	9711	9406	3567
16SV6-AF16-HL1-40m-S	49376	49203	48320	48320	10663	10419	346
16SV6-AF16-HL1-60m-L	52655	52435	51647	51647	16251	15687	1463
16SV6-AF16-HL1-60m-S	41132	40911	40207	40207	14436	14016	2345
16SV6-AF16-HL11-100m-L	89474	89050	87957	87957	33807	33326	546
16SV6-AF16-HL11-100m-S	70725	70423	69331	69331	12992	12280	41
16SV6-AF16-HL11-1m-L	73165	72802	71685	71685	28288	26821	3445
16SV6-AF16-HL11-1m-S	62493	62252	61337	61337	8486	7721	2830
16SV6-AF16-HL11-250m-L	76168	75769	74801	74801	33706	32435	179
16SV6-AF16-HL11-250m-S	88424	88063	86802	86802	16475	15221	2
16SV6-AF16-HL11-45m-L	57635	57354	56652	56652	20185	19266	2755
16SV6-AF16-HL11-45m-S	66301	66017	65058	65058	8531	7815	1044
16SV6-AF16-HL2-1m-L	35849	35689	35078	35078	10982	10694	2837
16SV6-AF16-HL2-1m-S	46788	46639	45844	45844	12363	12027	541
16SV6-AF16-HL2-20m-L	39339	39165	38478	38478	12194	11885	3960
16SV6-AF16-HL2-20m-S	49345	49175	48388	48388	12545	12112	684
16SV6-AF16-HL2-40m-L	46398	46173	45545	45545	15227	14898	6824
16SV6-AF16-HL2-40m-S	55233	55046	54184	54184	15859	15483	366
16SV6-AF16-HL2-80m-L	1083	474	329	329	62	45	0
16SV6-AF16-HL2-80m-S	55429	55222	54355	54355	14998	14652	46
16SV6-AF16-HL2b-1m-L	28117	28023	27621	27621	11204	10989	1168
16SV6-AF16-HL2b-1m-S	29090	28985	28620	28620	7837	7616	199
16SV6-AF16-HL2b-20m-L	2417	2394	2355	2355	1006	988	124
16SV6-AF16-HL2b-20m-S	29047	28937	28621	28621	8789	8662	693
16SV6-AF16-HL2b-40m-L	30770	30637	30232	30232	8961	8871	93
16SV6-AF16-HL2b-40m-S	233	194	171	171	27	25	0
16SV6-AF16-HL2b-80m-L	29281	29164	28792	28792	11502	11288	449
16SV6-AF16-HL2b-80m-S	50114	49917	49290	49290	10124	9709	0
16SV6-AF16-HL4-1m-L	54486	54277	53560	53560	16043	15705	2939
16SV6-AF16-HL4-1m-S	56724	56493	55630	55630	14234	13805	567
16SV6-AF16-HL4-20m-L	46529	46366	45788	45788	13462	13124	2166
16SV6-AF16-HL4-20m-S	44059	43838	43138	43138	11572	11321	434
16SV6-AF16-HL4-40m-L	50065	49845	49308	49308	19119	18692	15554
16SV6-AF16-HL4-40m-S	54315	54098	53275	53275	11082	10732	819
16SV6-AF16-HL4-60m-L	51857	51606	50897	50897	17722	16819	7679
16SV6-AF16-HL4-60m-S	58494	58218	57396	57396	12271	11824	141
16SV6-AF16-HL5.5-1m-L	73850	73511	72419	72419	24635	23723	9906
16SV6-AF16-HL5.5-1m-S	54641	54471	53660	53660	11087	10553	996
16SV6-AF16-HL5.5-20m-L	41316	40975	39926	39926	7715	7551	3518
16SV6-AF16-HL5.5-20m-S	46300	46118	45460	45460	8586	8100	931
16SV6-AF16-HL5.5-250m-L	58024	57726	56919	56919	20432	17588	4580

Pipeline Steps >>>>>							
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
16SV6-AF16-HL5.5-250m-S	89537	89247	88021	88021	14052	12706	11
16SV6-AF16-HL5.5-80m-L	20415	18850	16604	16604	524	511	300
16SV6-AF16-HL5.5-80m-S	63023	62838	61936	61936	13341	12849	247
16SV6-AF16-HL6-1m-L	79160	78851	77756	77756	26753	25882	10565
