TOWARDS UNDERSTANDING THE PROKARYOTIC CONTRIBUTIONS TO COBALAMIN CYCLING IN THE NORTHWEST ATLANTIC

by

Maria Alejandra Soto Rojas

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
March 2021

Table of Contents

List of Tables

List of Figures

Abstract

Cobalamin has the potential to limit primary productivity and shape the structure and ecological interactions of marine microbial communities. The identification of major sources and sinks of this vitamin is needed in order to understand its availability in the ocean. In this thesis, assembly-based and short-read-based approaches were combined to analyze metagenomic samples from the Scotian Shelf and Slope region of the Northwest Atlantic. This resulted in the first identification of major producers, remodelers and consumers of cobalamin and related compounds in this region. Mass-spectrometry tools to monitor the contribution of *Synechococcus*, an important cyanobacterial group, to the cobalamin cycle in the Northwest Atlantic were also identified. The implementation of these tools in culture experiments enabled the identification of environmental and physiological factors with potential to affect cyanobacterial contributions to cobalamin cycling in this region. In sum, this thesis is a step towards elucidating the influence that cobalamin may have on marine primary productivity and microbial ecological interactions in the Northwest Atlantic.

List of Abbreviations and Symbols Used

16s rRNA	16S Ribosomal RNA gene
Ado-B_{12}	Adenosylcobalamin
ANI	Average nucleotide identity
ANOVA	Analysis of Variance
AZMP	Atlantic Zone Monitoring Program
B_1	Thiamine
B_2	Riboflavin
BIO	Bedford Institute of Oceanography
Bp	Base pair
CN-B_{12}	Cyanocobalamin
CO2	Carbon dioxide
COG	Clusters of Orthologous Groups of proteins
DHPS	2,3-dihydroxypropane-1-sulfonate
DMB	5,6-dimethylbenzimidazole
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESI	Electrospray ionisation
GTDB	Genome Taxonomy Database
HMM	Hidden Markov model
HNLC	High Nutrient, Low Chlorophyll
HPLC	High performance liquid chromatography
LWR	Likelihood weight ratio
MAG	Metagenome-assembled genomes
MCM	Methylmalonyl CoA mutase
Me-B_{12}	Methylcobalamin
MetH	Methionine synthase
MLD	Mixed layer depth
MS	Mass spectrometer

N:P	Nitrogen to phosphorus ratio
NCMA	National Center for Marine Algae and Microbiota
OH-B_{12}	Hydroxocobalamin
OTU	Operational Taxonomic Unit
POC	Particulate Organic Carbon
PSHCP	*Prochlorococcus*/*Synechococcus* Hyper Conserved Protein
QC	Quality control
RFU	Relative fluorescence unit
RPKM	Reads Per Kilobase per Million
S18	30S ribosomal protein S18
SAM	S-adenosylmethionine
SCG	Single Copy Gene
SE	Surface euphotic
SOW	Synthetic Ocean Water
SRM	Selected reaction monitoring
V6-V8	Variable region 6 to Variable region 8 of 16S rRNA gene

Acknowledgements

I would like to thank to my supervisor, Dr. Erin Bertrand, for her guidance and support through this project. I learned a lot from her and I am very thankful for that. I would also like to thank to all Bertrand Lab members. They were all a big support through this project. I would like to thank to Dr. Elden Rowland for his help in mass spectrometry-based analyses, and Dr. Dhwani Desai for his help in metagenomic analyses. I would also like to thank to Cat Bannon, Loay Jabre, and Scott McCain for their help in different aspects that were essential to finish this thesis. Lastly, I would like to thank to my family for supporting me in everything I want to do.

Chapter 1: Introduction

Primary production in the ocean is an important process performed by phytoplankton that influences the carbon cycle through the fixation of atmospheric carbon dioxide (CO_2) into organic matter. Controls of phytoplankton growth have been the interest of researchers for decades given that phytoplankton are responsible for about 50% of the global net primary production (Field *et al*., 1998). Nutrient availability is one of the requirements for phytoplankton growth. Results from bottle incubation nutrient addition experiments with environmental samples from different regions have enabled the identification of nutrients that limit primary productivity in the ocean. Cobalamin (vitamin B_{12}) has been identified as a compound that can co-limit or secondarily limit primary productivity, as well as influence the composition of the microbial community, in certain regions like the Antarctic Peninsula of the Southern Ocean (Panzeca *et al*., 2006), Long island embayments (Sañudo-Wilhelmy *et al*., 2006; Gobler *et al*., 2007), the Ross Sea (Bertrand *et al*., 2007), HNLC areas of the Gulf of Alaska (Koch *et al*., 2011), the coastal McMurdo Sound (Bertrand *et al*., 2015), the Southeast Atlantic gyre (Browning *et al*., 2017), and the Northeast Atlantic (Barber-Lluch *et al*., 2019; Joglar *et al*., 2020).

Cobalamin is a complex compound consisting of a cobalt-containing corrin ring, a β-ligand and 5,6-dimethylbenzimidazole (DMB) as α-ligand. It is required by many prokaryotes and approximately half of surveyed species of eukaryotic phytoplankton (Croft *et al*., 2005; Tang *et al*., 2010) as a cofactor for essential enzymes like methionine synthase (MetH) and methylmalonyl CoA mutase (MCM) (Martens *et al*., 2002). Cobalamin biosynthesis, however, can be only accomplished by a limited number of bacteria and archaea, possibly due to the metabolic burden imposed by the more than 20 enzymes

required for its *de novo* synthesis (Rodionov *et al*., 2003; Shelton *et al*., 2019). The synthesis of cobalamin through a remodeling pathway is also possible. Remodeler organisms encode enzymes involved in the final steps of cobalamin synthesis, but do not have genes involved in the corrin ring synthesis. They are capable of the uptake and remodelling of products of cobalamin degradation or related compounds, such as pseudocobalamin (Helliwell *et al*., 2016; Heal *et al*., 2017; Shelton *et al*., 2019; Ma *et al*., 2020). Pseudocobalamin is produced and used by the vast majority of marine cyanobacteria and it is poorly bioavailable for most of the eukaryotic phytoplankton since it contains adenine as α-ligand rather than the DMB found in cobalamin (Helliwell *et al*., 2016; Heal *et al*., 2017). Nevertheless, the existence of pseudocobalamin remodelers makes pseudocobalamin pools in the ocean an important factor influencing cobalamin cycling.

Given the widespread demand for cobalamin in the ocean and the fact that microorganisms are its only source, this compound has the potential to not only affect phytoplankton growth but also to shape ecological interaction within the microbial community. The factors driving cobalamin limitation in marine environments as well as the nature of the community interactions influencing the cobalamin cycle remain poorly understood. The identification of cobalamin producers and consumers is an important step towards elucidating the extent of and controls on cobalamin's influence on primary productivity.

The Scotian Shelf, a region in the Northwest Atlantic Ocean, was selected as the study site for this thesis based on its socioeconomic and biogeochemical importance. This thesis has been divided in two research chapters with the following objectives: (Chapter 2) identify major cobalamin producers and prokaryotic consumers on the Scotian Shelf and Slope; (Chapter 3) identify principal cyanobacterial (*Synechococcus*) strains and select

appropriate peptides to track their contribution to cobalamin cycling on the Scotian Shelf and Slope.

In Chapter 2 major cobalamin producers (via *de novo* and remodeling pathways) and prokaryotic consumers on the Scotian Shelf were identified through the functional and taxonomic annotation of bulk metagenomic reads combined with the analysis of genome bins. Metagenomic read assembly, binning, and taxonomic inference was performed by Dr. Dhwani Desai. Dr. Desai also conducted taxonomic and functional classification of unassembled bulk metagenomic reads using an approach designed by all authors. The analysis of these metagenomic data, recruitment of metagenomic reads for bins coverage estimation, operon mining, analysis required for the comparison of the assembly and short read-based approaches, and the generation of all figures and tables was performed by me. This chapter was mainly written by me with the assistance of Dr. Desai in the methods sections where he was involved. Dr. Erin Bertrand and Dr. Julie LaRoche guided the structure of the analysis and provided feedback and suggestions for the improvement of this manuscript.

In Chapter 3 principal *Synechococcus* strains/clades on the Scotian Shelf were identified via 16S rRNA gene sequencing and the subsequent phylogenetic placement of *Synechococcus* OTU (operational taxonomic unit) sequences onto a reference phylogenetic tree. Appropriate peptides to track the contribution of these strains to cobalamin cycling through targeted proteomic analysis were selected and tested in controlled culture experiments to identify environmental conditions that might be affecting pseudocobalamin use and synthesis by *Synechococcus* on the Scotian Shelf and Slope, and to assess changes in protein expression that accompany changes in pseudocobalamin production. *Synechococcus* OTU sequences were obtained before this research by Dr. Desai and

previously published in Zorz *et al*. (2019). The phylogenetic placement of these OTU sequences for the identification of major *Synechococcus* strains on the Scotian Shelf and Slope was performed by me. The selection of peptides to target *Synechococcus* cobalamin-associated proteins was performed by me with advice from Dr. Elden Rowland and Dr. Bertrand. Culture experiments where designed and run by me, with guidance from Dr. Bertrand. I performed the subsequent harvesting of cells to obtain protein, metabolites and particulate organic carbon samples. The extraction of metabolites and protein from culture samples, and the preparation necessary to quantify the particulate organic carbon samples, was performed by me. The quantification of metabolites and proteins was performed by Dr. Rowland based on methods he developed in consultation with me and Dr. Bertrand. The analysis of the protein and metabolites data, and the generation of all figures and tables, was performed by me. This chapter was also mainly written by me with the assistance of Dr. Rowland in the methods sections where he was involved. Dr. Bertrand advised on experimental design and analysis and provided feedback and suggestions for the improvement of this chapter.

Chapter 2: Cobalamin producers and prokaryotic consumers in the Northwest Atlantic

2.1 Abstract

Cobalamin availability can influence primary productivity and ecological interactions in marine microbial communities. The characterization of cobalamin sources and sinks is a first step in investigating cobalamin dynamics and its importance for controlling productivity. This study identified major cobalamin sources and sinks on the Scotian Shelf and Slope, a region with biogeochemical and socioeconomic importance in the Northwest Atlantic Ocean. Functional and taxonomic annotation of bulk metagenomic reads, combined with analysis of genome bins, was used to determine which groups serve as cobalamin sources and sinks. Cobalamin synthesis potential was mainly attributed to Rhodobacteraceae, Thaumarchaeota, and cyanobacteria. Cobalamin remodeling potential was mainly attributed to Gammaproteobacteria, Alphaproteobacteria, and Verrucomicrobia, while potential cobalamin consumers include Flavobacteriia and Gammaproteobacteria. The use of these complementary approaches identified taxa involved in cobalamin cycling on the Scotian Shelf and revealed genomic information required for their further characterization. The Cob operon of Rhodobacterales bacterium HTCC2255, a strain with known importance in cobalamin cycling, was similar to a major cobalamin producer bin, suggesting that a related strain may represent a critical cobalamin source in this region. These results enable future inquiries that will enhance our understanding of how cobalamin shapes microbial interdependencies and productivity in this region.

2.2 Introduction

Cobalamin (vitamin B_{12}) is required by many bacteria, archaea, and approximately half of known species of eukaryotic phytoplankton in the ocean (Croft *et al*., 2005; Tang *et al*., 2010). It is used as a cofactor for essential enzymes like methionine synthase (MetH) and methylmalonyl CoA mutase (MCM) (Martens *et al*., 2002). Despite its importance and the widespread demand for this vitamin, its synthesis can only be accomplished by a limited number of bacteria and archaea (Rodionov *et al*., 2003; Shelton *et al*., 2019), making its availability critical for supporting productivity and shaping community interactions. Cobalamin is composed of a cobalt-containing corrin ring, a β-ligand and 5,6-dimethylbenzimidazole (DMB) as α-ligand. The limited number of cobalamin producers may be explained by the complexity of this structure and the fact that cobalamin *de novo* synthesis is a metabolically expensive process that requires more than 20 enzymes (Rodionov *et al*., 2003; Shelton *et al*., 2019). Cobalamin can also be produced without the machinery for the corrin ring synthesis by a subset of bacteria that can remodel degraded cobalamin or related compounds, such as pseudocobalamin, by encoding enzymes involved in the final steps of cobalamin synthesis, and/or salvage and repair, along with genes encoding proteins required for uptake of these related compounds (Figure 2.1) (Helliwell *et al*., 2016; Heal *et al*., 2017; Shelton *et al*., 2019; Ma *et al*., 2020). Given that microorganisms are the only source of cobalamin in the ocean, it is clear that it has the potential to shape ecological interaction between microbial groups. For example, culture experiments under cobalamin limitation have shown that specific mutualistic relationships can be developed between bacteria and eukaryotic algae, in which there is an interchange of organic matter for cobalamin (Croft *et al*., 2005; Kazamia *et al*., 2012; Grant *et al*., 2014;

Cruz-López and Maske, 2016). However, these specific and close interactions are not fully understood, and it remains unclear how quantitatively important they are in the ocean.

Bottle incubation nutrient addition experiments provide evidence that cobalamin availability influences phytoplankton growth in several regions of the ocean. In the Antarctic Peninsula of the Southern Ocean (Panzeca *et al*., 2006), the Ross Sea (Bertrand *et al*., 2007), HNLC (High Nutrient, Low Chlorophyll) areas of the Gulf of Alaska (Koch *et al*., 2011), and the coastal McMurdo Sound (Bertrand *et al*., 2015), the addition of cobalamin and iron to field samples resulted in the enhancement of phytoplankton biomass and changes in the microbial community, which suggest iron and cobalamin co-limitation of phytoplankton growth. Recent experiments have also begun to elucidate the role of cobalamin in the Atlantic. Cobalamin was found to secondarily limit phytoplankton growth in some regions of the Southeast Atlantic gyre when iron and nitrate limitation was alleviated (Browning *et al*., 2017). Additionally, in an upwelling system on the northeast Atlantic Ocean the addition of cobalamin either stimulated or negatively impacted the eukaryotic phytoplankton and prokaryote biomass depending on the structure of the microbial community encountered in a particular season and depth (Barber-Lluch *et al*., 2019; Joglar *et al*., 2020), providing more evidence of a link between cobalamin availability, community composition and ecological interactions.

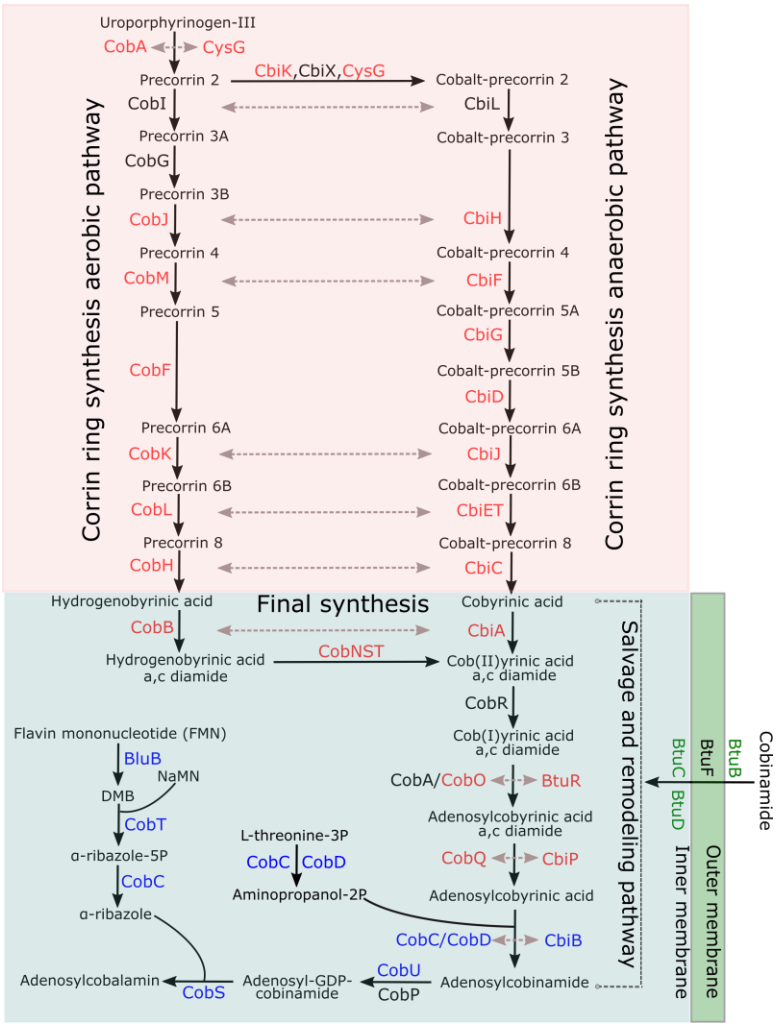

Figure 2.1. Cobalamin biosynthesis pathway in bacteria and archaea. Synthesis via *de novo* (aerobic or anaerobic corrin ring synthesis) and salvage and remodel pathways are depicted. Color denotates proteins used to identify cobalamin producers and remodelers from Scotian Shelf metagenomic samples (Table S2.1 in Appendix A). Assembled bins were considered to be producers when genes encoding more than 60% of proteins in red (involved corrin ring and final steps synthesis) were identified. Assembled bins were considered to be remodelers when genes encoding proteins in blue (involved in DMB synthesis and nucleotide loop assembly) were identified, but not those in red. Genes encoding proteins in green (involved in cobinamide uptake) can be found in bins for remodelers and consumers. Protein sequence homology is represented by horizontal double-headed arrows. Adapted from Lu *et al*. (2020), Fang *et al*. (2017) and Rodionov *et al*. (2003).

Despite the increase in numbers of studies trying to determine the role of cobalamin in limiting phytoplankton growth, information about cobalamin sources and sinks on the Northwest Atlantic remains limited. We believe that the identification and characterization of these organisms is an important step for the understanding of cobalamin's potential for

shaping the ecological interactions and composition of the microbial community, which ultimately can influence primary productivity. The Scotian Shelf, located in the Northwest Atlantic Ocean, is a region with significant biogeochemical and socioeconomic roles. Primary productivity here is categorized as moderately high (150-300 $gCm^{-2}yr^{-1}$) (Aquarone and Adams, 2009). This region contributes to the maintenance of important commercial fisheries (MacLean *et al*., 2013) and plays a key role on the carbon cycle (Shadwick and Thomas, 2014; Craig *et al*., 2015). It is oceanographically complex, being influenced by the southward flow of cold fresh water from the Nova Scotia Current (originated in the Gulf of St. Lawrence) and Shelf Break Current (extension of the Labrador Current), and the northward flow of warm salty water from the Gulf Stream (Loder *et al*., 1997; Hannah *et al*., 2001) (Figure 2.2). These currents set a trend of increasing temperature and salinity to the southwest, which can influence microbial biodiversity across the shelf (Zorz *et al*., 2019). Despite its importance, cobalamin sources and sinks remain uncharacterized in this region, and its role in controlling phytoplankton growth is just beginning to be explored.