16SV6-AF16-HL6-1m-S	59750	59602	58792	58792	9124	8389	1755
16SV6-AF16-HL6-20m-L	54965	54691	53818	53818	14857	13929	5949
16SV6-AF16-HL6-20m-S	70541	70320	69514	69514	10515	9684	3054
16SV6-AF16-HL6-50m-L	63887	63568	62328	62328	17662	17097	6875
16SV6-AF16-HL6-50m-S	57980	57826	57120	57120	9249	8616	1063
16SV6-AF16-HL6-80m-L	56961	56717	55915	55915	16803	16069	4656
16SV6-AF16-HL6-80m-S	79960	79674	78446	78446	11634	11113	407
16SV6-AF16-HL7-1m-L	74874	74564	73332	73332	26651	26078	5213
16SV6-AF16-HL7-1m-S	18105	18037	17741	17741	2520	2328	586
16SV6-AF16-HL7-20m-L	51535	51288	50438	50438	17393	16485	4449
16SV6-AF16-HL7-20m-S	56737	56515	55707	55707	6676	6083	1521
16SV6-AF16-HL7-50m-L	58495	58231	57501	57501	16918	15986	6497
16SV6-AF16-HL7-50m-S	58791	58581	57717	57717	8037	7654	1103
16SV6-AF16-HL7-80m-L	74839	74402	73533	73533	26001	23886	9172
16SV6-AF16-HL7-80m-S	58681	58421	57631	57631	11742	10936	382
16SV6-AF16-HL8-100m-L	60126	59840	58823	58823	19227	18137	2748
16SV6-AF16-HL8-100m-S	76204	75960	74724	74724	12061	11073	69
16SV6-AF16-HL8-1m-L	74150	73819	72260	72260	27090	26213	4216
16SV6-AF16-HL8-1m-S	75518	75271	74098	74098	10812	10051	2533
16SV6-AF16-HL8-250m-L	54050	53743	52885	52885	18607	16533	441
16SV6-AF16-HL8-250m-S	66222	65941	64911	64911	14408	13371	7
16SV6-AF16-HL8-60m-L	51988	51784	51149	51149	16819	16089	7610
16SV6-AF16-HL8-60m-S	73299	73010	71924	71924	11638	10982	1419
16SV6-AS14-HL1-1m-L	47051	46873	46775	46772	25686	24693	5587
16SV6-AS14-HL1-1m-Se	44118	44024	43826	43797	17925	17306	805
16SV6-AS14-HL1-1m-Sf	25902	25844	25734	25721	10883	10481	267
16SV6-AS14-HL1-20m-L	6700	6636	6608	6607	3723	3569	238
16SV6-AS14-HL1-20m-Sd	28582	28531	28347	28326	12064	11579	267
16SV6-AS14-HL1-20m-Sf	42357	42245	42033	42006	18116	17570	669
16SV6-AS14-HL1-40m-L	868	837	806	806	318	295	29
16SV6-AS14-HL1-40m-Sd	71254	71085	70792	70732	28774	27576	611
16SV6-AS14-HL1-40m-Se	40854	40748	40484	40472	15557	15174	449
16SV6-AS14-HL1-60m-L	36640	36506	36401	36400	20515	19346	3857
16SV6-AS14-HL1-60m-Sd	38526	38420	38255	38233	15327	14612	184
16SV6-AS14-HL1-60m-Se	30161	30099	29838	29810	11440	10446	85
16SV6-AS14-HL2-1m-L	43272	43125	42947	42945	26159	24387	7449
16SV6-AS14-HL2-1m-Sd	67586	67429	67129	67083	27063	25928	1531
16SV6-AS14-HL2-1m-Sf	32215	32121	31938	31921	12569	11890	555
16SV6-AS14-HL2-20m-L	14115	14068	14043	14043	8029	7364	2091
16SV6-AS14-HL2-20m-Sd	1462	1458	1419	1417	524	460	8
16SV6-AS14-HL2-20m-Sf	26191	26122	25961	25946	11984	11672	339
16SV6-AS14-HL2-40m-L	59305	59101	58896	58892	34517	31148	8485
16SV6-AS14-HL2-40m-Sd	50047	49958	49727	49689	21118	20213	765
16SV6-AS14-HL2-40m-Se	29345	29256	29089	29056	12142	11735	312
16SV6-AS14-HL2-80m-L	15010	14973	14929	14929	8266	7646	3837