Possible cobalamin sources on the Scotian Shelf and Slope include Alpha and Gammaproteobacteria, since they comprise a large percentage of the microbial community of the Scotian Shelf based on 16S rRNA genes amplicon analyses (Zorz *et al*., 2019), and many known cobalamin producers are from these groups (Sañudo-Wilhelmy *et al*., 2014; Bertrand *et al*., 2015; Doxey *et al*., 2015; Shelton *et al*., 2019). Thaumarcheota may also be a significant contributor, since they have been identified as important cobalamin producers in regions with cold and deep water columns like the Arctic and North Atlantic ocean (Doxey *et al*., 2015). The cyanobacterial groups *Synechococcus* and *Prochlorococcus* were found to be abundant during the fall (Zorz *et al*., 2019), suggesting

that they may be important sources of pseudocobalamin. Possible cobalamin sinks in this region include eukaryotic phytoplankton and Bacteroidetes, groups that are widely known as cobalamin consumers (Croft *et al*., 2005; Sañudo-Wilhelmy *et al*., 2014; Shelton *et al*., 2019). The class Flavobacteriia (which belongs to the Bacteriodetes phylum) is an important component of the microbial community of the Scotian Shelf, especially in spring (Zorz *et al*., 2019), while diatoms and dinoflagellates were found to dominate the eukaryotic phytoplankton community in spring and fall, respectively (Li *et al*., 2006; Dasilva *et al*., 2014). However, the metabolic capacities of these microbial groups present on the Scotian Shelf and Slope remain poorly characterized, and their role in the cobalamin cycle needs to be further explored to understand how this vitamin might be affecting ecological interactions, community composition, and productivity.

The role of uncultured microorganism in different nutrient cycles can be examined by sequencing and analyzing metagenomic samples to taxonomically and functionally characterize the microbial community in a specific environment. Previous studies have used profile hidden Markov models (HMMs) to identify major players in the cobalamin synthesis in aquatic and soil systems by searching for homologs of genes implicated in the cobalamin biosynthesis pathway against sequences from unassembled metagenomes (Doxey *et al*., 2015; Lu *et al*., 2020). These studies have revealed interesting patterns about cobalamin producers and have increased our knowledge about the ecology of these microbial groups. However, a major drawback of short-read-based analyses is the annotation accuracy, which can lead to misleading conclusions due to the presence of false positives (Temperton and Giovannoni, 2012), and its inability to associate specific functional genes with other relevant genomic context that could allow a more precise characterization. The assembly and binning of metagenomic reads (usually based on

sequence similarity and coverage across samples) to recover metagenome-assembled genomes (MAGs) from environmental samples has been implemented as a way to improve the taxonomic and functional annotation of taxa as well as to facilitate downstream analyses like operon mining, more complex functional predictions, and phylogenetic placement. Nevertheless, this approach can be computationally expensive and biased towards abundant organisms (Temperton and Giovannoni *et al*., 2012; Thomas *et al*., 2012; Ayling *et al*., 2019). Furthermore, the presence of strain variants with similar sample coverage makes the assembly and binning of some taxa challenging since those reads might get either confused as sequencing errors and removed from the analysis or can be assembled into chimeric contigs (Temperton and Giovannoni *et al*., 2012; Ayling *et al*., 2019). Facing the drawbacks of both short-read-based and MAGs analyses, a way to get a comprehensive view of the metabolic capacities of the organisms of interest is to use both approaches since they offer complementary strengths.

Further characterization of major sources and sinks of cobalamin is needed for elucidating the extent of its influence on primary productivity and its role shaping microbial interdependencies in the ocean. This study focuses on the identification of prokaryotic cobalamin producers and consumers on the Scotian Shelf as a path towards understanding the role of heterotrophic bacteria and archaea in cobalamin cycling. Sequencing and analyzing metagenomes from Scotian Shelf and Slope samples through two complementary approaches, consisting of the assembly of short metagenomic reads and a short-read-based analysis, enables identification and quantification of major cobalamin producers and prokaryotic consumers in this region.

2.3 Methods

2.3.1 *Sampling and metagenomes sequencing*

Water samples from different stations and depths of the Scotian Shelf were obtained during the 2016 fall and spring AZMP (Atlantic Zone Monitoring Program) cruises. ~4 L of water were collected at each station in Niskin bottles using a CTD Rosette and filtered through two consecutive polycarbonate filters (3.0 µm and 0.2 µm) to isolate microbial cells for DNA extraction. Samples were stored at -80˚C until their analysis. 42 representative DNA samples (0.2 µm size fraction) from 9 stations were selected for metagenomic sequencing and analysis (Figure 2.2A; Table S2.2 in Appendix A). DNA extraction was performed as described elsewhere (Zorz *et al*., 2019) and the samples were then sent to the Integrated Microbiome Resource (IMR) (https://imr.bio) for metagenomic sequencing via NextSeq (4X depth, 150+150 bp PE).

CTD observations from each cruise were provided by the Bedford Institute of Oceanography (BIO). These datasets were used to generate Chlorophyll-a depth profiles (proxy for photosynthetic algae abundance in the water column) and to determine the mixed layer depth (MLD) for each station where DNA samples were taken (Figure 2.2B). The MLD was determined as the depth at which a change of 0.03 kg m^{-3} was observed relative to the density value at 10 m (de Boyer Montégut *et al*., 2014; see Figure S2.1 in Appendix A for density depth profile). The chlorophyll-a depth profile and MLD were used to estimate the photic zone depth and categorize the depth at which each sample was taken as follows: samples above the MLD and the chlorophyll-a signal were deemed to be within the 'surface euphotic zone', samples below the MLD but above the loss of the chlorophyll-a signal were at the 'euphotic zone', and samples below the MLD and depth at which the chlorophyll-a signal could not be resolved were at the 'deep zone'. Additionally, samples

were grouped by season (fall or spring) and geographical location relative to the shelf break (coast or offshore).

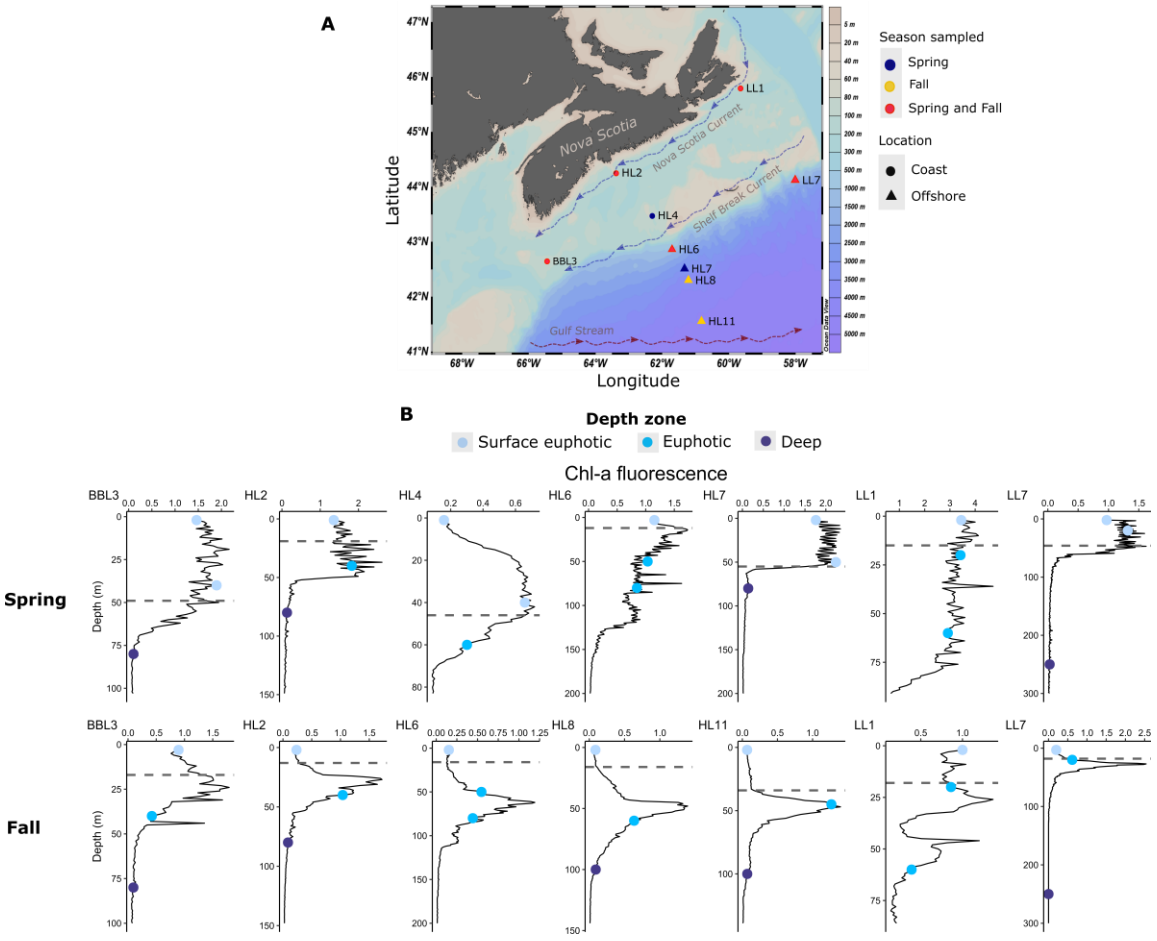

Figure 2.2. (A) Location of Scotian Shelf and Slope stations sampled in 2016 during fall and spring for metagenomic sequencing. Bottom depth and direction of the main currents affecting the region are also displayed. Blue arrows represent cool water flow and red arrows represents warm water flow. **(B)** Chlorophyll-a depth profile for each station used to classify each sample. Dots represent depth at which the sample was taken, and colour denotes classification. The grey dashed line represents the MLD.

2.3.2 *Identification of major cobalamin producers, remodelers, and consumers on the Scotian Shelf and Slope*

The identification of major players in the cobalamin cycle on the Scotian Shelf was done using two different approaches:

2.3.2.1 Assembly-based analyses

Non-redundant bins were analysed for presence/absence of genes involved in cobalamin synthesis, remodeling, utilization and uptake (Table S2.1 in Appendix A) to annotate each bin according to its functional potential as cobalamin producer, remodeler or consumer. Major cobalamin-associated bins on the Scotian Shelf were determined by analyzing the percent of reads functionally annotated as cobalamin-related that were mapped to each bin, per sample, as described below.

2.3.2.1.1 Metagenomic reads assembly, binning, and taxonomic inference

Metagenomic read assembly, binning and taxonomic inference was done following the pipeline described in http://merenlab.org/data/tara-oceans-mags/ using the Anvi'o platform (Eren *et al*., 2015; Delmont and Eren, 2018). Reads from Scotian Shelf and Slope metagenomic samples were grouped in sets based on similar location (coastal or offshore) and season (spring and fall). Bulk metagenomic reads from each group of samples were quality filtered and pooled into a co-assembly using MEGAHIT (Li *et al*., 2015). Anvi'o was used to generate a contigs database from the assembled contigs. This contigs database contains information about k-mer frequencies, identifies open reading frames using Prodigal (Hyatt *et al*., 2010), and enables the functional annotation of genes using the Clusters of Orthologous Groups of proteins (COGs) database (Tatusov *et al*., 2000). Contigs were then binned using CONCOCT, manually curated, and dereplicated using dRep (Alneberg *et al*., 2014; Olm *et al*., 2017). The taxonomic inference of the dereplicated bins was made using the Single Copy Core gene (SCG) analysis in Anvi'o which utilizes the Genome Taxonomy Database (GTDB) for bacterial taxonomy (Parks *et al*., 2018). The SCG analysis also provided estimation of the percent completion and percent redundancy for each bin based on presence/absence of single-copy genes and multiple occurrences of

single-copy genes in a bin, respectively (Eren *et al.*, 2015). Bins with redundancy ≤ 10% and completion ≥ 70% were annotated as MAGs (Table S2.3 in Appendix A).

2.3.2.1.2 Recruitment of metagenomic reads for bins coverage estimation

Quality filtered metagenomic reads annotated as cobalamin-related genes (see below) from each sample were mapped against the non-redundant collection of bins concatenated in a single FASTA file using Bowtie2 (Langmead and Salzberg, 2012). The resultant SAM files were then converted to BAM files using samtools (Li *et al*., 2009). Anvi'o was used to generate a contigs database from the same FASTA file and a profile database from each BAM file which were then merged in a single profile database as described in http://merenlab.org/2016/06/22/anvio-tutorial-v2/ (Eren *et al*., 2015; Delmont and Eren, 2018). This database contains sample-specific information about contigs and enables the coverage estimation of each bin across all the samples included in the merged profile database. The percent of cobalamin-related reads mapped to each bin per sample was obtained from the "Percent recruitment" output given by Anvi'o. To normalize the data, this value was divided by the total length (nucleic acids) of the cobalamin-related genes found in that bin.

2.3.2.1.3 Operon mining and genetic comparison with known cobalamin producers

To further investigate the genetic potential for cobalamin synthesis of representative bins assigned as cobalamin producer, these bins were uploaded to the Rast web-tool (https://rast.nmpdr.org/rast.cgi) for analysis and comparison of the Cob/Cbi operon arrangement with organisms with similar sequences that might have been previously identified as cobalamin producers. Organism with similar Cob/Cbi gene sequences were

found by using BLAST and the sequences of the Cob/Cbi genes of the cobalamin producer bin as query.

Whole-genome similarities between two organisms with similar Cob/Cbi operon arrangement was also assessed by estimating their average nucleotide identity (ANI) using FastANI (Jain *et al*., 2018).

2.3.2.2 Short read analysis

Recognizing that rare taxa or taxa represented by closely related strains might not get properly assembled into bins, the taxonomic assignment and functional annotation of bulk metagenomic reads was done in an attempt to account for other taxa that may be important contributors to the cobalamin cycle on the Scotian Shelf and Slope.

2.3.2.2.1 Taxonomic and functional classification of unassembled bulk metagenomic reads

The quality-controlled bulk metagenomic reads were annotated for taxonomic affiliations as well as functions using an in-house pipeline. Briefly, the reads were mapped to the Refseq Complete genomes database (O'Leary *et al*., 2016) using the program Kraken2 (Wood *et al*., 2019) using a confidence score threshold of 0.2. The reads were also mapped to the Refseq protein database using the program mmseqs2 (Steinegger and Soding, 2017). The counts for each function were normalized by the gene length of Refseq hit and the total number of reads mapping to any function to calculate the Reads Per Kilobase per Million (RPKM). The RPKM for each function was further split into contributions by various taxonomic categories to generate a stratified RPKM table using in-house python scripts.

2.3.2.2.2 Comparison of assembly and short read-based approaches

Reads that failed to map to the bins were extracted from the BAM files generated for each sample using samtools (Li *et al*., 2009). To determine how representative the identified bins are of the major taxonomic groups involved in the cobalamin cycle on the Scotian Shelf, the total number of cobalamin-assigned reads and the number of unmapped cobalamin-assigned reads per sample were counted to calculate the percent of cobalamin-assigned reads that were/were not mapped to the bins. To get information about the taxa that were not represented in the bin-based analysis but that might have the potential to be major contributors to the cobalamin cycle, the ID of the reads that failed to map to the bins was associated to their taxonomic annotation from the bulk metagenomic reads analysis.

2.3.3 *Data visualization*

Metagenomic and CTD data visualizations were generated using R (R Core Team, 2020) with the *ggplot2* package (Wickham, 2014). The contrast on Figure 2.3 was increased to facilitate the visualization of the data. The Scotian Shelf and Slope map was generated using Ocean Data View (V. 5.3.0). Some elements on these figures were added using Inkscape (V. 1.0) to highlight trends or add important data. Inkscape was also used to create Figure 2.1 and Figure S2.3.

2.4 Results and discussion

2.4.1 *Water column properties*

In general, fall samples were characterized by a shallower MLD compared to spring samples (Figure 2.2B), likely due to the higher surface water temperatures observed in fall (Figure S2.2 in Appendix A) that increase stratification of the water column and reduce the wind-driven mixing of surface and deep waters. The water temperature of the stations

closer to the Gulf Stream was higher compared to other stations due the input of warm water, as expected (Figure 2.2A, Figure S2.2 in Appendix A). The water temperature of the spring samples taken at the deep zone was higher compared to the samples at the surface euphotic and euphotic zones. The opposite was observed for the fall samples, where temperature decreased with depth (Figure S2.2 in Appendix A). Spring on the Scotian Shelf is characterized by nutrient-rich upper mixed waters whereas fall upper mixed waters can be described as oligotrophic; although, similar nutrient concentrations can be found below these waters in spring and fall (Dasilva *et al*., 2014; Zorz *et al*., 2019). A decrease in phosphate concentrations have been also described in offshore waters compared to coast waters (Zorz *et al*., 2019). Based on these observations, samples were divided into environmental, geographic and seasonal groups: spring and fall, coastal and offshore, surface euphotic, euphotic, and deep samples for subsequent analyses, as described above.

2.4.2 *Major cobalamin producers, remodelers, and consumers on the Scotian Shelf and Slope*

93% of bins recovered were found to be involved in the cobalamin cycle. 69% of these bins were classified as cobalamin consumers, 17% as remodelers, and 14% as producers. Bins with the highest percentage of cobalamin-assigned reads mapped were Gammaproteobacteria, Alphaproteobacteria, Betaproteobacteria, Flavobacteriia, Actinobacteria, Archaea and Cyanobacteria bins (Figure 2.3). Notably, all archaeal bins recovered were Thermoplasmatota bins classified as cobalamin consumers. Therefore, contrary to what was expected, this bin-based analysis suggests that Thaumarchaeota are not major cobalamin producers in this region. Furthermore, the number of cyanobacterial bins recovered as well as the percent of cobalamin-assigned reads mapped to them (Figure 2.3) was lower than expected given the high abundance and distribution of *Synechococcus*

and *Prochlorococcus* across the Scotian Shelf and Slope (Zorz *et al*., 2019). The assembly of closely related genomes with similar sample coverage can be very challenging (Temperton and Giovannoni *et al*., 2012; Ayling *et al*., 2019), leading to an underestimation of the abundance of these taxa in metagenomic samples. This might have been the case for cyanobacteria on the Scotian Shelf since *Synechococcus* and *Prochlorococcus* are comprised of closely related strains clustered in clades that can often be found occupying the same environment (Rocap *et al*., 2002).

To further interrogate these bin-based identifications of the major players in the cobalamin cycling on the Scotian Shelf, the percent of bulk reads functionally annotated as cobalamin related genes that mapped to the bins was calculated. For all locations, less that the 50% of these reads were mapped to bins (Figure 2.4A). Furthermore, the taxonomic classification of the cobalamin related reads that were not mapped to bins revealed that specific taxonomic groups were underrepresented by the bin-based analysis (discussed below, Figure 2.4B), suggesting that this analysis only captured a subset of the organisms involved in cobalamin cycling on the Scotian Shelf and Slope. Therefore, the bin-based analysis was complemented by performing the taxonomic and functional classification of unassembled bulk metagenomic reads (see Figure 2.5) to characterize the prokaryotic microbial community involved in the cobalamin cycle. Although the taxonomic classification for reads functionally annotated as cobalamin related was limited to only 9% (on average) of these reads in all locations (Table S2.4 in Appendix A), this complementary analysis allowed the identification of additional players in the cobalamin cycle, as discussed below.