16SV6-AS14-HL2-80m-Se	3163	3149	3048	3046	1222	1126	24
16SV6-AS14-HL2-80m-Sf	38210	38108	37900	37878	15784	15249	470
16SV6-AS14-HL4-1m-L	66227	66007	65889	65887	31354	30344	15181
16SV6-AS14-HL4-1m-Se	20581	20541	20399	20384	7491	6841	357
16SV6-AS14-HL4-1m-Sf	36356	36287	36080	36052	13586	12543	794
16SV6-AS14-HL4-20m-L	60079	59814	59739	59738	27549	26660	11430
16SV6-AS14-HL4-20m-Sd	41075	40983	40725	40691	17099	16285	1214
16SV6-AS14-HL4-20m-Se	22048	21985	21788	21774	8378	7673	353
16SV6-AS14-HL4-40m-L	79732	79345	79190	79187	35124	33652	13499
16SV6-AS14-HL4-40m-Sd	28790	28708	28503	28484	11176	10585	142
16SV6-AS14-HL4-40m-Sf	49047	48925	48607	48584	18670	17453	256
16SV6-AS14-HL4-60m-L	64608	64391	64263	64258	28561	25782	1244
16SV6-AS14-HL4-60m-Sd	32745	32658	32463	32433	11695	10598	17
16SV6-AS14-HL4-60m-Sf	33804	33720	33436	33409	12150	11086	28

Pipeline Steps >>>>>>							FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	FINAL All Reads Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
16SV6-AS14-HL5.5-1m-L	41822	41643	41549	41544	17582	16783	7001
16SV6-AS14-HL5.5-1m-Sd	42567	42459	42170	42140	16644	15627	1191
16SV6-AS14-HL5.5-1m-Se	36222	36147	35883	35855	13872	12808	758
16SV6-AS14-HL5.5-20m-L	32043	31890	31825	31824	15128	14750	5958
16SV6-AS14-HL5.5-20m-Sd	22407	22354	22168	22153	9078	8474	376
16SV6-AS14-HL5.5-20m-Sf	26162	26112	25948	25932	11542	11063	824
16SV6-AS14-HL5.5-250m-L	42278	42097	42017	42016	17535	15506	3015
16SV6-AS14-HL5.5-250m-Se	31871	31795	31613	31592	11816	10615	43
16SV6-AS14-HL5.5-250m-Sf	21462	21418	21254	21236	7644	6758	41
16SV6-AS14-HL5.5-80m-L	64332	64172	64069	64066	33299	32193	20715
16SV6-AS14-HL5.5-80m-Sd	8891	8861	8770	8768	3523	3338	85
16SV6-AS14-HL5.5-80m-Sf	21067	21023	20923	20915	7301	6669	100
16SV6-AS14-HL8-100m-L	83203	82891	82725	82724	40255	38438	23485
16SV6-AS14-HL8-100m-Sd	64384	64251	63859	63815	17675	14822	56
16SV6-AS14-HL8-100m-Se	49878	49786	49485	49459	13713	11345	44
16SV6-AS14-HL8-1m-L	39218	39038	38948	38947	15771	14834	5209
16SV6-AS14-HL8-1m-Sd	24728	24689	24533	24517	9408	8693	90
16SV6-AS14-HL8-1m-Sf	43164	43061	42819	42788	16705	15436	198
16SV6-AS14-HL8-20m-L	53402	53234	53142	53138	22010	20347	9471
16SV6-AS14-HL8-20m-Sd	39096	39015	38799	38776	15063	13944	203
16SV6-AS14-HL8-20m-Se	38949	38849	38600	38571	14534	13423	228
16SV6-AS14-HL8-250m-L	53881	53589	53441	53437	23756	21891	5644
16SV6-AS14-HL8-250m-Sd	25332	25278	25091	25074	8353	6808	11
16SV6-AS14-HL8-250m-Se	31835	31761	31665	31641	10360	8039	5
16SV6-AS16-HL1-1m-L	22605	22403	22363	22363	10777	10633	5644
16SV6-AS16-HL1-1m-S	24457	24272	24234	24233	8724	8476	518
16SV6-AS16-HL1-20m-L	30061	29801	29767	29767	15689	15542	5113
16SV6-AS16-HL1-20m-S	25221	25025	24964	24964	8119	7820	403
16SV6-AS16-HL1-40m-L	26815	26591	26543	26542	14737	14546	3761
16SV6-AS16-HL1-40m-S	27194	26978	26927	26927	8921	8550	64
16SV6-AS16-HL1-60m-L	24044	23868	23807	23807	12662	12355	5125
16SV6-AS16-HL1-60m-S	24986	24829	24785	24785	7876	7557	111
16SV6-AS16-HL2-1m-L	15401	15218	15193	15193	8675	8587	2353
16SV6-AS16-HL2-1m-S	20651	20529	20503	20503	7800	7654	441
16SV6-AS16-HL2-20m-L	23814	23670	23643	23642	12338	12270	5843
16SV6-AS16-HL2-20m-S	27276	27150	27114	27114	10562	10342	462
16SV6-AS16-HL2-40m-L	37313	37081	37046	37046	17243	17103	10809
16SV6-AS16-HL2-40m-S	24380	24283	24249	24249	9480	9234	208
16SV6-AS16-HL2-80m-L	27421	27260	27224	27224	11729	11432	2634
16SV6-AS16-HL2-80m-S	23298	23163	23126	23126	6200	5689	21
16SV6-AS16-HL2b-1m-L	22282	22081	22023	22023	10648	10561	3701
16SV6-AS16-HL2b-1m-S	30130	29920	29853	29852	12055	11877	272
16SV6-AS16-HL2b-20m-L	24986	24812	24695	24695	13934	13847	6408
16SV6-AS16-HL2b-20m-S	29542	29327	29263	29263	11682	11485	407
16SV6-AS16-HL2b-40m-L	29605	29385	29330	29330	15176	15063	796
16SV6-AS16-HL2b-40m-S	28012	27792	27725	27725	11979	11807	121
16SV6-AS16-HL2b-80m-L	28436	28100	28023	28023	11700	11435	1202
16SV6-AS16-HL2b-80m-S	26408	26234	26177	26177	7860	7549	8
16SV6-AS16-HL4-1m-L	13864	13736	13715	13715	5646	5506	1201
16SV6-AS16-HL4-1m-S	24130	23992	23938	23938	6104	5922	156
16SV6-AS16-HL4-20m-L	23343	23179	23144	23144	8772	8538	1238
16SV6-AS16-HL4-20m-S	34406	34204	34155	34154	9575	9157	217
16SV6-AS16-HL4-40m-L	25552	25358	25319	25319	9649	9509	4402
16SV6-AS16-HL4-40m-S	33496	33327	33278	33278	10180	9766	286
16SV6-AS16-HL4-60m-L	41238	40980	40937	40936	15765	15393	1414
16SV6-AS16-HL4-60m-S	25057	24858	24831	24831	6701	6182	44
16SV6-AS16-HL5.5-1m-L	21310	21164	21126	21126	7854	7682	1304
16SV6-AS16-HL5.5-1m-S	31880	31684	31612	31612	10704	10399	342
16SV6-AS16-HL5.5-20m-L	26557	26407	26358	26358	12430	12336	2437
16SV6-AS16-HL5.5-20m-S	29017	28863	28807	28807	9281	8886	317
16SV6-AS16-HL5.5-250m-L	31693	31470	31389	31389	11696	10620	483
16SV6-AS16-HL5.5-250m-S	36450	36246	36186	36186	7525	6685	2
16SV6-AS16-HL5.5-80m-L	23521	23390	23354	23353	9492	9208	1946

Pipeline Steps >>>>>							FINAL
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	FINAL All Reads Remove Rare ASVs	Chloroplast AND Cyanobacterial Reads
16SV6-AS16-HL5.5-80m-S	29677	29478	29413	29413	8699	8249	133
16SV6-AS16-HL6-1m-L	20169	20037	20011	20011	7842	7691	1176
16SV6-AS16-HL6-1m-S	38427	38265	38194	38194	11472	11077	228
16SV6-AS16-HL6-20m-L	19936	19827	19798	19798	8027	7860	706
16SV6-AS16-HL6-20m-S	19472	19389	19359	19359	6530	6330	78
16SV6-AS16-HL6-50m-L	35495	35334	35292	35291	15220	14831	4705
16SV6-AS16-HL6-50m-S	14667	14600	14570	14570	4071	3882	29
16SV6-AS16-HL6-80m-L	20659	20539	20499	20499	9586	9376	1834
16SV6-AS16-HL6-80m-S	28958	28765	28715	28715	7710	7258	76