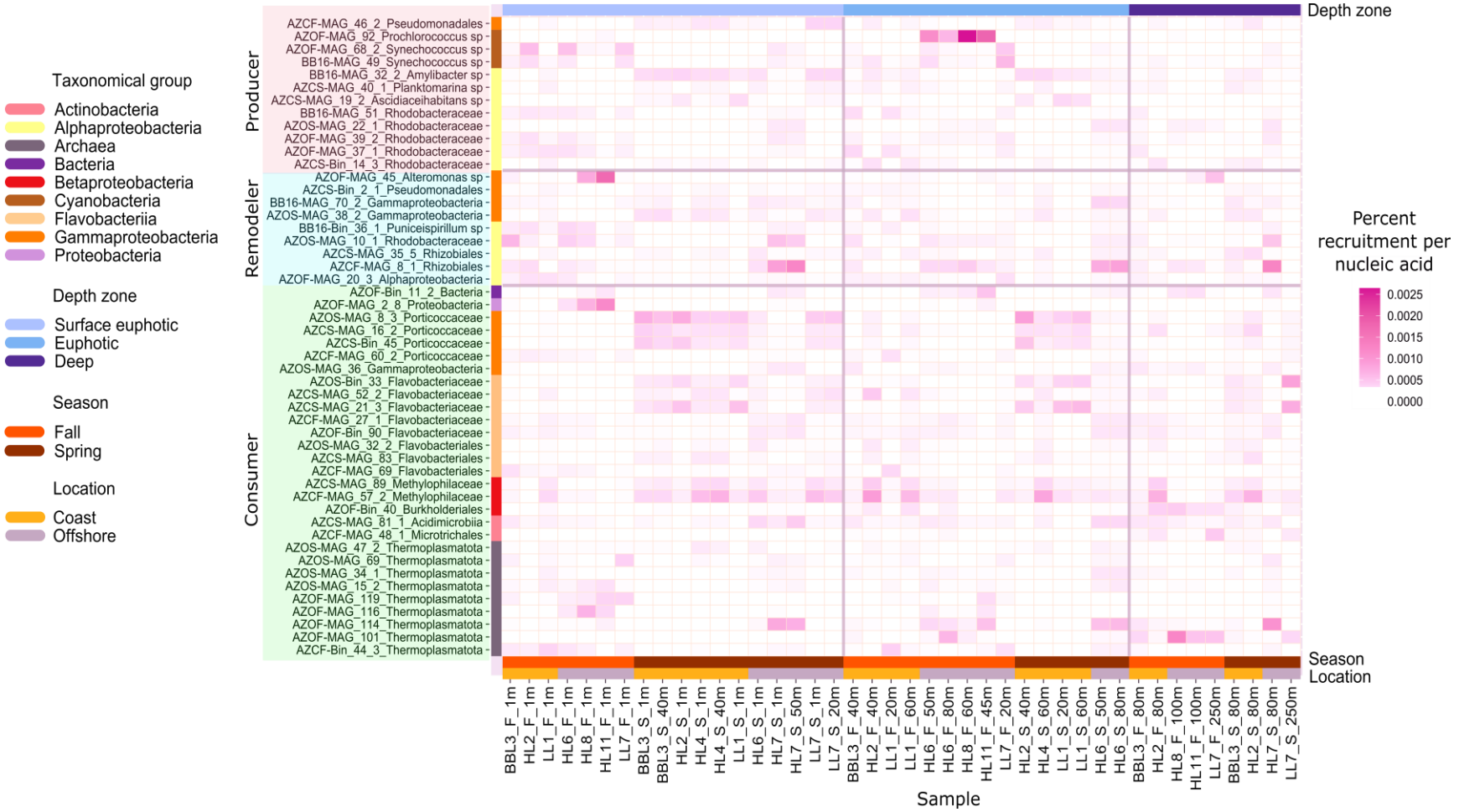

Figure 2.3. Percent of cobalamin-assigned metagenomic reads, normalized to the summed length of cobalamin-associated genes in each bin, mapped to those bins annotated as cobalamin producers, remodelers and consumers across Scotian Shelf and Slope samples. Samples have been grouped by depth zone, season and location. Only the top 50 most dominant bins are displayed. For a complete list of bins recovered, along with the cobalamin related genes found on their genomes, see Table S2.3 in Appendix A.

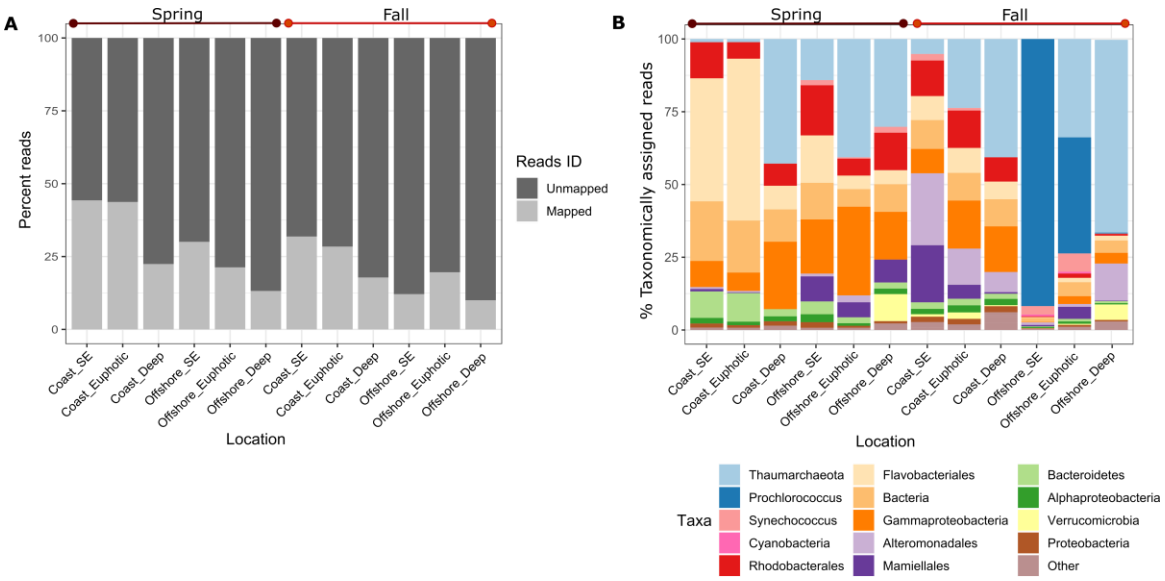

Figure 2.4. (A) Percent of bulk metagenomic reads functionally assigned as cobalamin related genes that did or did not map to the identified bins and **(B)** the taxonomic annotation of unmapped reads. Reads have been grouped by the location and season where metagenomic samples were taken. SE: Surface euphotic.

2.4.2.1 Cobalamin consumers

Bins classified as cobalamin consumers have the potential to express genes involved in cobalamin uptake and use, but not cobalamin synthesis genes. Consumer bins with high percentages of cobalamin-assigned reads mapped included Gammaproteobacteria, Flavobacteriales, Methylophiliaceae, Actinobacteria, and Thermoplasmatota bins (Figure 2.3). Given that the majority of bacteria (either cobalamin consumer, remodeler or producer) have at least one cobalamin dependent enzyme (Shelton *et al.*, 2019), the contributors to the pool of cobalamin use and uptake genes on the Scotian Shelf are not only cobalamin consumers (taxa with no potential for cobalamin synthesis), but will also include other groups (Figure 2.5). Therefore, tentatively classifying taxa as cobalamin consumers through the short read analysis requires the identification of taxonomic groups that only contributed to the pool of cobalamin uptake and utilization genes and not to cobalamin synthesis. This approach is limited, however, to coarse taxonomic resolution

relative to the bin-based analyses. However, taxonomic groups with only cobalamin use and uptake potential identified through this coarse short read-based analysis were Flavobacteriales and Bacteroidetes, which were also both identified as important consumers through the bin-based analysis. This approach notably misses other important consumers due to the lower taxonomic resolution, as expected. In addition, none of the cobalamin-associated reads were taxonomically assigned to Thermoplasmatota (Figure 2.5). Thermoplasmatota reads could have been part of the group of reads that could not be assigned to specific taxa (Table S2.4 in Appendix A), which would explain the lack of cobalamin-associated reads assigned to them.

General trends show a similar spatio-temporal distribution between bacterial consumer bins and taxa that contributed only to the pool of cobalamin-dependent genes. The number of reads assigned to these taxa as well as the percent of cobalamin-related reads mapped to bacterial consumer bins in coastal-spring samples at the surface euphotic zone are higher compared to the fall (coast and offshore) samples at the same depth zone (Figure 2.3 and Figure 2.5). This trend might be responding to a lower availability of inorganic nutrients (and possibly cobalamin) due to the shallower MLD that limits nutrient supply from deeper waters. Alternatively, predation or competition pressures can be occurring due to the shift towards a eukaryotic microbial community dominated by Dinoflagellates (mostly mixotrophs) during this time (Li *et al*., 2006; Dasilva *et al*., 2014). An increase in the percent of reads mapped to consumers can be observed during the fall (coast and offshore) at the euphotic and deep zones compared with the surface euphotic zone, possibly reflecting an increase in nutrient availability below the MLD (Figure 2.3). Contrasting results can be observed for the Thermoplasmatota bins, which present a higher percent of cobalamin-assigned reads mapped during the fall and spring-offshore samples

at all depth zones (Figure 2.3). Thermoplasmatota is an archaeal phylum recently described by Rinke *et al*. (2019) that include the also newly described order *Candidatus* Poseidoniales (Marine Group II archaea). It has been suggested that *Ca.* Poseidoniales may be abundant archaea in surface waters, though they can also be present in deep waters. This order is also comprised by both genera adapted to oligotrophic and nutrient-rich environments (Rinke *et al*., 2019). Given the spatio-temporal distribution of the Thermoplasmatota bins recovered (Figure 2.3), this suggests that different genera may be inhabiting during spring and fall due to the differences in nutrient availability, and possibly in surface and deep waters in the Scotian Shelf and Slope.

2.4.2.2 Cobalamin remodelers

Remodelers can express genes involved in the lower ligand (DMB) synthesis and nucleotide loop assembly (conversion of adenosylcobyrinic acid to adenosylcobalamin), but not complete corrin ring synthesis, which allow them to synthesize cobalamin through the uptake of cobalamin precursors (cobinamide) or modification of a cobalamin analog (pseudocobalamin), but not through *de novo* synthesis. The modification of pseudocobalamin is done by the interchange of its lower ligand (adenine) by DMB available in the water column or synthesized by the remodeler (Escalante-Semerena, 2007; Helliwell *et al*., 2016; Heal *et al*., 2017). Bins annotated as cobalamin remodelers with the highest percent of cobalamin-assigned reads mapped include a diversity of Alphaproteobacterial as well as Gammaproteobacterial bins (Figure 2.3). The entire set of genes implicated in the salvage and remodeling pathway could not be found in any of these bins, perhaps due to assembly or annotation errors of those genes in certain taxa. However, the majority of the remodeler bins contained CobT and BtuB in their genomes, indicating potential for DMB activation and cobinamide uptake, respectively. CobT has been found

to be essential for the incorporation of external DMB (Crofts *et al*., 2013; Helliwell *et al*.,2016). Therefore, if DMB is available in the water column, the presence of CobT and BtuB in remodeler bins may indicate the ability to perform cobalamin salvage and remodeling. Cobalamin-assigned reads predominately mapped to remodeler bins in fall and spring-offshore samples at the photic zones (Figure 2.3), where the contribution of cyanobacteria (pseudocobalamin sources) to the pool of corrin ring synthesis genes is likely high (Figure 2.5). This suggests that pseudocobalamin remodeling might be an important mechanism used to obtain cobalamin by remodeler organisms. Heterotrophic sources of cobalamin are also present at these same locations (Figure 2.3 and Figure 2.5); therefore, it is possible that these remodelers can also uptake cobalamin from readily available dissolved pools or use cobalamin precursors or products of its degradation to synthesize cobalamin. A similar scenario can be described for the deep zone, where pseudocobalamin sources do not contribute to the pool of cobalamin synthesis genes (Figure 2.5). However, the prevalence and importance of pseudocobalamin/cobinamide remodeling on the Scotian Shelf and Slope needs to be further explored. Data characterizing the concentrations of particulate and dissolve cobalamin, pseudocobalamin and DMB through the water column as well as studies describing the expression patterns of genes involved in cobalamin remodeling, may help to clarify the role of remodelers in the cobalamin cycle on this region.

The identification of pseudocobalamin/cobinamide remodelers through the short read analysis is challenging since the majority of reads annotated as genes implicated in the final synthesis, salvage and repair of cobalamin as well as genes implicated in the DMB synthesis, were assigned to taxa that also possess corrin ring synthesis capacity. Not being able to tell whether metagenomic reads came from the same organism (genome) is an important limitation in this case, and the low taxonomic resolution of the short read analysis

prevents differentiation between members of the same taxa with all the genes necessary to synthesize cobalamin and the ones with only final synthesis, salvage and repair genes. Nevertheless, taxonomic groups with pseudocobalamin/cobinamide remodeling capacity and without corrin ring synthesis capacity, according to the short read analysis, included select Gammaproteobacteria (Alteromonadales, Oceanospirilalles, Vibrionales) across all samples and also Verrucomicrobia in offshore-deep samples, suggesting that these groups may contain important cobalamin remodelers that were not all identified in bin-based analyses (Figure 2.5).

2.4.2.3 Cobalamin producers

Bins annotated as cobalamin producers with the highest percent of cobalamin-associated reads mapped were Rhodobacteraceae (including *Ascidiaceihabitans sp*., *Planktomarina sp*., and *Amylibacter sp*.) and cyanobacteria bins. A lower percent of reads also mapped to Gammaproteobacteria bins (Figure 2.3). Although Gammaproteobacteria were predicted as major *de novo* cobalamin producers on the Scotian Shelf, they were mainly classified as cobalamin consumers and remodelers as well, as described above. Alphaproteobacteria (Rhodobacteraceae) are indeed important taxa with cobalamin synthesis capacity in this region, especially in spring and fall-coast samples at surface euphotic and euphotic zones, as shown by both the percent of cobalamin-assigned reads mapped to these bins and the contribution of Rhodobacterales to the pool of cobalamin synthesis genes at these locations (Figure 2.3 and Figure 2.5). Unlike the bin-based analysis, the taxonomic assignment of unmapped cobalamin-assigned reads highlights the presence and dominance of Thaumarchaeota at the deep and euphotic depth zones, where we were expecting this taxon to be present on the Scotian Shelf and Slope (Figure 2.4B). Thaumarchaeota are also a major contributor to the cobalamin synthesis genes below the

surface euphotic zone (Figure 2.5), which makes them a potential source of cobalamin in sub-surface and deep waters. Short read analysis also confirms that cyanobacteria where underrepresented on the bin-based analysis. A large percentage of the reads that could not mapped to bins in the fall-offshore samples at the surface euphotic and euphotic zones were assigned to cyanobacteria, especially *Prochlorococcus* (Figure 2.4B). The percent of reads mapped to *Synechococcus* and *Prochlorococcus* bins is higher in fall-photic samples, as expected. This is consistent with the locations where cyanobacteria are major contributors to the pool of cobalamin synthesis genes (Figure 2.3 and Figure 2.5). *Synechococcus* can also be observed in spring samples, reflecting the larger geographic and temporal distribution that this cyanobacterium can have due to the different ecotypes adapted to various environmental conditions (e.g. temperature) that have been described within this genus (Ahlgren and Rocap, 2012). Depth was another factor determining the distribution of cyanobacteria. A higher number of reads associated to cobalamin synthesis genes at the euphotic zone were assigned to *Prochlorococcus* compared to the reads assigned to *Synechococcus* (Figure 2.5). Furthermore, the percent of cobalamin-assigned reads mapped to the *Prochlorococcus* bin was the highest between 45m and 60m, while the percent of these reads mapped to *Synechococcus* was higher above 20m (Figure 2.3). The prevalence of *Prochlorococcus* at deeper waters can be explained by the presence of low-light adapted *Prochlorococcus* ecotypes (Moore *et al*., 1998; Zinser *et al*., 2007).

The assembly and binning of metagenomic reads enables the analysis of the genomes and associated metabolic potential of the bins recovered. This advantage was used to further investigate the genetic potential for cobalamin synthesis of the producer bins identified here by looking for similarities in the sequences and arrangement of their Cob/Cbi operon with cobalamin producers that have been previously characterized. The

sequence and arrangement of the Cob/Cbi gene clusters of the heterotrophic bacteria cobalamin producer bin with the highest percent of cobalamin-assigned reads mapped (MAG_32_2_*Amylibacter sp.*) have a high similarity with Rhodobacterales bacterium HTCC2255, a bacterium with known importance in cobalamin cycling (Figure S2.3 in Appendix A). A phylogenetic study, based on 31 conserved proteins, suggested that Amylibacter species and Rhodobacterales bacterium HTCC2255 are closely related (Knobloch *et al*., 2020). An average nucleotide identity (ANI) of 82% between MAG_32_2_*Amylibacter sp.* and Rhodobacterales bacterium HTCC2255 genomes confirmed that these organisms are closely related, though not the same species (Konstantinidis and Tiedje, 2005; Goris *et al.*, 2007).

Rhodobacterales bacterium HTCC2255 has been suggested to be the most abundant contributor to the Cob/Cbi genes identified in metagenomic samples across the ocean (Doxey et al., 2015). It was also one of the taxa that dominated the expression of cobalamin synthesis genes on a coastal environment (Gómez-Consarnau *et al.*, 2018). Notably, this bacterium is known to degrade 2,3-dihydroxypropane-1-sulfonate (DHPS), a metabolite released by certain diatoms that may promote bacteria-phytoplankton cross-feeding interactions that supply cobalamin to phytoplankton (Durham *et al*., 2015). The similarities between these two genomes suggests that there may be important functional similarities between these strains, and that MAG_32_2_ *Amylibacter sp.* may be a critically important cobalamin producer on the Scotian Shelf. Rhodobacterales bacterium HTCC2255 is available as a cultured isolate and thus it can be grown under controlled conditions that may allow future experiments to understand cobalamin production and the factors that affect it in the Northwest Atlantic, should efforts to isolate the organism corresponding to MAG_32_2 fail.

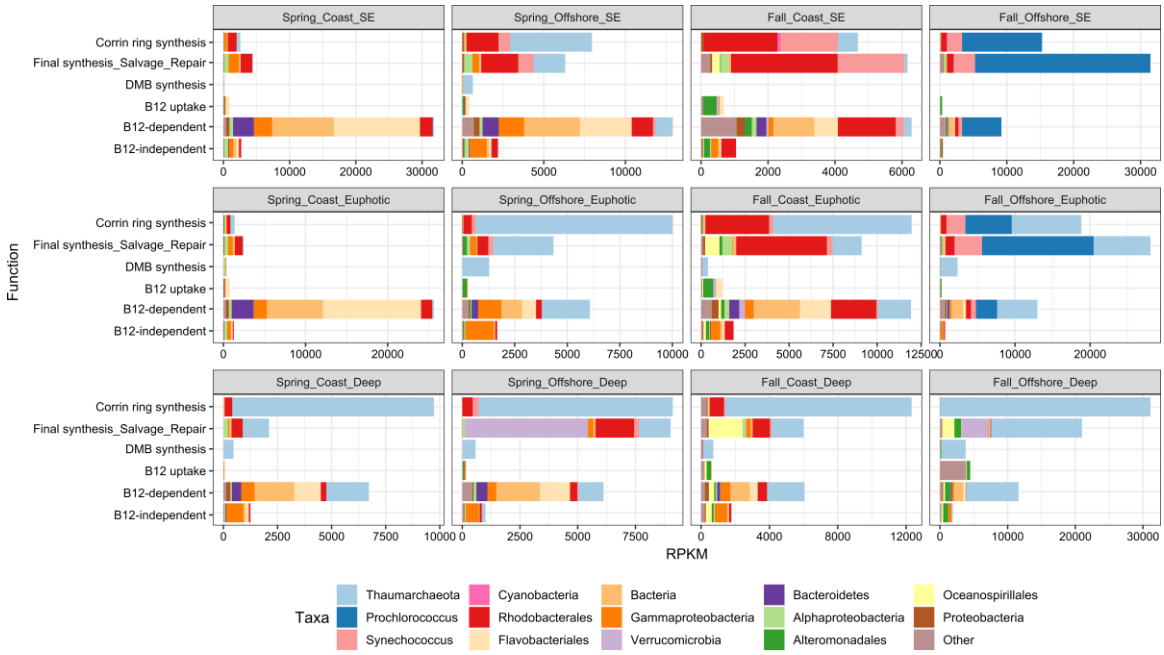

Figure 2.5. Taxonomic contributions to the pool of genes involved in the cobalamin synthesis, uptake, use and remodeling on the Scotian Shelf and Slope. Taxonomic contributions are given by the number of reads per kilo base per million reads (RPKM) assigned to each taxa per gene. SE: Surface euphotic.

2.5 Conclusion

The complementary results given by the genome-bins and bulk metagenomic read analyses allowed the characterization of organisms with the potential for being major cobalamin producers, remodelers and consumers on the Scotian Shelf and Slope, and the identification of important organisms for future studies to elucidate controls on cobalamin cycling. While the short read analysis allowed identification of taxa that could not be properly assembled, the bin-based analysis provided us with the genomic context required for confident classification as cobalamin consumers or remodelers, and also offers important information that can be used for further genomic, functional and ecological characterization through other bioinformatic analyses or laboratory experiments with similar strains. An important conclusion from this work is that relatively high levels of taxonomic resolution, and

accompanying genomic context, are required to confidently describe the cobalamin-related roles of specific groups, particularly members of Gamma and Alphaproteobacteria, since important members of these groups can be cobalamin producers, consumers or remodelers. This identification of major contributors to cobalamin cycling on the Scotian Shelf and Slope can form a base upon which to build understanding of the cobalamin dynamics in this region, enabling future analysis using metatranscriptomic, metaproteomic and metabolomic approaches. These approaches would clarify the temporal and geographic importance of the taxa identified in this study by characterizing gene expression patterns and metabolite availability in the water column of the Scotian Shelf and Slope.

Chapter 3: Mass spectrometry-based insights into *Synechococcus* contributions to cobalamin cycling in the Northwest Atlantic

3.1 Abstract

Cobalamin availability can influence phytoplankton growth and affect the structure and ecological interactions of marine microbial communities. Identifying the organisms that play a role in cobalamin cycling can improve our understanding of what factors and processes limit primary productivity in the ocean. Although most cyanobacteria produce and use pseudocobalamin, a molecule that is poorly bioavailable for most of the eukaryotic phytoplankton, they can still contribute to cobalamin cycling in the ocean due to the existence of organisms with the capacity to chemically remodel pseudocobalamin into cobalamin. *Synechococcus*, one of the most abundant cyanobacterial genera in the ocean, produce pseudocobalamin and thus have the potential to be an important contributor to the marine cobalamin cycle. However, factors that might be affecting pseudocobalamin production and quotas in *Synechococcus* are unknown. In this study, we identify and implement protein and metabolite measurements that can be used to assess the role of *Synechococcus* in cobalamin cycling in the Northwest Atlantic Ocean. Phylogenetic placement of previously published *Synechococcus* OTU sequences from Scotian Shelf and Slope environmental samples enabled the identification of clade III as the dominant *Synechococcus* clade in this region, followed by clade I and clade X. This identification guided the selection of peptides specific to proteins involved in the pseudocobalamin utilization and synthesis (MetH and CobO, respectively) that target strains from these major clades and could be quantified using selected reaction monitoring (SRM) mass spectrometry. These specific peptides, as well as pseudocobalamin, were quantified in *Synechococcus* WH 8102 (Clade III strain) cultures under different conditions. These

experiments revealed that CobO presented invariant expression through the conditions tested, whereas MetH abundance was significantly increased under low temperature and in the morning at exponential phase. Low temperature also significantly affected the growth rate, cellular carbon and protein content of *Synechococcus sp.* WH 8102. Cellular quotas of pseudocobalamin significantly decreased with low temperature and increased in stationary phase. These results revealed environmental and physiological conditions with the potential to affect the pseudocobalamin pool on the Scotian Shelf and Slope. Furthermore, this study provides tools to monitor and interpret CobO and MetH protein expression patterns, as well as and pseudocobalamin concentration, in environmental samples from this region.

3.2 Introduction

Cobalamin is an organometallic compound that can only be synthesized by a subset of bacteria and archaea but is widely required by many prokaryotes and about half of surveyed eukaryotic phytoplankton (Rodionov *et al*., 2003; Croft *et al*., 2005; Tang *et al*., 2010; Shelton *et al*., 2019). A key cobalamin-dependent enzyme is methionine synthase (MetH) which catalyzes the regeneration of methionine, a proteinogenic amino acid and key component of the essential methyl donor S-adenosylmethionine (SAM) (Fontecave *et al*., 2004; Combs, 2012). Insufficient concentrations of cobalamin in the ocean can limit the growth of auxotrophic organisms, such as eukaryotic phytoplankton, and potentially affect primary productivity and the structure of marine microbial communities. Bottle incubation nutrient addition experiments have provided evidence that cobalamin availability can influence the abundance and distribution of bacteria and phytoplankton in different oceanic regions (Panzeca *et al*., 2006; Bertrand *et al*., 2007; Koch *et al*., 2011; Moore *et al*., 2013; Bertrand *et al*., 2015; Browning *et al*., 2017; Barber-Lluch *et al*., 2019;

Joglar *et al*., 2020). Culture experiments under limiting concentrations of cobalamin have also suggested that this vitamin might be playing a role influencing mutualistic interactions between eukaryotic phytoplankton and cobalamin producers, in which cobalamin is supplied in exchange for organic matter (Croft *et al*., 2005; Kazamia *et al*., 2012; Grant *et al*., 2014; Cruz-López and Maske, 2016). Pseudocobalamin remodeling, performed by organisms with partial cobalamin synthesis capacity, may be another way in which microorganisms can interact to obtain cobalamin.

Pseudocobalamin is a cobalamin analog that is poorly bioavailable for most of the eukaryotic phytoplankton since it contains adenine as α-ligand instead of the 5,6-dimethylbenzimidazole (DMB) found in cobalamin. It is, however, produced and used by the vast majority of cyanobacteria which lack the genes necessary for the DMB synthesis (Helliwell *et al*., 2016; Heal *et al*., 2017). Some heterotrophic bacteria and eukaryotic algae have the capacity to remodel pseudocobalamin to produce cobalamin by expressing genes involved in the final synthesis of cobalamin, including DMB synthesis and activation. They can produce their own DMB or obtain it from dissolved pools and interchange it for the adenine in pseudocobalamin (Helliwell *et al*., 2016; Heal *et al*., 2017; Ma *et al*., 2020). Therefore, despite the fact that cyanobacteria are not a direct source of cobalamin, they can still have an important role in cobalamin cycling in the ocean due to the existence of these remodelers. However, the prevalence of and controls on pseudocobalamin remodeling in the ocean, as well as the factors affecting pseudocobalamin synthesis, remain poorly characterized.

Synechococcus is one of the most abundant cyanobacterial genera in the ocean (Flombaum *et al*., 2013), and thus, represent a major source of pseudocobalamin (Helliwell *et al*., 2016; Heal *et al*., 2017). It is comprised of many strains that can be grouped into

distinct clades using phylogenetic clustering (Ahlgren and Rocap, 2012). These clades appear to thrive under different levels of light, temperature, macronutrients, and metals. For example, strains that belong to clades I and IV are known to inhabit cold high nutrient waters; clades II, III, and X are more abundant in warm oligotrophic waters, and clades CRD1 and CRD2 have shown an elevated tolerance to low iron availability (Ahlgren and Rocap, 2012; Sohm et at., 2016). *Synechococcus* was identified as an important contributor to the microbial community on the Scotian Shelf, a region in the Northwest Atlantic characterized by moderately high values of primary productivity (150-300 $gCm^{-2}yr^{-1}$) (Aquarone and Adams, 2010). Prevalence of this cyanobacterium is particularly high in coastal euphotic zone waters during the fall, and to a lesser extent in spring and fall offshore euphotic zone waters (as described in Figure 2.3 and Figure 2.5 in Chapter 2). Water temperature where *Synechococcus* was present range from 10 – 26 ºC, where the colder temperatures can be found during the spring (Zorz *et al*., 2019; Figure S2.2 in Appendix A). Nutrient concentrations tend to be higher at spring surface waters compared to fall surface waters (Dasilva *et al*., 2014; Zorz *et al*., 2019). This strong spatiotemporal variability suggests that different *Synechococcus* clades may be important at different times of year, in different Scotian Shelf and Slope locations. To date, differences in *Synechococcus* clade contributions to pseudocobalamin production have not been investigated. However, identifying the dominant clades in this region is required to guide design of experiments to explore pseudocobalamin cycling on the Scotian Shelf.

In this study, 16S rRNA gene sequencing and phylogenetic placement of previously defined *Synechococcus* sequences (Zorz *et al*., 2019) were used to identify the principal *Synechococcus* strains on the Scotian Shelf and Slope. Culture experiments on a well-characterized representative strain belonging to the dominant clade (WH 8102) were

performed to identify environmental conditions that might be affecting pseudocobalamin use and synthesis by *Synechococcus* in this region. To do this, peptides for targeted, quantitative proteomic monitoring of CobO and MetH protein expression patterns were identified, enabling examination of pseudocobalamin production and use, respectively. These analyses were then coupled with particulate pseudocobalamin measurements. The quantification of peptides and pseudocobalamin was performed using selected reaction monitoring (SRM) mass spectrometry. This method enables quantification and detection of specific peptides and metabolites of interest in complex samples with high precision and sensitivity (Lange *et al*., 2008). In sum, these analyses aim to reveal environmental variables that exert control on pseudocobalamin production by cyanobacteria in the Northwest Atlantic, and offer a set of mass spectrometry-based tools to monitor their contributions to cobalamin cycling.

3.3 Material and methods

3.3.1 *Identification of major Synechococcus strains/clades on the Scotian Shelf and Slope*

Major *Synechococcus* strains on the Scotian Shelf and Slope were identified by following a methodology that allows placement of short reads/sequences from environmental samples into a reference tree that is phylogenetically resolved (Matsen *et al*., 2010). A guide/reference tree was constructed using the 16S rRNA gene sequences that are available for marine *Synechococcus* strains on the GenBank database. The sequences retrieved were aligned using ClustalW (Larkin *et al*., 2007) and manually curated. A maximum likelihood phylogenetic tree was inferred in RaxML using a GTR+Γ model followed by a 1000 bootstrap analysis (Stamatakis, 2014). Consensus sequences from 49 OTUs (operational taxonomic units; 97% similarity) identified as *Synechococcus* from the

V6-V8 region of the 16S rRNA gene from Scotian Shelf samples were obtained by Zorz *et al*., 2019. These OTU sequences were aligned to the *Synechococcus* 16S rRNA gene reference alignment using PaPaRa (Berger and Stamatakis, 2011) and placed on the phylogenetic tree previously inferred using pplacer (Matsen *et al*., 2010). The OTU sequences can be placed in one or multiple branches of the tree with a certain probability which is given by a likelihood weight ratio (LWR) value (i.e. a high LWR value means that the placement has a high probability to belong to that branch). The sum of the LWR of all placements of a single sequence should be equal to 1. OTU sequences were only assigned to a clade if that sequence was only placed on the branches of that clade or if the placements of a sequence have an accumulated LWR > 0.6 across the branches of that clade.

3.3.2 *Identification of proteins and selection of peptides*

A literature search was performed to identify proteins involved in pseudocobalamin production (Cob/Cbi proteins; see Figure 2.1 in Chapter 2) and utilization (MetH and the class II ribonucleotide reductase – NrdJ). Highly conserved proteins that are not directly related to pseudocobalamin metabolism were also searched to be used as reference proteins. Ideal reference proteins should be conserved, present only in the organisms of interest (*Synechococcus sp.*), and have stable expression under different environmental conditions (Alexander *et al*., 2012; Li *et al*., 2011). The quantification of peptides from these proteins could enable estimation of the protein contribution of *Synechococcus* in environmental samples. This can be used to normalize measurements of peptides from proteins of interest to adequately compare and interpret protein expression patterns of these proteins in environmental samples.

Peptides that were both conserved in a representative number of the *Synechococcus* strains of interest and unique to proteins involved in pseudocobalamin-related processes as

well as potential reference proteins, were selected to monitor *Synechococcus* contributions to cobalamin cycling. To do this, BLASTp was used to search for homologues of these *Synechococcus sp.* WH 8102 proteins in other available marine *Synechococcus* genomes using the GenBank nr database. The resultant amino acid sequences were aligned using COBALT (Papadopoulos and Agarwala, 2007) and digested in silico with trypsin to look for a potential conserved peptide within the sequences aligned. Once a potential conserved peptide was found, its sequence was introduced in the Unipept Tryptic Peptide Analysis search (Gurdeep *et al*., 2019) to determine if the peptide is present in a significant number of the *Synechococcus* strains of interest. Peptides were selected for further analyses and mass spectrometry testing when they met the peptide selection guidelines for SRM targeted assays described in Hoofnagle *et al*. (2016).

3.3.3 *Culture experiment to test Synechococcus peptides*

Culture experiments were performed to evaluate the expression pattern of the selected peptides, and to identify changes in pseudocobalamin synthesis and utilization under relevant conditions that may be experienced by cyanobacterial communities on the Scotian Shelf and Slope. *Synechococcus sp.* WH 8102 (CCMP2370) obtained from the National Center for Marine Algae and Microbiota (NCMA) was used as study organism. This strain was maintained axenically at 24±1°C in modified Synthetic Ocean Water (SOW) media (Dupont *et al*., 2008, Table S3.1 in Appendix A) under ~20 µmol photosynthetically active radiation m^{-2} s^{-1} in a light:dark cycle of 12:12 h, unless otherwise described (Table 3.1). Cultures were tested for axenicity once a week using SYBR Green followed by flow cytometry screening. To do this, 2 µL of 1:100 SYBR Green were added to 0.2 mL culture aliquots to stain possible bacterial contaminants cells. Given that *Synechococcus* cells are also stained, a BD Accuri™ C6 flow cytometer (BD Biosciences,

San Jose, CA) was used to differentiate between *Synechococcus* and heterotrophic bacteria cells by plotting red (FL3) versus green (FL1) fluorescence.

Experiments were performed on biological triplicates and consisted of four treatments as displayed in Table 3.1. Semicontinuous culture is designed to maintain the cells at exponential phase. This was achieved by removing a certain volume of culture (at exponential phase) and adding the same volume of new media to dilute the cell concentration to a fixed value. This was done every other day. Using this method *Synechococcus* cells were maintained in exponential phase for ~7 generations, under the conditions described, before harvesting. 50 mL of culture was harvested for protein extraction and quantification, 40 mL for metabolite (pseudocobalamin) extraction and quantification, and 20 mL for particulate organic carbon (POC) extraction and quantification. 0.2 µm polycarbonate filters were used to harvest protein samples, 0.2 µm Nylon filters for metabolites, and GF/F glass microfiber filters for POC. When culture samples where about to be harvested, relative fluorescence units (RFU) were measured in a 10-AU Fluorometer (Turner Designs, San Jose, CA), and cell counts were obtained using a BD Accuri™ C6 flow cytometer. For the treatments in batch culture, cell counts and RFU were measured every other day. When cells were at early exponential phase, 35 mL were harvested for proteins, 17 mL for metabolites, and 8 mL for POC at the three points of the diel cycle. When cells were at early stationary phase, 20 mL were harvested for proteins, 10 mL for metabolites, and 5 mL for POC at the three points of the diel cycle (Table 3.1). The three points of the diel cycle evaluated where: 1.5 h after the lights went on (9:30), 1 h before the lights went off (18:00), and 3 h after the lights went off (22:00).

Table 3.1. Experimental treatments performed to explore changes in protein abudance and pseudocobalamin content in *Synechococcus sp.* WH 8102.

Treatment	Culture type	Harvesting time
Control (24±1 ºC; 26 N:P)	Semicontinuous	After ~7 generations
Low temperature (17±1 ºC)		
Low N:P ratio (10 N:P)		
Diel cycle (harvest at 9:30, 18:00 and 22:00)	Batch	Exponential phase
		Stationary phase

3.3.4 *Protein extraction and digestion*

Protein sample filters were submerged in 750 µL of SDS extraction buffer (2% SDS, 0.1 M Tris/HCl pH 7.5, 5% glycerol, 5 mM EDTA) and incubated for 10 min on ice. Samples were then heated for 15 min at 95 ºC in a ThermoMixer® C (Eppendorf, Mississauga, ON) at 350 RPM. Cells were sonicated on ice for 1 min with a Q125 Sonicator (Qsonica Sonicators, Newton, CT) at 50% amplitude and 125 W (pulse 15 s ON, 15 s OFF). Samples were incubated at room temperature for 30 min and vortexed every 10 min. The filter was then removed and samples were centrifuged at 15,000 x g for 30 min at room temperature. 4x volume of ice-cold acetone was added to the supernatant for overnight precipitation of proteins at -20 ºC. The protein pellet was washed with 400 µL of ice-cold acetone three times and once with ice-cold methanol. Following each wash, samples were centrifuged at 15,000 x g for 30 min at room temperature and the supernatant was removed. Pellet was dried down in a Vacufuge plus (Eppendorf, Mississauga, ON) for ~15 min at room temperature. Protein was resuspended in 20 µL of 8 M urea and left to dissolve for 10 min at room temperature. 80 µL of freshly made 50 mM ammonium bicarbonate was gradually added to each sample to get a final solution of 1.6 M urea and 40 mM ammonium bicarbonate. Aliquots of 50 µg of protein were removed from each sample and diluted to 100 µL with 50 mM ammonium bicarbonate for protein digestion. 5 µL of 100 mM

Dithiothreitol (DTT) was added to the solution and samples were incubated at 56 ºC for 30 min in a ThermoMixer® C. Samples were then vortexed and incubated in dark for 30 min at room temperature after the addition of 9 µL of 200 mM iodoacetamide. Samples were vortexed and incubated for 5 min at room temperature after the addition of 5 µL of 100 mM DTT. 10 µL of 0.1 µg/µL of Pierce™ Trypsin Protease (Thermo Scientific™, Waltham, MA) (dissolved in 50 mM ammonium bicarbonate) was added to each sample to give a 1 to 50 ratio of protease to substrate. Samples were vortexed and digested at 37 ºC overnight. 0.5 µL of formic acid (or enough volume to get a pH < 3) was added to each sample to stop the proteins digestion.

3.3.5 *Peptide quantification - targeted liquid chromatography-mass spectrometry*

Targeted mass spectrometry was performed using a Dionex Ultimate 3000 UPLC system interfaced to a TSQ Quantiva triple-stage quadrupole mass spectrometer (MS) (Thermo Scientific™, Waltham, MA), fitted with a heated, low flow capillary ESI probe (HESI-II). The MS was operated with a spray voltage of 3500 V, sheath gas 5, auxiliary gas 2, ion transfer tube 325 ºC, vaporizer gas 70˚C and a Chrom filter setting of 10 s. 1 µg protein samples were spiked with 20 fmol of each heavy isotope-labeled peptide standard (described below) and loaded onto 5 mm x 0.3 mm I.D. C18 trapping column at 20 µl/min and then separated over a 150 x 0.3 mm ID reverse phase column (Acclaim C18, 2 µm, 100 Å), 4 – 43% B over 40 min, 5 µl/min, 50 °C. Mobile phase A was 0.1% formic acid; B was 80% acetonitrile, 0.08% formic acid. Isotopically labeled, heavy internal standard versions of each peptide (Table S3.2 in Appendix A) were synthesized by Thermo Scientific™ at greater than 95% purity as determined by HPLC. 100 µM stock solution of peptide standards were prepared in mixtures of acetic acid, acetonitrile and water, or 50 mM ammonium bicarbonate in the case of acidic peptides. SRM transitions were optimized by

syringe infusing 1 μM solutions and executing the Quantiva transition optimization tool. The method contained 80 transitions, 15 msec dwell time, Q1 and Q3 resolution was set to 0.7 (FWHM), automatically calibrated RF lens setting and a collision gas pressure 2.5 mTorr. Each sample was analyzed via triplicate injections. SRM parameters and details on heavy isotope labeled internal standard peptides can be found in Table S3.2 in Appendix A.