16SV6-AS16-HL7-1m-L	23783	23609	23563	23563	9932	9614	5618
16SV6-AS16-HL7-1m-S	42140	41930	41844	41844	10473	9947	720
16SV6-AS16-HL7-20m-L	38172	37911	37845	37845	17188	16770	5071
16SV6-AS16-HL7-20m-S	41124	40901	40813	40813	9756	9129	437
16SV6-AS16-HL7-50m-L	35308	35112	35049	35049	14219	13736	5028
16SV6-AS16-HL7-50m-S	51335	51033	50940	50940	13329	12425	530
16SV6-AS16-HL7-80m-L	24376	24247	24210	24210	9543	9164	4551
16SV6-AS16-HL7-80m-S	39933	39735	39667	39665	10027	9428	461
AF17-HL01-1m-L-501a	67162	65907	65857	65857	34406	30880	13095
AF17-HL01-1m-S-501b	134260	131898	131768	131768	57488	55787	3460
AF17-HL01-20m-L-502a	60797	59793	59767	59767	29914	24738	4962
AF17-HL01-20m-S-502b	150180	147605	147525	147525	72868	71568	2614
AF17-HL01-40m-L-503a	72127	70823	70787	70787	40049	37072	1853
AF17-HL01-40m-S-503b	170579	167850	167734	167734	75264	73642	592
AF17-HL01-60m-L-504a	41539	40656	40589	40589	22355	19903	880
AF17-HL01-60m-S-504b	171316	168283	168142	168142	71919	69973	155
AF17-HL02-1m-L-505a	10823	10510	10505	10505	5501	5367	2086
AF17-HL02-1m-S-505b	198796	195736	195650	195650	88618	87133	6645
AF17-HL02-20m-L-506a	30789	30205	30195	30195	16811	16360	4586
AF17-HL02-20m-S-506b	162244	159450	159376	159376	69830	68609	3056
AF17-HL02-40m-L-507a	83857	82329	82297	82297	42295	36558	2746
AF17-HL02-40m-S-507b	143859	141510	141432	141432	62392	61222	464
AF17-HL02-80m-L-508a	25116	24606	24586	24586	13987	13495	212
AF17-HL02-80m-S-508b	78625	76633	76519	76519	28496	28085	21
AF17-HL04-1m-L-509a	37475	36553	36536	36536	20348	20056	5115
AF17-HL04-1m-S-509b	134488	131917	131837	131837	59501	58734	3855
AF17-HL04-20m-L-510a	30923	30197	30186	30186	17047	16746	4246
AF17-HL04-20m-S-510b	99455	97531	97487	97487	41548	40886	2129
AF17-HL04-40m-L-511a	57585	56307	56285	56285	31397	30581	3455
AF17-HL04-40m-S-511b	150835	148134	148054	148054	60160	59036	1157
AF17-HL04-60m-L-512a	42954	41946	41918	41918	21394	17932	675
AF17-HL04-60m-S-512b	134715	132198	132125	132125	50142	48702	150
AF17-HL06-1m-L-513a	34192	33353	33331	33331	18103	17707	4662
AF17-HL06-1m-S-513b	216173	212790	212659	212659	85870	83975	5023
AF17-HL06-20m-L-514a	39472	38695	38675	38675	21122	20428	6093
AF17-HL06-20m-S-514b	176811	173887	173809	173809	59406	56920	2610
AF17-HL06-50m-L-515a	92115	90454	90417	90417	47920	46421	11635
AF17-HL06-50m-S-515b	175945	173167	173081	173081	55243	53268	2035
AF17-HL06-80m-L-516a	31587	30862	30840	30840	17146	16218	2146
AF17-HL06-80m-S-516b	176060	173074	172941	172941	46350	44574	1098
AF17-HL07-1m-L-517a	11844	11382	11361	11361	5891	5567	2003
AF17-HL07-1m-S-517b	185303	181795	181674	181674	48650	46141	3516
AF17-HL07-20m-L-518a	21227	20704	20692	20692	10715	10114	2886
AF17-HL07-20m-S-518b	206203	202651	202524	202524	45897	42823	3032