3.3.6 *Particulate metabolites extraction and quantification*

Internal standard stock solutions of heavy cyanocobalamin (CN-B_{12}) (1 μg/mL), heavy thiamine (B_1) (1 mg/mL), heavy riboflavin (B_2) (0.1 mg/mL) and heavy biotin (B_7) (0.25 mg/mL) used for quantification were prepared in the dark and stored at -80 ºC. 1 mg/mL authentic standards stock solutions of B_1, hydroxocobalamin (OH-B_{12}), adenosylcobalamin (Ado-B_{12}), CN-B_{12} and methylcobalamin (Me-B_{12}) used for calibration curves (see below) were prepared and stored at -80 ºC.

Before metabolites were extracted, enough heavy-CN-B_{12} was spiked to the metabolites samples to give a final concentration of 3 nM in the final extract after sample resuspension. Enough heavy B_1, B_2 and B_7 were also spiked to the samples such that the final extract had a final concentration of 15 nM. Metabolites were extracted from culture samples following the methodology described by Heal *et al*. (2017) with some minor modifications. Filters were transferred to 2 mL beadbeating tubes with a mixture of 100 and 400 μm silica beads, and treated with an acidic acetonitrile:methanol:water solvent mixture (40:40:20 with 0.1% formic acid). Metabolites were extracted by bead beating 3 times at 6 m/s for 40 s over a 20 min period using a dry ice cooled adapter and a Fastprep-24 bead beater (MP Biomedical, Santa Ana, CA). Samples were centrifuged briefly in a benchtop microfuge and the supernatant collected. Filters were rinsed once with 0.3 mL of

the solvent mixture and twice with 0.3 mL methanol. After each rinse, the supernatant was pooled with the primary extract. Solvent was removed using a Vacufuge (Eppendorf, Mississauga, ON) at room temperature. Samples were re-suspended in 20 mM ammonium formate, 0.1% formic acid, 2% acetonitrile on ice. Enough buffer was added to give the equivalent of extracts from 12,000 cells/µL. Samples were then vortexed and centrifuged at 10,000 x g for 3 min at 4 ºC and then diluted 2-fold with buffer in conical polypropylene HPLC vials (Phenomenex, Torrance, CA). A quality control (QC) sample was prepared by mixing equal volumes of each sample. The QC sample was injected periodically throughout the mass spectrometry analysis to monitor instrument response and compound degradation. Calibration curves with authentic cobalamin standards were prepared using the QC sample as a matrix (pseudocobalamin standards are not available). Triplicate injections were performed for 0, 0.1, 0.5, 1, 5 and 10 fmol of each authentic standard. Limits of quantitation and limits of detection were calculated as 10x and 3x the variation in the QC sample without spike, respectively.

HPLC-MS was used to quantify cobalamins using a Dionex Ultimate-3000 LC system coupled to the electrospray ionization source of a TSQ Quantiva triple-stage quadrupole mass spectrometer in SRM mode, operating under the following conditions: Q1 and Q3 resolution 0.7 (FWHM), 50 ms dwell time, spray voltage 3500 (positive ion mode), sheath gas 6, auxiliary gas 2, ion transfer tube 325 ºC, vaporizer temperature 100 ºC. Triplicate 5 µL injections were performed onto a 150 x 0.3 mm ID column (Acclaim PepMap RSLC, C-18, 2 µm, 100 Å) with a 5 x 0.3 mm ID guard column in front, held at 50 ºC and subject to an HPLC gradient of 2 – 32% B over 6 min, then 32 - 60% B over 0.5 min (A, 20 mM ammonium formate, 0.1% formic acid; B, 0.1% formic acid in acetonitrile)

at 10 µl per min. The total run time including washing and equilibration was 12 min. The metabolites transition list is shown in Table S3.3 in Appendix A.

Given that pseudocobalamin standards are not available, the following assumptions were made for pseudocobalamin quantification using cobalamin standards: (1) cobalamin and pseudocobalamin ionize at the same rate, (2) cobalamin and pseudocobalamin fragments are generated in a similar fashion, (3) cobalamin and pseudocobalamin's different elution times do not greatly affect the response.

3.3.7 *Statistical analysis*

Statistical analyses were performed using R (R Core Team, 2020). A One-way ANOVA followed by a Post-hoc Tuckey's Test was used to determine the effects of low temperature and low N:P ratio in *Synechococcus sp.* WH 8102 growth rate, cellular carbon and protein, pseudocobalamin cellular quotas, and MetH and CobO abundance. A Two-way ANOVA followed by a Post-hoc Tuckey's Test was used to determine the effects of growth phase and diel cycle on the same variables. Differences were considered significant when a P-value < 0.05 was observed.

3.4 Results and discussion

3.4.1 *Major Synechococcus clades/strains on the Scotian Shelf and Slope*

Phylogenetic placement of 49 OTU sequences identified as *Synechococcus* from Scotian Shelf and Slope environmental samples enabled the identification of major *Synechococcus* clades in this region. 63% of these OTU sequences were placed in *Synechococcus* clade III, clade I, and clade X (31%, 20%, and 12%, respectively) (Figure 3.1), suggesting that the majority of *Synechococcus* strains that inhabit the Scotian Shelf and Slope belong to these clades. It has been described that clade III and X can be found

simultaneously in warm oligotrophic waters (Sohm *et al*., 2016). Therefore, these clades may be dominating the *Synechococcus* population on the Scotian Shelf and Slope during fall, where nutrients can be depleted, and water temperatures range from 15 ºC to 26 ºC, at the euphotic zone. Clade I strains are prevalent in colder and nutrient rich waters (Ahlgren and Rocap, 2012; Sohm el at., 2016), and thus, these strains may be dominating the *Synechococcus* population on the Scotian Shelf and Slope during spring, given the higher nutrient concentrations and lower water temperatures (< 10 ºC) that can be found at the euphotic zone in this season compared to fall (Dasilva *et al*., 2014; Zorz *et al*., 2019; Figure S2.2 in Appendix A).

The identification of these important *Synechococcus* clades on the Scotian Shelf and Slope guided the selection of peptides that could target a high number (if not all) of marine *Synechococcus* strains of interest (clade III, clade I, and clade X strains), and may enable the characterization of their contribution to cobalamin cycling in this region. Furthermore, the identification of clade III as the dominant clade guided the selection of the study organism (*Synechococcus sp.* WH 8102, clade III member) that was used to explore environmental variables that could affect the pseudocobalamin synthesis and utilization on the Scotian Shelf and Slope.

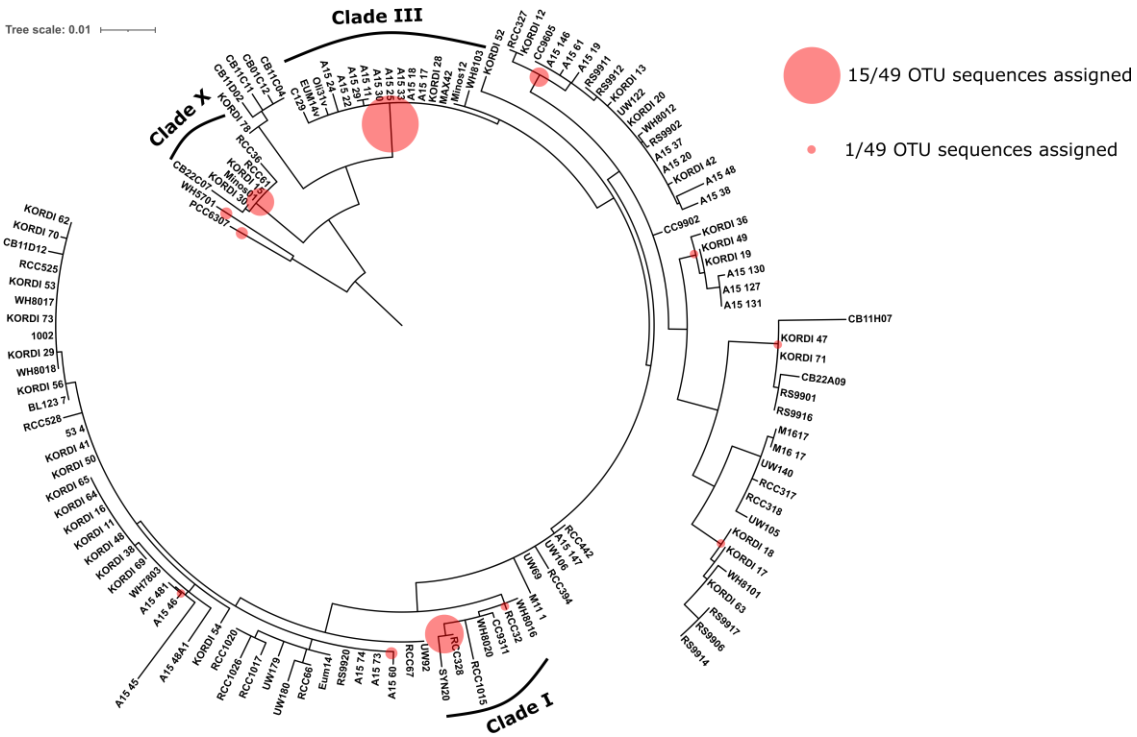

Figure 3.1. Phylogenetic placement of 49 OTUs identified as *Synechococcus* from Scotian Shelf and Slope samples. The size of the circles is related to the number of OTU sequences allocated to that clade (if the circle is placed on a node) or strain (if the circle is placed on a branch). The clades with higher number of sequences assigned are specified. The guide tree shown is a maximum likelihood tree based on *Synechococcus* strains 16S rRNA gene sequences.

3.4.2 *Identification of proteins and peptides to monitor Synechococcus contributions to cobalamin cycling*

Only one peptide from the cobalamin adenosyltransferase (CobO) and one peptide from MetH were present in a representative number of clade I, III, and X strains, and met the mass spectrometry selection criteria. These two peptides were selected for further analyses and appear to be the best candidates to monitor pseudocobalamin utilization and synthesis on the Scotian Shelf and Slope. Unfortunately, neither of them targets clade X strains with available genomes. The CobO peptide is also unable to target the clade I strains that could be analyzed, while the MetH peptide can target 67% of these strains (Figure 3.2). However, these peptides may be present in a higher number of *Synechococcus* strains

inhabiting the Scotian Shelf and Slope since this analysis is restricted to those clade members that have sequenced genomes.

The selection of potential reference proteins for this study was based only on how conserved they were in *Synechococcus* genomes, given that the response of these proteins to different environmental conditions is mostly unknown and poorly described in the literature. The potential reference proteins selected were the *Prochlorococcus/Synechococcus* Hyper Conserved Protein (PSHCP) and 30S ribosomal protein S18. Unfortunately, *Synechococcus*-specific peptides from these proteins could not be found as all peptides evaluated were also present in *Prochlorococcus* genomes. Therefore, if these proteins are used as reference proteins, the quantification of their peptides will serve as proxy for the protein contributions from both *Synechococcus* and *Prochlorococcus* in environmental samples. One peptide from each of these proteins was selected for monitoring and analysis. These peptides were highly conserved among *Synechococcus* strains, including the 100% of clade X and clade III strains analyzed. The S18 peptide is also present in the 100% of clade I strains analyzed, while the PSHCP peptide in 67% of these strains (Figure 3.2). However, the expression patterns of these proteins need to be examined before considering them as adequate reference proteins.

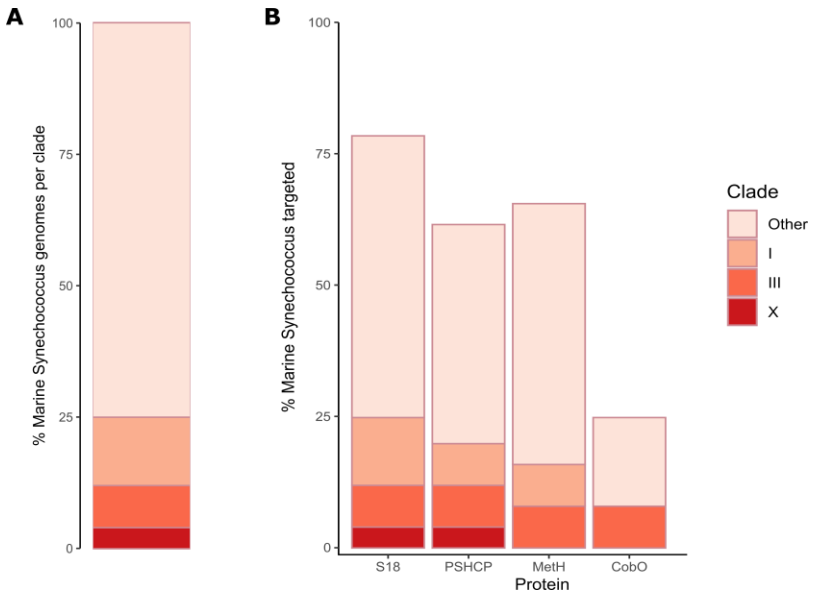

Figure 3.2. (A) Percent of marine *Synechococcus* genomes available associated with the major *Synechococcus* clades identified on the Scotian Shelf and Slope, and **(B)** the percentage of these genomes targeted by the selected peptides belonging to potential reference proteins (S18, PSHCP), and proteins involved in the pseudocobalamin utilization (MetH) and synthesis (CobO).

3.4.3 *Culture experiments in Synechococcus sp. WH 8102*

Culture experiments were performed to determine whether pseudocobalamin utilization and synthesis on the Scotian Shelf and Slope may be affected by different environmental conditions. Changes in the expression of CobO and MetH peptides, as well as in pseudocobalamin cellular content, were monitored in *Synechococcus sp.* WH 8102 (hereafter WH 8102), a representative strain from clade III. These experiments were also designed to test the suitability of the potential reference proteins selected (S18, PSHCP) by addressing their stability under the environmental conditions tested.

The growth rate of WH 8102 was significantly reduced by low temperature (17 ºC) ($P < 0.001$, Figure 3.3A). Clade III strains can inhabit cold waters (< 20 ºC); however, their abundance in such locations is not as high as it can be in regions with warmer water temperatures (Zwirglmaier *et al*., 2008; Ahlgren and Rocap, 2012; Sohm el at., 2016).

Decrease in growth rate at low temperature in WH 8102 is consistent with what has been previously described for this strain (Mackey *et al*., 2013; Varkey *et al*., 2016). Low temperature can decrease the fluidity of the membrane and the abundance of proteins involved in photosynthesis, which can lead to a reduced growth (Mackey *et al*., 2013; Varkey *et al*., 2016). It has been suggested that WH 8102, and possibly other tropical strains, do not have the machinery necessary to induce proper molecular responses to adapt to cold temperatures and recover maximum growth rate (Varkey *et al*., 2016). Low temperature also induced a significant increase in carbon ($P < 0.001$ and protein ($P < 0.01$) cellular content in WH 8102 (Figure 3.3B and Figure 3.3C). The "temperature size rule" (TSR) states that cell size decreases with temperature (Atkinson, 1994). This trend has been demonstrated in many organisms, including picophytoplankton (Morán *et al*., 2010), suggesting that the higher carbon and protein content in WH 8102 at low temperature can be explained by an increase in cell size. Furthermore, low temperatures are often associated with a decrease in the efficiency of some enzymes. An increase in the synthesis of these enzymes can be a response to maintain growth under low temperature, which can also explain the higher protein concentration in WH 8102 cells at 17 ºC compared to 24 ºC. Notably, WH 8102 cells grow faster under a lower N:P ratio (10 N:P) compared to control conditions (26 N:P), though this difference was not significant (Figure 3.3A). Clade III members have been described as oligotrophs (Zwirglmaier *et al*., 2008). Therefore, a high concentration of nitrogen may negatively affect WH 8102 growth. Cellular carbon and protein content of WH 8102 were also not significantly affected by a lower N:P ratio (Figure 3.3B and Figure 3.3C). Diel cycle and growth phase did not have a significant effect on WH 8102 protein content (Figure 3.3C), contrary to what is observed for cellular carbon, where its concentration significantly increased as the day progresses (from 9:30 to 22:00)

(P < 0.05, Figure 3.3B). This is consistent with the increase in biomass that cells go through to prepare for mitosis, which have been shown to finalize at dawn in cyanobacteria (Wang and Levin, 2009; Waldbauer *et al*., 2012). A significant increase in cellular carbon content is also observed as WH 8102 cells enter to stationary phase (P < 0.0001, Figure 3.3B). Although a decrease in cell size is known to be induced during stationary phase (Nystrom, 2004), the increased cellular carbon concentration can be explained by a possible increase in carbon-rich compounds. For example, stationary phase in phytoplankton cells was shown to be accompanied with an increase in lipid content (Schwenk *et al*., 2013).

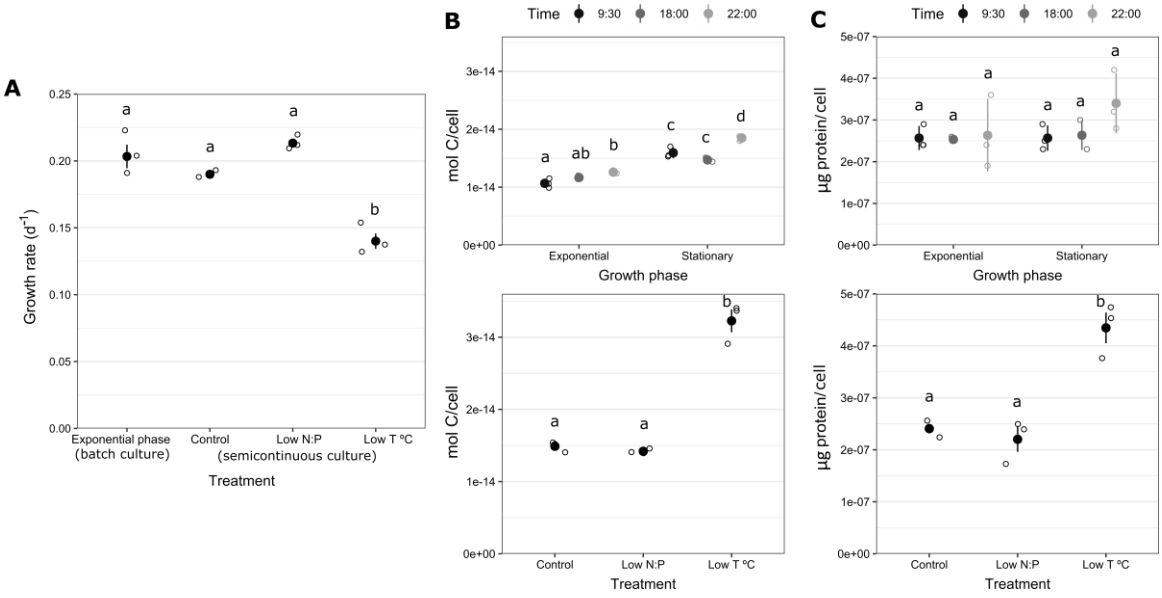

Figure 3.3. Changes in **(A)** growth rate, and **(B)** carbon and **(C)** protein cellular content in *Synechococcus sp.* WH 8102 after growth under different experimental treatments. Mean values are represented by solid circles, individual biological replicates values by open circles, and the standard deviation by lines. The type of culture (batch or semicontinuous) used for each treatment is displayed in A. Different letters over data points indicate statistically significant differences (P < 0.05) between pairs of means based on the post-hoc Tukey's Test. Shared letters indicate no significant difference. See Table S3.4 and Table S3.5 in Appendix A for P-values.