AF17-HL07-50m-L-519a	60009	58659	58594	58594	28276	26586	9452
AF17-HL07-50m-S-519b	239013	235391	235258	235258	55041	52321	3858
AF17-HL07-80m-L-520a	45892	45059	45004	45004	27559	26306	8943
AF17-HL07-80m-S-520b	179871	176792	176672	176672	42021	39871	2919
AS17-HL02-1m-L-421a	115273	113164	113107	113107	55638	48517	3183
AS17-HL02-1m-S-421c	147688	145225	145138	145138	81209	80546	2040
AS17-HL02-20m-L-422a	23735	23240	23234	23234	13740	13623	1199
AS17-HL02-20m-S-422c	72187	70773	70705	70705	40696	40469	926
AS17-HL02-40m-L-423a	30823	30240	30234	30234	18454	18337	1604
AS17-HL02-40m-S-423c	3354	3216	3208	3208	1986	1979	19

Pipeline Steps >>>>>>							
Sample	Initial Pairs	Cutadapt	Pear	Quality Filtered	Deblur	FINAL All Reads Remove Rare ASVs	FINAL Chloroplast AND Cyanobacterial Reads
AS17-HL02-80m-L-424a	28315	27730	27721	27721	16270	16174	2583
AS17-HL02-80m-S-424c	84091	82496	82446	82446	43547	43334	614
AS17-HL02R-1m-L-453a	210	192	191	191	115	115	22
AS17-HL02R-1m-S-453c	1387	1331	1317	1317	773	751	84
AS17-HL02R-20m-L-454a	8197	8043	8041	8041	4521	4486	765
AS17-HL02R-20m-S-454c	3275	3194	3188	3188	1852	1825	140
AS17-HL02R-40m-L-455a	6841	6716	6714	6714	4237	4203	87
AS17-HL02R-40m-S-455c	733	654	646	646	375	369	17
AS17-HL02R-80m-L-456a	78187	77034	76965	76965	44835	44568	5422
AS17-HL02R-80m-S-456c	157240	154853	154787	154787	60524	59080	257
AS17-HL04-1m-L-418a	18944	18550	18542	18542	11053	10999	353
AS17-HL04-1m-S-418c	78811	77348	77292	77292	42400	42036	989
AS17-HL04-20m-L-419a	27648	27099	27094	27094	15992	15921	500
AS17-HL04-20m-S-419c	80990	79413	79358	79358	46503	46373	559
AS17-HL04-60m-L-420a	107174	105044	104981	104981	59142	58618	4574
AS17-HL04-60m-S-420c	149422	146768	146675	146675	69529	68878	526
AS17-HL06-1m-L-414a	19318	18934	18912	18912	11600	11522	2443
AS17-HL06-1m-S-414c	106187	104205	104150	104150	52439	52016	976
AS17-HL06-20m-L-415a	13942	13686	13672	13672	8406	8362	1920
AS17-HL06-20m-S-415c	39485	38781	38741	38741	18296	17995	108
AS17-HL06-50m-L-416a	31657	31123	31103	31103	18380	18177	4057
AS17-HL06-50m-S-416c	201360	197896	197759	197759	86444	84735	591
AS17-HL06-80m-L-417a	94523	92662	92370	92370	50620	49564	8451
AS17-HL06-80m-S-417c	177396	174483	174335	174335	67336	65630	279
AS17-HL07-1m-L-410a	102653	100972	100935	100935	62956	62670	7273
AS17-HL07-1m-S-410c	51119	50174	50135	50135	19887	19274	131
AS17-HL07-20m-L-411a	27118	26687	26673	26673	16722	16645	5001
AS17-HL07-20m-S-411c	54050	53193	53151	53151	22556	21789	179
AS17-HL07-50m-L-412a	32796	32271	32259	32259	18798	18678	2292
AS17-HL07-50m-S-412c	134164	132025	131953	131953	50318	48625	361
AS17-HL07-80m-L-413a	103688	102028	101925	101925	51293	50798	3731
AS17-HL07-80m-S-413c	114245	112557	112479	112479	44880	43834	269