3.4.3.1 Changes in protein abundance and pseudocobalamin cellular content in Synechococcus sp. WH 8102

Cyanobacterial cell cycle is divided in three main periods that occur across the day-night cycle. At each of these stages, cells perform different metabolic processes that can differentially affect the expression of some proteins. B period starts after new cells are "born" and extends until later afternoon, when DNA replication takes place (C period). During this period cyanobacterial cells accumulate biomass, perform photosynthesis, and the synthesis of proteins is increased (Wang and Levin, 2009; Waldbauer *et al*., 2012; Welkie *et al*., 2019). This increase in protein synthesis is consistent with the observed increase in MetH abundance in the morning (9:30) at exponential phase (Figure 3.4) given that it synthesizes the amino acid methionine, though this increase was only significant when compared to MetH abundance in the night (22:00) at stationary phase ($P < 0.05$). In addition to being a component of structural proteins, methionine is also needed to synthesize SAM (Combs, 2012). SAM is a methyl donor with an important role controlling gene expression and transcription (Fontecave *et al*., 2004), and thus, an increase in methionine synthesis can be also linked to an increase in these functions during the B period. The abundance of MetH is not significantly affected by a lower N:P ratio; however, the abundance of this protein significantly increased with low temperature ($P < 0.001$, Figure 3.4). Cold temperatures may decrease the efficiency of MetH. Therefore, an increase on its abundance can be a response to maintain the necessary methionine cellular quotas to sustain growth. CobO abundance is constant through the conditions tested, though the variability in the low temperature experiment is high and the expression of this protein may in fact be affected by this condition, but this requires further examination (Figure 3.4). In

sum, these data suggest that it is possible that pseudocobalamin synthesis may not be influenced by any of the conditions tested.

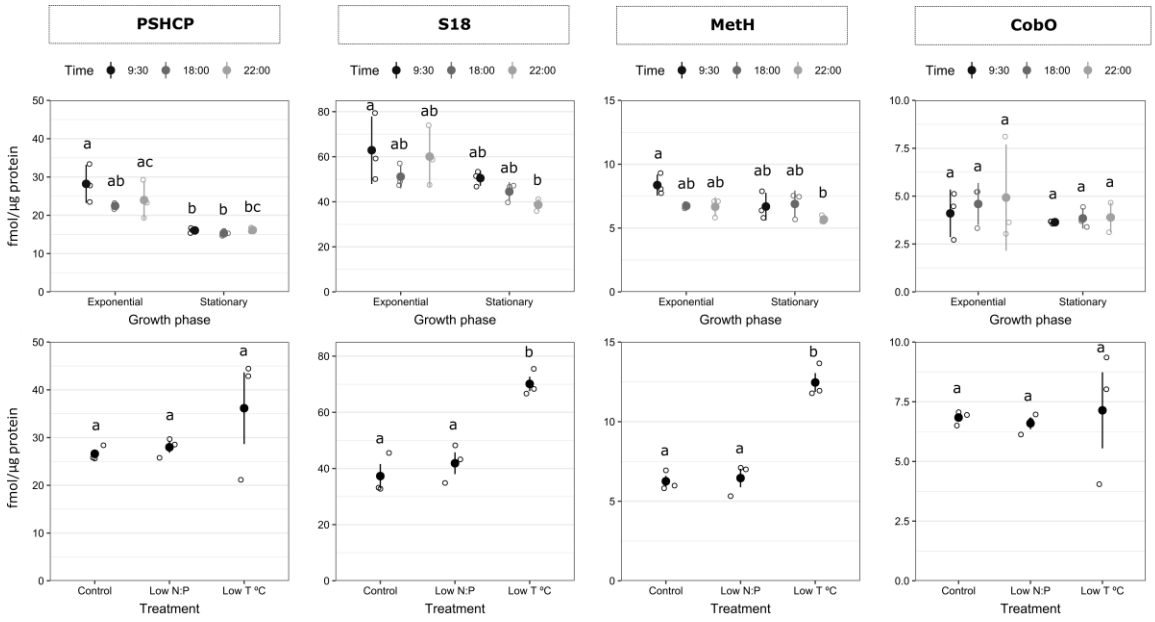

Figure 3.4. Changes in protein expression in *Synechococcus sp.* WH 8102 under different experimental treatments. Mean values are represented by solid circles, individual biological replicates values by open circles, and the standard deviation by lines. Different letters over data points indicate statistically significant differences ($P < 0.05$) between pairs of means based on the post-hoc Tukey's Test. Shared letters indicate no significant difference. See Table S3.6 and Table S3.7 in Appendix A for P-values.

Me-pseudocobalamin was the most abundant pseudocobalamin form in WH 8102 cells followed by OH- and Ado-. This result is consistent with the only other available pseudocobalamin measurements in this strain (Heal *et al*., 2017), where pseudocobalamin was only measured at one growth condition. Total pseudocobalamin content per cellular carbon quantified in WH 8102 at each experimental treatment, however, ranged from 3 to 40 times higher than what was reported by Heal *et al*. (2017). This may be the result of different pseudocobalamin quantification methods. Pseudocobalamin standards (which are now unavailable) were used to quantify pseudocobalamin by Heal *et al*. (2017), which may result in a more precise quantification of this metabolite compared to a quantification using the cobalamin standards that were used for this study (see Methods section). Alternatively,

differences in WH 8102 growth conditions between the two studies may have influenced the pseudocobalamin quantified given that, as described below, the concentration of this metabolite in WH 8102 can greatly vary with temperature. Nevertheless, as the same quantification method was used to quantify pseudocobalamin in each culture sample here, pseudocobalamin cellular content in WH 8102 can be compared between experimental treatments, and the differences observed can be attributed to biological rather than technical factors.

Total pseudocobalamin content per cellular carbon is significantly affected by growth phase and low temperature ($P < 0.01$, Figure 3.5), opposite to the invariant abundance of CobO through all experimental treatments. This suggests that CobO abundance may not be a good indicator for pseudocobalamin cellular content in WH 8102. The concentration of all pseudocobalamin forms were significantly lower at low temperature compared to the control treatment ($P < 0.05$, $P < 0.01$ for Ado-pseudocobalamin, Figure 3.5). It is possible that cell requirements for Ado- and OH-pseudocobalamin were lower under this condition. However, MetH is the only known sink of Me-pseudocobalamin in *Synechococcus*, and thus, the upregulation of this protein suggests a higher demand for this metabolite, in contrast to the decrease that was observed. This surprising result suggests that, in addition to MetH, there may be additional factors or cellular processes (currently unknown) controlling Me-pseudocobalamin quotas in WH 8102 cells. These data also suggest that low temperature resulted in a greater percentage of the available Me-pseudocobalamin being used by MetH.

Growth phase also influenced pseudocobalamin content in WH 8102. Although a significant difference between individual time points was not observed, Me-pseudocobalamin (and total pseudocobalamin) content per cellular carbon (on average)

during exponential phase was lower and significantly different than stationary phase ($P < 0.01$, Figure 3.5). This result is also surprising. The opposite trend was expected given that MetH abundance patterns suggests that the Me-pseudocobalamin demand is higher at exponential phase (Figure 3.4). This further suggests that additional, currently unknow, cellular mechanisms may be affecting Me-pseudocobalamin quotas in WH 8102 cells.

Though the decrease in Me-pseudocobalamin cellular quotas in WH 8102 at exponential phase and low temperature was not accompanied by a decrease in CobO abundance, changes in the regulation of pseudocobalamin production under these conditions cannot be ruled out. Contrary to what can be observed in many heterotrophic bacteria, Cob/Cbi genes in *Synechococcus* (including WH 8102) are not clustered in operons (Rodionov *et al*., 2003). Therefore, these genes may not be as tightly co-regulated in *Synechococcus*. The same environmental or physiological condition may have a different effect on different Cob genes, and thus, the lack of a significant change in CobO abundance in WH 8102 under low temperature does not necessarily imply that other genes involved in the pseudocobalamin synthesis were not affected by this condition. Further investigations are required to examine which proteins in the biosynthesis pathway may serve as the rate limiting step, and to examine what environmental factors control their expression.

Interestingly, OH- ($P < 0.05$) and ado-pseudocobalamin ($P < 0.0001$) quotas in WH 8102 cells were significantly affected by the diel cycle. The concentration of ado-pseudocobalamin peaked during late afternoon (18:00) at exponential and stationary phase (Figure 3.5). This may be the result of an increased requirement for ado-pseudocobalamin by the class II ribonucleotide reductase (NrdJ). NrdJ is a protein that utilizes this pseudocobalamin form as a cofactor to synthesize deoxyribonucleotides, which are the

components of DNA, and therefore, are needed for DNA replication (Jordan and Reichard, 1998). DNA replication is likely occurring during the late afternoon (Waldbauer *et al*., 2012), consistent with the observed peak in ado-pseudocobalamin at this time. A peak in OH-pseudocobalamin can be also observed during late afternoon at exponential phase ($P < 0.05$), although it is unclear what might be driving this increase.

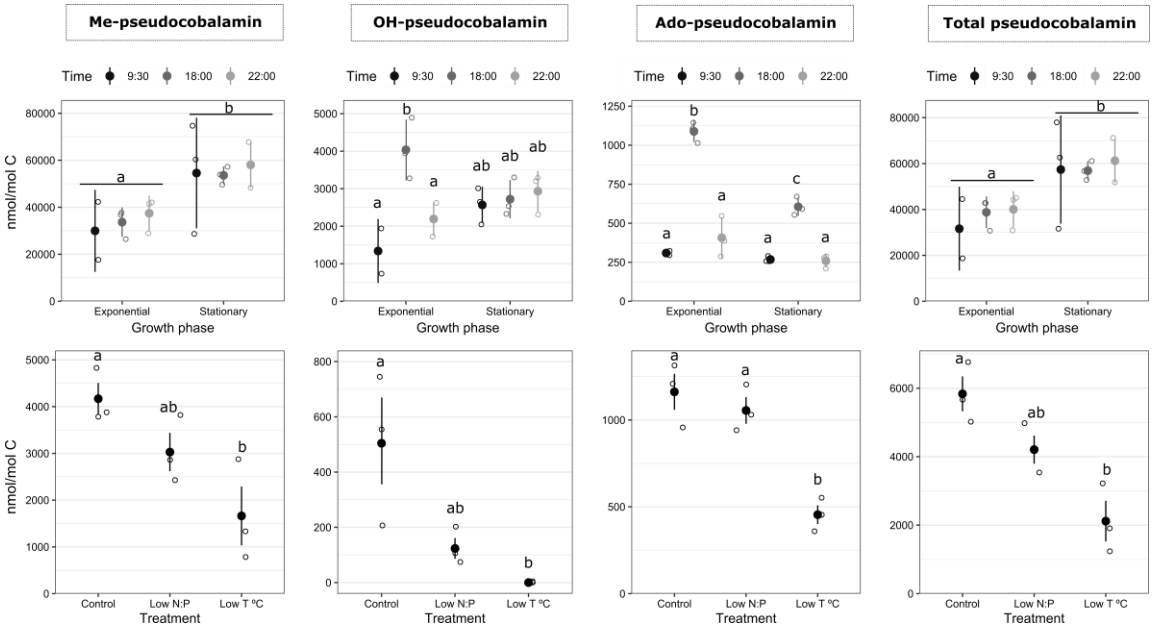

Figure 3.5. Changes in pseudocobalamin (OH-, Me-, Ado-, and total) content per total cellular carbon in *Synechococcus sp.* WH 8102 under different experimental treatments. Mean values are represented by solid circles, individual biological replicates values by open circles, and the standard deviation by lines. Different letters over data points indicate statistical significance ($P < 0.05$) between pairs of means based on the post-hoc Tukey's Test. The horizontal line above multiple groups indicates that statistical significance was not observed for individual data points but for the mean values in each growth phase, regardless of timepoint. See Table S3.6 and Table S3.7 in Appendix A for P-values.

These culture experiments also revealed that the proteins chosen as potential reference proteins may not be adequate for this purpose since they their abundances vary under different culture conditions. S18 abundance significantly decreased at stationary phase ($P < 0.05$) and increased in low temperature ($P < 0.01$). PSHCP abundance was also significantly reduced at stationary phase ($P < 0.0001$) but did not change with low

temperature. However, similar to CobO, the variability in the low temperature experiment is high (Figure 3.4). This suggests that it is possible that PSHCP abundance may significantly change with low temperature, but requires further investigation. The synthesis of proteins is usually reduced at stationary phase; therefore, the lower expression of these protein at stationary phase compared to exponential phase was expected given that S18 is a ribosomal protein and PSHCP has been suggested to be a group-specific ribosomal protein linked with photosystem assembly (Whidden *et al*., 2014). Nevertheless, considering that *Synechococcus* cells population in the ocean may be at both growth phases, the measurement of S18 and PSHCP can still be used to estimate and compare the protein contribution of *Synechococcus* between different environmental samples. Although, as both proteins may also be affected by temperature, it is likely that misleading results can be obtained when comparing samples taken at different water temperatures (e.g. fall vs spring samples) if they are used as sole reference proteins.

3.5 Conclusion

The identification of major *Synechococcus* clades on the Scotian Shelf and Slope allowed the selection of appropriate peptides to monitor the pseudocobalamin synthesis and utilization in this region. These results provided background information and proteomic tools that can be used to understand MetH and CobO protein expression patterns in Scotian Shelf and Slope environmental samples. Furthermore, low temperature was identified as a condition that can significantly decrease the pseudocobalamin cellular content in clade III strains, and therefore, has the potential to affect pseudocobalamin pools in the ocean. These same experiments can be performed with a clade I representative strain to determine whether the conditions tested might induce a different response across different *Synechococcus* strains. This study can be complemented with the analysis of additional

metabolites and proteins that are related to the pseudocobalamin synthesis and utilization pathways (e.g. SAM, methionine, other Cob proteins) to further understand the molecular mechanisms behind the responses observed to the different culture conditions tested.

Chapter 4: Conclusion

The identification of organisms that are playing a key role in the cobalamin cycle is a step forward towards understanding the influence that cobalamin may have on marine primary productivity and microbial ecological interactions. Using the Scotian Shelf and Slope as a study site, Chapter 2 of this thesis implemented a metagenomic approach that allowed the identification of then main prokaryotes with cobalamin synthesis, remodeling and utilization capacity in the Northwest Atlantic. This approach also offered genomic information about these key players that can be used to further understand their functional and ecological role in cobalamin cycling. Chapter 3 of this thesis implemented culture experiments, and metaproteomic and pseudocobalamin quantification tools, to gain insights into the contributions of *Synechococcus* to cobalamin cycling in the Northwest Atlantic. Culture condition with the potential to affect pseudocobalamin pools in this region were also identified.

Future studies can aim to characterize the spatio-temporal contributions of taxa, identified here, to cobalamin cycling in the Northwest Atlantic. This can be done by identifying specific peptides that can target cobalamin-associated proteins in these organisms (and using the already identified *Synechococcus*-specific peptides), and coupling the measurements of those peptides with the quantification of particulate and dissolve metabolites (such as cobalamin, pseudocobalamin, DMB) in Scotian Shelf and Slope environmental samples. Bottle incubation nutrient addition experiments can also be performed to further understand how cobalamin availability may be playing a key role determining the activity and abundance of different players in the cobalamin cycle by identifying changes in community composition, as well as in the abundance pattern of

metabolites and taxa-specific proteins involved in this cycle, that accompany the addition of different concentrations of cobalamin.

Future efforts can be also placed into the isolation of the identified organisms with pseudocobalamin synthesis and remodeling capacity. The isolation of these organisms would allow the implementation of culture experiments that can explore the nature of the microbial interactions that are involved in the pseudocobalamin remodeling in the Scotian Shelf and Slope. Given that some remodelers identified can encode the cobalamin-independent methionine synthase (MetE), it would be interesting to explore what mechanism, expression of MetE or pseudocobalamin remodeling, is preferred under cobalamin limitation and other culture conditions.

Future studies can also include the isolation of the cobalamin producers that were identified here to perform culture experiments to explore the effects of environmental conditions in proteins and metabolites involved in cobalamin cycling. In particular, culture experiments with MAG_32_2_*Amylibacter sp.* or closely related strains can offer valuable information about what is controlling the cobalamin synthesis on the Scotian Shelf and Slope. Additionally, the capacity of MAG_32_2_*Amylibacter sp.* to be involved in mutualistic interactions with eukaryotic phytoplankton can be examined by co-culturing MAG_32_2_*Amylibacter sp.* with currently available diatom isolates from the Scotian Shelf and Slope.

Additionally, culture experiments in *Prochlorococcus* and *Synechococcus* clade I representatives can be performed to assess whether pseudocobalamin synthesis and utilization in these organisms are differentially affected by different environmental conditions, and how these responses compare with what was observed in *Synechococcus sp.* WH 8102. Culture experiments can include the conditions tested in Chapter 3 and

additional experiments such as growth under high irradiance, cobalt limitation, and lower N:P ratio. Although cobalt limitation has not been assessed in the Northwest Atlantic, these environmental factors may be affecting the cyanobacterial population in this region.

In sum, this thesis provided tools and information that can be used to characterize the ecological interactions that may be involved in the cobalamin cycle, and to further understand how cobalamin availability can influence primary productivity in the Northwest Atlantic.

References

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*(JUN), 1–24. https://doi.org/10.3389/fmicb.2012.00213

Alexander, H., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. L., & Dyhrman, S. T. (2012). Identifying reference genes with stable expression from high throughput sequence data. *Frontiers in Microbiology*, *3*(NOV), 1–10. https://doi.org/10.3389/fmicb.2012.00385

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146. https://doi.org/10.1038/nmeth.3103

Aquarone, M.C., & Adams, S. (2010). The UNEP Large Marine Ecosystem Report, A Perspective on Changing conditions in LMEs of the World's Regional Seas: XIX-60 Scotian Shelf: LME #8. 182:795-804. UNEP. UNEP Regional Seas Report and Studies

Atkinson, D., Ciotti, B. J., & Montagnes, D. J. S. (2003). Protists decrease in size linearly with temperature: ca. 2.5%°C -1. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1533), 2605–2611. https://doi.org/10.1098/rspb.2003.2538

Ayling, M., Clark, M. D., & Leggett, R. M. (2020). New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics*, *21*(2), 584–594. https://doi.org/10.1093/bib/bbz020

Barber-Lluch, E., Hernández-Ruiz, M., Prieto, A., Fernández, E., & Teira, E. (2019). Role of Vitamin B12 in the microbial plankton response to nutrient enrichment. *Marine Ecology Progress Series*, *626*, 29–42. https://doi.org/10.3354/meps13077

Berger, S. A., & Stamatakis, A. (2011). Aligning short reads to reference alignments and trees. *Bioinformatics*, *27*(15), 2068–2075. https://doi.org/10.1093/bioinformatics/btr320

Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., Post, A. F., Sipler, R. E., Spackeen, J. L., Xu, K., Bronk, D. A., Hutchins, D. A., Allen, A. E., & Karl, D. M. (2015). Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(32), 9938–9943. https://doi.org/10.1073/pnas.1501615112

Bertrand, E. M., Saito, M. A., Rose, J. M., Riesselman, C. R., Lohan, M. C., Noble, A. E., Lee, P. A., & DiTullio, G. R. (2007). Vitamin B12 and iron colimitation of phytoplankton growth in the Ross Sea. *Limnology and Oceanography*, *52*(3), 1079–1093. https://doi.org/10.4319/lo.2007.52.3.1079

Browning, T. J., Achterberg, E. P., Rapp, I., Engel, A., Bertrand, E. M., Tagliabue, A., & Moore, C. M. (2017). Nutrient co-limitation at the boundary of an oceanic gyre. *Nature*, *551*(7679), 242–246. https://doi.org/10.1038/nature24063

Craig, S. E., Thomas, H., Jones, C. T., Li, W. K. W., Greenan, B. J. W., Shadwick, E. H., & Burt, W. J. (2015). The effect of seasonality in phytoplankton community composition on CO2 uptake on the Scotian Shelf. *Journal of Marine Systems*, *147*, 52–60. https://doi.org/10.1016/J.JMARSYS.2014.07.006

Combs, G. F. (2012). Vitamin B12. In *The Vitamins* (pp. 377–394). Elsevier. https://doi.org/10.1016/B978-0-12-381980-2.00017-7

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., & Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature*, *438*(7064), 90–93. https://doi.org/10.1038/nature04056

Crofts, T. S., Seth, E. C., Hazra, A. B., & Taga, M. E. (2013). Cobamide structure depends on both lower ligand availability and CobT substrate specificity. *Chemistry and Biology*, *20*(10), 1265–1274. https://doi.org/10.1016/j.chembiol.2013.08.006

Cruz-López, R., & Maske, H. (2016). The vitamin B1 and B12 required by the marine dinoflagellate Lingulodinium polyedrum can be provided by its associated bacterial community in culture. *Frontiers in Microbiology*, *7*(MAY), 1–13. https://doi.org/10.3389/fmicb.2016.00560

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. https://doi.org/10.1093/plankt/fbt123

de Boyer Montégut, C., G., Madec, A. S., Fischer, A., Lazar, and D. Iudicone, 2004: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. Journal of geophysical research, *109*, C12003, https://doi.org/10.1029/2004JC002378.

Delmont, T. O., & Eren, E. M. (2018). Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. *PeerJ*, *2018*(1), 1–23. https://doi.org/10.7717/peerj.4320

Doxey, A. C., Kurtz, D. A., Lynch, M. D. J., Sauder, L. A., & Neufeld, J. D. (2015). Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME Journal*, *9*(2), 461–471. https://doi.org/10.1038/ismej.2014.142

Dupont, C. L., Barbeau, K., & Palenik, B. (2008). Ni uptake and limitation in marine Synechococcus strains. *Applied and Environmental Microbiology*, *74*(1), 23–31. https://doi.org/10.1128/AEM.01007-07

Durham, B.P., Sharma, S., Luo, H., Smith, C.B., Amin, S.A., Bender, S.J., Dearth, S.P., Van Mooy, B.A., Campagna, S.R., Kujawinski, E.B., Armbrust, E.V., & Moran, M.A. (2015). Cryptic carbon and sulfur cycling between surface ocean plankton. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(2), 453–457. http://dx.doi.org /10.1073/pnas.1413137112.

Eren, A. M., Esen, O. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platformfor 'omics data. *PeerJ*, *2015*(10), 1–29. https://doi.org/10.7717/peerj.1319

Escalante-Semerena, J. C. (2007). Conversion of cobinamide into adenosylcobamide in bacteria and archaea. *Journal of Bacteriology*, *189*(13), 4555–4560. https://doi.org/10.1128/JB.00503-07

Fang, H., Kang, J., & Zhang, D. (2017). Microbial production of vitamin B12: A review and future perspectives. *Microbial Cell Factories*, *16*(1), 1–14. https://doi.org/10.1186/s12934-017-0631-y

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

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

Fontecave, M., Atta, M., & Mulliez, E. (2004). S-adenosylmethionine: Nothing goes to waste. *Trends in Biochemical Sciences*, *29*(5), 243–249. https://doi.org/10.1016/j.tibs.2004.03.007

Gómez-Consarnau L., Sachdeva, R., Gifford, S.M., Cutter L.S., Fuhrman, J.A., Sañudo-Wilhelmy, S.A., & Moran, M.A. (2018) Mosaic patterns of B-vitamin synthesis and utilization in a natural marine microbial community. *Environmental Microbiology, 20*(8), 2809-2023. https://doi.org/ 10.1111/1462-2920.14133

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International journal of systematic and evolutionary microbiology*, *57*(Pt 1), 81–91. https://doi.org/10.1099/ijs.0.64483-0

Gurdeep Singh, R., Tanca, A., Palomba, A., Van Der Jeugt, F., Verschaffelt, P., Uzzau, S., Martens, L., Dawyndt, P., & Mesuere, B. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of Proteome Research*, *18*(2), 606–615. https://doi.org/10.1021/acs.jproteome.8b00716

Grant, M. A. A., Kazamia, E., Cicuta, P., & Smith, A. G. (2014). Direct exchange of vitamin B 12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures. *ISME Journal*, *8*(7), 1418–1427. https://doi.org/10.1038/ismej.2014.9

Gobler, C.J., Norman, C., Panzeca, C., Taylor, G.T., & Sañudo-Wilhelmy, S.A. (2007). Effect of B-vitamins (B1, B12) and inorganic nutrients on algal bloom dynamics in a coastal ecosystem. *Aquatic Microbial Ecology*, *49*, 181-194. https://doi.org/10.3354/ame01132

Hannah, C. G., Shore, J. A., & Loder, J. W. (2001). Seasonal circulation on the Western and Central Scotian Shelf. *Journal of Physical Oceanography*, *31*(2), 591–615. https://doi.org/10.1175/1520-0485(2001)031<0591:SCOTWA>2.0.CO;2

Heal, K. R., Qin, W., Ribalet, F., Bertagnolli, A. D., Coyote-Maestas, W., Hmelo, L. R., Moffett, J. W., Devol, A. H., Armbrust, E. V., Stahl, D. A., & Ingalls, A. E. (2017). Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *Proceedings of the National Academy of Sciences*, *114*(2), 364–369. https://doi.org/10.1073/PNAS.1608462114

Helliwell, K. E., Lawrence, A. D., Holzer, A., Kudahl, U. J., Sasso, S., Kräutler, B., Scanlan, D. J., Warren, M. J., & Smith, A. G. (2016). Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B12. *Current Biology*, *26*(8), 999–1008. https://doi.org/10.1016/j.cub.2016.02.041

Hoofnagle, A. N., Whiteaker, J. R., Carr, S. A., Kuhn, E., Liu, T., Massoni, S. A., … Paulovich, A. G. (2016). Recommendations for the generation, quantification, storage and handling of peptides used for mass spectrometry-based assays. *Clinical Chemistry*, *62*(1), 48–69. http://doi.org/10.1373/clinchem.2015.250563

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

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 1–8. https://doi.org/10.1038/s41467-018-07641-9

Joglar, V., Prieto, A., Barber-Lluch, E., Hernández-Ruiz, M., Fernández, E., & Teira, E. (2020). Spatial and temporal variability in the response of phytoplankton and prokaryotes to B-vitamin amendments in an upwelling system. *Biogeosciences*, *17*(10), 2807–2823. https://doi.org/10.5194/bg-17-2807-2020

Kazamia, E., Czesnick, H., Nguyen, T. T. Van, Croft, M. T., Sherwood, E., Sasso, S., Hodson, S. J., Warren, M. J., & Smith, A. G. (2012). Mutualistic interactions between vitamin B12-dependent algae and heterotrophic bacteria exhibit regulation. *Environmental Microbiology*, *14*(6), 1466–1476. https://doi.org/10.1111/j.1462-2920.2012.02733.x

Knobloch, S., Jóhannsson, R., & Marteinsson, V. Þ. (2020). Genome analysis of sponge symbiont 'Candidatus Halichondribacter symbioticus' shows genomic adaptation to a host-dependent lifestyle. *Environmental microbiology*, *22*(1), 483–498. https://doi.org/10.1111/1462-2920.14869

Koch, A. L. (2001). Oligotrophs versus copiotrophs. *BioEssays*, *23*(7), 657–661. https://doi.org/10.1002/bies.1091

Koch, F., Marcoval, M. A., Panzeca, C., Bruland, K. W., Sañudo-Wilhelmy, S. A., & Gobler, C. J. (2011). The effect of vitamin B12 on phytoplankton growth and community structure in the Gulf of Alaska. *Limnology and Oceanography*, *56*(3), 1023–1034. https://doi.org/10.4319/lo.2011.56.3.1023

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2567–2572. https://doi.org/10.1073/pnas.0409727102

Lange, V., Picotti, P., Domon, B., & Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: A tutorial. *Molecular Systems Biology*, *4*(222). https://doi.org/10.1038/msb.2008.61

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Li, X., Bai, H., Wang, X., Li, L., Cao, Y., Wei, J., Liu, Y., Liu, L., Gong, X., Wu, L., Liu, S., & Liu, G. (2011). Identification and validation of rice reference proteins for western blotting. *Journal of Experimental Botany*, *62*(14), 4763–4772. https://doi.org/10.1093/jxb/err084

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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. https://doi.org/10.1098/rspb.2006.3529

Loder, John W.; Petrie, Brian; Gawarkiewicz, G. (1998). The coastal ocean off northeastern North America: a large-scale view. *The Sea*, *11*, 105–133.

Lu, X., Heal, K. R., Ingalls, A. E., Doxey, A. C., & Neufeld, J. D. (2020). Metagenomic and chemical characterization of soil cobalamin production. *ISME Journal*, *14*(1), 53–66. https://doi.org/10.1038/s41396-019-0502-0

Ma, A. T., Tyrell, B., & Beld, J. (2020). Specificity of cobamide remodeling, uptake and utilization in Vibrio cholerae. *Molecular Microbiology*, *113*(1), 89–102. https://doi.org/10.1111/mmi.14402

Mackey, K. R. M., Paytan, A., Caldeira, K., Grossman, A. R., Moran, D., Mcilvin, M., & Saito, M. A. (2013). Effect of temperature on photosynthesis and growth in marine Synechococcus spp. *Plant Physiology*, *163*(2), 815–829. https://doi.org/10.1104/pp.113.221937

MacLean, M., Breeze, H., Walmsley, J., & Corkum, J. eds. (2013). State of the Scotian Shelf Report. Can. Tech. Rep. Fish. Aquat. Sci. 3074: xvi + 352 p.

Martens, J. H., Barg, H., Warren, M., & Jahn, D. (2002). Microbial production of vitamin B12. *Applied Microbiology and Biotechnology*, *58*(3), 275–285.

Matsen, F.A., Konder, R.B., & Armbrust E.V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, *11*, 538. http://dx.doi.org/10.1186/1471-2105-11-538

Moore, C. M., Mills, M. M., Arrigo, K. R., Berman-Frank, I., Bopp, L., Boyd, P. W., Galbraith, E. D., Geider, R. J., Guieu, C., Jaccard, S. L., Jickells, T. D., La Roche, J., Lenton, T. M., Mahowald, N. M., Marañón, E., Marinov, I., Moore, J. K., Nakatsuka, T., Oschlies, A., … Ulloa, O. (2013). Processes and patterns of oceanic nutrient limitation. *Nature Geoscience*, *6*(9), 701–710. https://doi.org/10.1038/ngeo1765

Moore, L.R., Rocap, G., & Chisholm, S.W. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature*. *393*(6684), 464-7. doi: 10.1038/30965.

Morán, X. A. G., López-Urrutia, Á., Calvo-Díaz, A., & LI, W. K. W. (2010). Increasing importance of small phytoplankton in a warmer ocean. *Global Change Biology*, *16*(3), 1137–1144. https://doi.org/10.1111/j.1365-2486.2009.01960.x

Nordlund, P., & Reichard, P. (2006). Ribonucleotide reductases. *Annual Review of Biochemistry*, *75*, 681–706. https://doi.org/10.1146/annurev.biochem.75.103004.142443

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research*, *44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME Journal*, *11*(12), 2864–2868. https://doi.org/10.1038/ismej.2017.126

Panzeca, C., Tovar-Sanchez, A., Agustí, S., Reche, I., Duarte, C. M., Taylor, G. T., & Sañudo-Wilhelmy, S. A. (2006). B vitamins as regulators of phytoplankton dynamics. *Eos*, *87*(52), 4–6. https://doi.org/10.1029/2006EO520001

Papadopoulos, J. S., & Agarwala, R. (2007). COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, *23*(9), 1073–1079. https://doi.org/10.1093/bioinformatics/btm076

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996. https://doi.org/10.1038/nbt.4229

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rinke, C., Rubino, F., Messer, L. F., Youssef, N., Parks, D. H., Chuvochina, M., Brown, M., Jeffries, T., Tyson, G. W., Seymour, J. R., & Hugenholtz, P. (2019). A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME Journal*, *13*(3), 663–675. https://doi.org/10.1038/s41396-018-0282-y

Rocap, G., Distel, D. L., Waterbury, J. B., & Chisholm, S. W. (2002). Resolution of Prochlorococcus and Synechococcus ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, *68*(3), 1180–1191. https://doi.org/10.1128/AEM.68.3.1180-1191.2002

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., & Gelfand, M. S. (2003). Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes. *Journal of Biological Chemistry*, *278*(42), 41148–41159. https://doi.org/10.1074/jbc.M305837200

Sañudo-Wilhelmy, S. A., Gómez-Consarnau, L., Suffridge, C., & Webb, E. A. (2014). The role of B vitamins in marine biogeochemistry. *Annual Review of Marine Science*, *6*, 339–367. https://doi.org/10.1146/annurev-marine-120710-100912

Schwenk, D., Seppälä, J., Spilling, K., Virkki, A., Tamminen, T., Oksman-Caldentey, K. M., & Rischer, H. (2013). Lipid content in 19 brackish and marine microalgae: Influence of growth phase, salinity and temperature. *Aquatic Ecology*, *47*(4), 415–424. https://doi.org/10.1007/s10452-013-9454-z

Shadwick, E. H., & Thomas, H. (2014). Seasonal and spatial variability in the CO2 system on the Scotian Shelf (Northwest Atlantic). *Marine Chemistry*, *160*, 42–55. https://doi.org/10.1016/J.MARCHEM.2014.01.009

Shelton, A. N., Seth, E. C., Mok, K. C., Han, A. W., Jackson, S. N., Haft, D. R., & Taga, M. E. (2019). Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME Journal*, *13*(3), 789–804. https://doi.org/10.1038/s41396-018-0304-9

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. *ISME Journal*, *10*(2), 333–345. https://doi.org/10.1038/ismej.2015.115

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Tang, Y. Z., Koch, F., & Gobler, C. J. (2010). Most harmful algal bloom species are vitamin B1 and B12 auxotrophs. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(48), 20756–20761. https://doi.org/10.1073/pnas.1009566107

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33–36. https://doi.org/10.1093/nar/28.1.33

Temperton, B., & Giovannoni, S. J. (2012). Metagenomics: Microbial diversity through a scratched lens. *Current Opinion in Microbiology*, *15*(5), 605–612. https://doi.org/10.1016/j.mib.2012.07.001

Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, *2*(1), 3. https://doi.org/10.1186/2042-5783-2-3

Varkey, D., Mazard, S., Ostrowski, M., Tetu, S. G., Haynes, P., & Paulsen, I. T. (2016). Effects of low temperature on tropical and temperate isolates of marine Synechococcus. *ISME Journal*, *10*(5), 1252–1263. https://doi.org/10.1038/ismej.2015.179

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

Wang, J. D., & Levin, P. A. (2009). Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology*, *7*(11), 822–827. https://doi.org/10.1038/nrmicro2202

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

Whidden, C. E., DeZeeuw, K. G., Zorz, J. K., Joy, A. P., Barnett, D. A., Johnson, M. S., Zhaxybayeva, O., & Cockshutt, A. M. (2014). Quantitative and functional characterization of the hyper-conserved protein of Prochlorococcus and marine Synechococcus. *PLoS ONE*, *9*(10). https://doi.org/10.1371/journal.pone.0109327

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org.

Wood, D.E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology, 20*, 257. https://doi.org/10.1186/s13059-019-1891-0

Zinser, E. R., Johnson, Z. I., Coe A., Karaca E., Veneziano D., & Chisholm S. W. (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnology and Oceanography*, *52*, 2205-2220.

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*(FEB), 1–24. https://doi.org/10.3389/fmicb.2019.00281

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

Appendix A: Supplementary Material

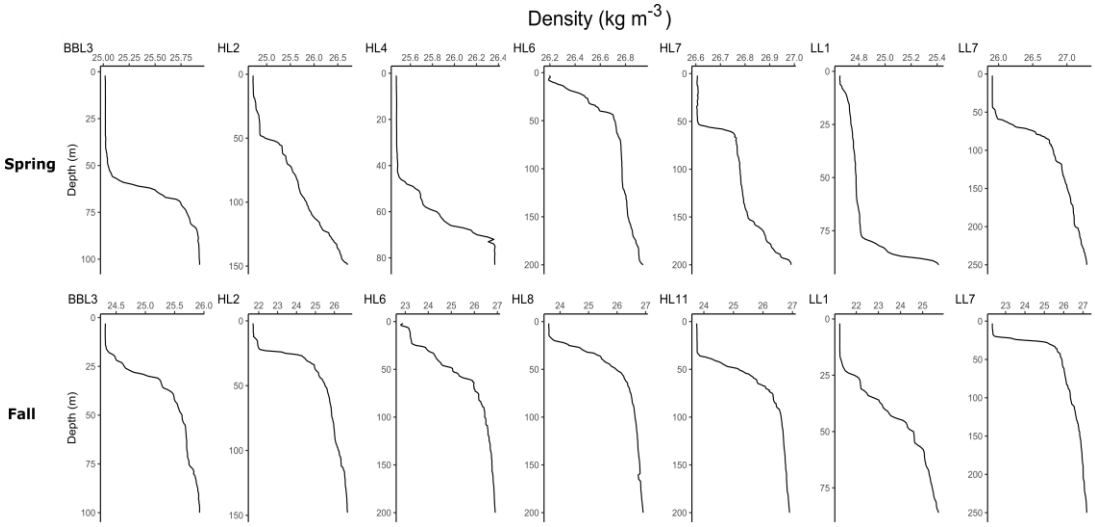

Figure S2.1. Density depth profile of the Scotian Shelf and Slope stations samples in 2016 during fall and spring that were used for metagenomic sequencing.

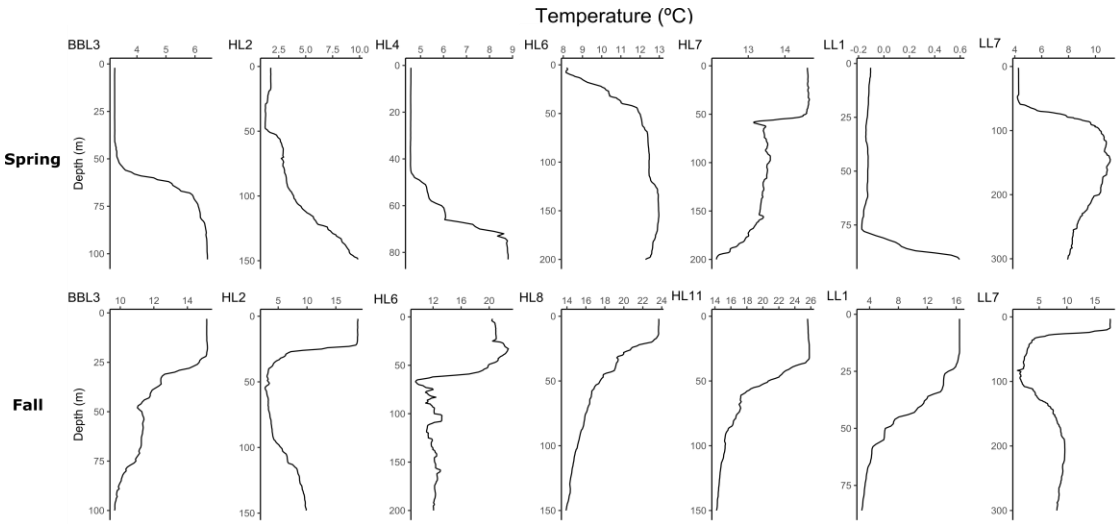

Figure S2.2. Temperature depth profile of the Scotian Shelf and Slope stations samples in 2016 during fall and spring that were used for metagenomic sequencing.

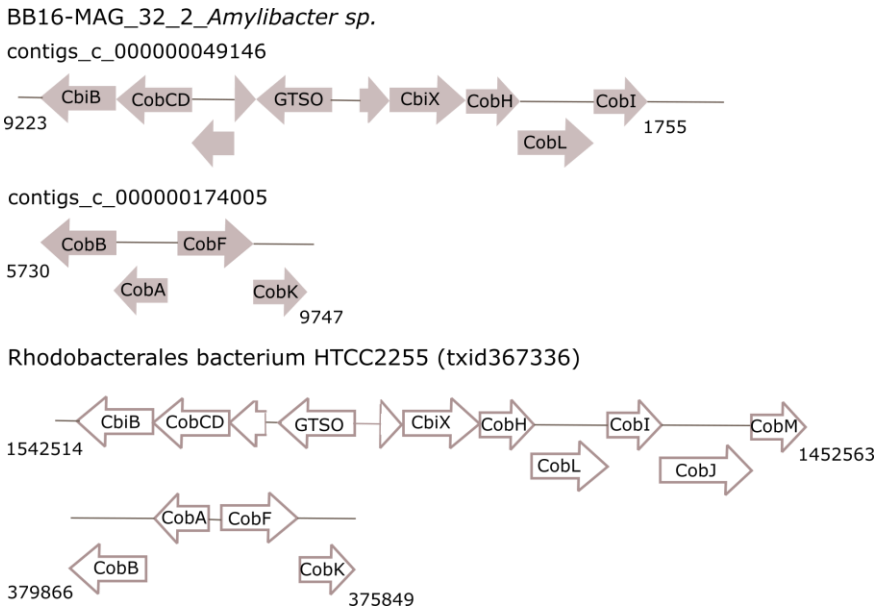

Figure S2.3. Cob/Cbi genes clusters arrangement of MAG_32_2_*Amylibacter sp.* and Rhodobacterales bacterium HTCC2255.

Table S2.1. Proteins involved in the cobalamin cycle included in the COG databases.

COG ID	Functional annotation	Protein	Role in B_{12} cycling
COG0646	Methionine synthase I (cobalamin-dependent), methyltransferase domain	MetH	B_{12}-dependent
COG1410	Methionine synthase I, cobalamin-binding domain	MetH	
COG2185	Methylmalonyl-CoA mutase, C-terminal domain/subunit (cobalamin-binding)	MCM	
COG0620	Methionine synthase II (cobalamin-independent)	MetE	B_{12}-independent
COG2073	Cobalamin biosynthesis protein CbiG	CbiG	Corrin ring synthesis
COG4822	Cobalamin biosynthesis protein CbiK, Co2+ chelatase	CbiK	
COG1010	Precorrin-3B methylase	CobJ/CbiH	
COG1648	Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain)	CysG/CobA	
COG2082	Precorrin isomerase	CobH/CbiC	
COG2099	Precorrin-6x reductase	CobK/CbiJ	
COG2241	Precorrin-6B methylase 1	CobL/CbiET	
COG2242	Precorrin-6B methylase 2	CobL/CbiET	
COG2243	Precorrin-2 methylase	CobF	
COG2875	Precorrin-4 methylase	CobM/CbiF	
COG1797	Cobyrinic acid a,c-diamide synthase	CobB/CbiA	Final synthesis
COG1429	Cobalamin biosynthesis protein CobN, Mg-chelatase	CobN	
COG2096	Cob(I)alamin adenosyltransferase	CobO	
COG2109	ATP:corrinoid adenosyltransferase	BtuR	
COG1492	Cobyric acid synthase	CobQ/CbiP	
COG1903	Cobalamin biosynthesis protein CbiD	CbiD	
COG4206	Outer membrane cobalamin receptor protein	BtuB	Uptake
COG4139	ABC-type cobalamin transport system, permease component	BtuC	
COG4138	ABC-type cobalamin transport system, ATPase component	BtuD	
COG0368	Cobalamin synthase	CobS	Remodeling
COG4547	Cobalamin biosynthesis protein CobT	CobT	
COG2087	Adenosyl cobinamide kinase/adenosyl cobinamide phosphate guanylyltransferase	CobU	
COG1865	Adenosylcobinamide amidohydrolase	CbiZ	
COG1270	Cobalamin biosynthesis protein CobD/CbiB	CobD/CbiB	

Table S2.2. Scotian Shelf and Slope DNA samples selected for metagenomic sequencing.

Station	Sample depth (m)	latitude	longitude	Season sampled	Location
BBL3	1, 40, 80	42.76	-65.48	Both	Coast
HL2	1, 40, 80	44.27	-63.32	Both	Coast
HL4	1, 40, 60	43.48	-62.45	Spring	Coast
HL6	1, 50, 80	42.85	-61.73	Both	Offshore
HL7	1, 50, 80	42.53	-61.4	Spring	Offshore
HL8	1, 60, 100	42.36	-61.34	Fall	Offshore
HL11	1, 45, 100	41.78	-60.91	Fall	Offshore
LL7	1, 20, 250	44.13	-58.18	Both	Offshore
LL1	1, 20, 60	45.83	-59.85	Both	Coast

Table S2.3. List of genome bins recovered form Scotian Shelf and Slope metagenomic samples. Bins have been classified as cobalamin consumers, producers or remodelers according to the genes involved on these processes found in their genomes.

MAG/bin name	% completion	% redundancy	Group	Supporting genes
AZOF-MAG_86_1_Microtrichales	90	0	Consumer	MetH, MCM
AZOF-MAG_86_2_Microtrichales	87	1	Consumer	MetH, MCM
AZCF-MAG_48_1_Microtrichales	90	3	Consumer	MetH, MCM, MetE
AZCS-MAG_81_1_Acidimicrobiia	93	4	Consumer	BtuB
BB16-MAG_68_2_Microbacteriaceae	92	1	Consumer	MetH, MCM
BB16-MAG_75_1_Microbacteriaceae	93	1	Consumer	MetH
AZCF-MAG_17_1_Planctomycetota	83	3	Consumer	MetH, MCM
AZCF-MAG_94_Bacteria	87	4	Consumer	MetH, MCM, BtuB
AZOF-Bin_14_4_*Marinisoma atlanticum*	96	17	Consumer	MCM, BtuB
AZOF-Bin_40_Burkholderiales	85	37	Consumer	MetH, MCM, BtuB, BtuR
AZCS-MAG_74_Bacteroidetes	89	0	Consumer	MetH, MCM, BtuB
BB16-MAG_69_Bacteroidetes	86	1	Consumer	MetH, MCM, BtuB
AZOS-MAG_102_2_Cytophagales	94	0	Consumer	BtuB
AZCS-MAG_78_Cytophagales	93	0	Consumer	MetH, BtuB
BB16-MAG_18_Saprospiraceae	97	4	Consumer	MetH, MCM, BtuB
BB16-MAG_29_Saprospiraceae	82	6	Consumer	MetH, MCM, BtuB
BB16-MAG_66_Burkholderiales	94	3	Consumer	MetH, MCM, BtuB, MetE
AZCF-MAG_57_2_Methylophilaceae	83	3	Consumer	MetH
BB16-MAG_33_2_Methylophilaceae	94	3	Consumer	MetH, BtuB
AZCS-MAG_89_Methylophilaceae	83	3	Consumer	MetH, BtuB
AZOF-Bin_15_1_Dehalococcoidetes	85	28	Consumer	MetH, BtuB
AZCF-MAG_27_1_Flavobacteriaceae	86	0	Consumer	BtuB
AZCF-MAG_28_2_Cryomorphaceae	82	1	Consumer	MetH, BtuB
AZCS-MAG_88_Cryomorphaceae	93	1	Consumer	MetH, MCM, BtuB
AZCF-MAG_69_Flavobacteriales	83	3	Consumer	MetH, MCM, BtuB

MAG/bin name	% completion	% redundancy	Group	Supporting genes
AZCS-MAG_65_Flavobacteriales	86	3	Consumer	MCM, BtuB
AZCS-MAG_83_Flavobacteriales	82	3	Consumer	MetH, MCM, BtuB
AZOF-MAG_72_1_Flavobacteriaceae	99	3	Consumer	BtuB
AZOS-MAG_101_Cryomorphaceae	99	3	Consumer	MetH, MCM, BtuB
AZCS-MAG_87_Cryomorphaceae	83	4	Consumer	MetH, MCM, BtuB
BB16-MAG_77_Cryomorphaceae	92	4	Consumer	MetH, MCM, BtuB
AZCF-MAG_55_1_Flavobacteriales	92	6	Consumer	MCM, BtuB
AZOS-MAG_32_2_Flavobacteriales	86	6	Consumer	MetH, MCM, BtuB
AZOF-MAG_54_2_Flavobacteriaceae	89	8	Consumer	BtuB
AZOF-Bin_90_Flavobacteriaceae	75	11	Consumer	BtuB
AZOS-Bin_14_1_Cryomorphaceae	82	24	Consumer	MetH, MCM, BtuB
BB16-MAG_89_2_Flavobacteriaceae	86	3	Consumer	MetH, MCM, BtuB
AZCF-MAG_115_Flavobacteriaceae	93	0	Consumer	MetH, BtuB
AZOF-MAG_62_1_Flavobacteriaceae	85	0	Consumer	MetH, BtuB
BB16-MAG_20_3_Flavobacteriaceae	90	0	Consumer	MetH, BtuB
AZCS-MAG_34_2_Flavobacteriaceae	82	1	Consumer	MCM, BtuB
AZCF-MAG_47_2_Flavobacteriaceae	85	3	Consumer	BtuB
AZCS-MAG_52_2_Flavobacteriaceae	90	4	Consumer	BtuB
AZCF-MAG_32_2_Flavobacteriaceae	80	6	Consumer	BtuB
BB16-MAG_97_Flavobacteriaceae	83	7	Consumer	MCM, BtuB
AZCS-MAG_21_3_Flavobacteriaceae	92	10	Consumer	MetH, MCM, BtuB
BB16-MAG_20_2_Flavobacteriaceae	75	10	Consumer	MetH, MCM, BtuB
AZOF-Bin_62_3_Flavobacteriaceae	85	11	Consumer	BtuB
AZOS-Bin_33_Flavobacteriaceae	85	13	Consumer	MetH, BtuB
AZCF-Bin_4_1_Flavobacteriaceae	93	14	Consumer	MetH, BtuB
AZOF-Bin_94_Flavobacteriaceae	97	15	Consumer	MetH, BtuB
AZOS-Bin_51_Flavobacteriaceae	66	23	Consumer	MCM, BtuB

MAG/bin name	% completion	% redundancy	Group	Supporting genes
BB16-Bin_80_Flavobacteriaceae	87	27	Consumer	MCM, BtuB
BB16-MAG_38_1_Flavobacteriaceae	96	0	Consumer	MetH, MCM, BtuB
AZCF-MAG_60_2_Porticoccaceae	94	3	Consumer	MetH, MCM, BtuB
AZOS-MAG_41_Pseudohongiellaceae	94	3	Consumer	MetH, BtuB, BtuR
BB16-MAG_59_2_Gammaproteobacteria	97	4	Consumer	BtuB
BB16-MAG_61_Porticoccaceae	82	4	Consumer	MetH, MCM, BtuB, BtuR
AZCS-MAG_59_Pseudohongiellaceae	89	7	Consumer	MetH, MCM, BtuB, BtuR
AZCS-MAG_91_Porticoccaceae	92	7	Consumer	MetH, MCM, BtuB, BtuR
AZOS-MAG_8_3_Porticoccaceae	77	7	Consumer	MetH, MCM, BtuB
AZCS-MAG_16_1_Porticoccaceae	90	8	Consumer	MetH, MCM, BtuB, BtuR
AZCS-MAG_16_2_Porticoccaceae	75	8	Consumer	MetH, MCM, BtuB, BtuR
AZCS-Bin_45_Porticoccaceae	83	14	Consumer	MetH, MCM, BtuB, BtuR
AZCF-Bin_25_2_Woeseiaceae	90	20	Consumer	MetH, BtuB
AZOF-Bin_11_2_Bacteria	86	20	Consumer	MetH, MCM, BtuB, MetE
AZCS-Bin_43_2_Porticoccaceae	89	27	Consumer	MetH, MCM, BtuB, MetE, BtuR
AZCF-MAG_41_1_Halieaceae	92	10	Consumer	MetH, MCM, BtuB
BB16-Bin_39_Halieaceae	89	11	Consumer	MetH, MCM, BtuB, BtuR
AZCF-Bin_33_1_Halieaceae	85	17	Consumer	MetH, MCM, BtuB
AZOS-Bin_39_Halieaceae	94	38	Consumer	MetH, MCM, BtuB, BtuR
AZCF-Bin_9_1_Halieaceae	87	23	Consumer	MetH, MCM, BtuB
AZCF-MAG_59_Planctomycetaceae	96	1	Consumer	MetH, BtuB
AZCF-MAG_10_Planctomycetaceae	89	6	Consumer	MetH
AZOS-MAG_36_Gammaproteobacteria	77	1	Consumer	MetH, MCM, BtuB, MetE
BB16-MAG_52_Gammaproteobacteria	96	3	Consumer	MetH, BtuB, BtuR
AZOF-MAG_2_8_Proteobacteria	87	10	Consumer	MetH, MCM, BtuB
BB16-MAG_62_Opitutales	82	0	Consumer	MetH, BtuB
AZOF-MAG_83_Verrucomicrobiales	92	3	Consumer	MetH, BtuB, MetE

MAG/bin name	% completion	% redundancy	Group	Supporting genes
AZCF-MAG_39_3_Thermoplasmatota	92	3	Consumer	MCM
AZCF-MAG_44_1_Thermoplasmatota	93	3	Consumer	MCM
AZCF-Bin_44_3_Thermoplasmatota	93	13	Consumer	MCM
AZOF-Bin_48_2_Thermoplasmatota	92	16	Consumer	MCM
AZOF-MAG_57_1_Thermoplasmatota	80	5	Consumer	MCM
AZOF-MAG_62_2_Thermoplasmatota	97	3	Consumer	MCM
AZOF-MAG_96_2_Thermoplasmatota	93	4	Consumer	MCM
AZOF-MAG_101_Thermoplasmatota	80	1	Consumer	MCM
AZOF-MAG_110_Thermoplasmatota	93	5	Consumer	MCM
AZOF-MAG_114_Thermoplasmatota	89	1	Consumer	MCM
AZOF-MAG_116_Thermoplasmatota	92	7	Consumer	MCM
AZOF-MAG_119_Thermoplasmatota	96	1	Consumer	MCM, BtuB
AZOS-MAG_15_2_Thermoplasmatota	84	7	Consumer	MCM
AZOS-MAG_34_1_Thermoplasmatota	84	3	Consumer	MCM
AZOS-MAG_47_2_Thermoplasmatota	93	1	Consumer	MCM
AZOS-MAG_79_1_Thermoplasmatota	93	4	Consumer	MCM
AZOS-MAG_88_Thermoplasmatota	92	3	Consumer	MCM
AZOS-MAG_69_Thermoplasmatota	88	8	Consumer	MCM
AZOS-MAG_63_Microtrichales	76	6	remodeler	CobD, CobU, CobO, CobQ, BtuR, MetH, MCM
AZOF-MAG_37_2_Parvibaculales	96	1	remodeler	CobT, CobO, BtuB, MetH, MCM
BB16-MAG_68_1_Parvibaculales	85	4	remodeler	CobT, CobO, BtuB, MetH, MCM
AZCF-MAG_8_1_Rhizobiales	97	6	remodeler	CobT, CobO, BtuB
AZCF-MAG_91_Parvibaculales	92	6	remodeler	CobT, CobO, BtuB, MetH, MCM
AZCS-MAG_35_5_Rhizobiales	90	7	remodeler	CobT, CobO, BtuB, MCM
AZOF-MAG_20_3_Rickettsiales	85	8	remodeler	CobT, BtuB, MetH
BB16-Bin_36_1_*Puniceispirillum sp.*	79	17	remodeler	CobT, CobS, CobO, MetH, MCM, MetE
AZOS-MAG_10_1_Rhodobacteraceae	87	6	remodeler	CobT, CobO, MetH, MCM

MAG/bin name	% completion	% redundancy	Group	Supporting genes
AZCS-MAG_6_3_Pseudomonadales	99	0	remodeler	CobT, CobO, BtuB, MetH
BB16-MAG_70_2_Gammaproteobacteria	89	0	remodeler	CobT, BtuB, MetH
AZOF-MAG_42_3_Gammaproteobacteria	93	1	remodeler	CobT, BtuB
BB16-MAG_40_Pseudomonadales	80	1	remodeler	CobT, BtuB, MetH
AZCS-MAG_41_2_Gammaproteobacteria	96	3	remodeler	CobT, CobO, BtuB, MetH
AZOS-MAG_19_1_Pseudomonadales	87	4	remodeler	CobT, CobO, BtuB, MetH, MCM
AZCS-MAG_11_1_Pseudohongiellaceae	75	8	remodeler	CobT, BtuR, BtuB, MetH
AZOF-MAG_20_5_Proteobacteria	92	8	remodeler	CobT, BtuB
AZOS-MAG_38_2_Gammaproteobacteria	86	10	remodeler	CobT, BtuB, MetH
AZCF-Bin_38_2_Gammaproteobacteria	86	14	remodeler	CobT, BtuB, MetH
AZCS-Bin_2_1_Pseudomonadales	86	20	remodeler	CobT, CobO, BtuB, MetH, MCM
AZOF-MAG_45_*Alteromonas sp.*	94	6	remodeler	CobD, CobO, BtuB, MetH, MetE
AZOF-MAG_103_Pseudoalteromonas	85	4	remodeler	CobT, CobD, CobQ, BtuR, BtuB, MetH, MetE
AZCS-Bin_6_1_*Oceanicoccus sp.*	93	13	remodeler	CobT, CobS, CbiZ, CobU, BtuB, MetH, MCM
AZCF-MAG_114_*Alcanivorax sp.*	100	0	remodeler	CobD, CobS, CobU, CobO, BtuR, CobQ, BtuB, MetH, MetE
BB16-MAG_50_*Planktomarina sp.*	100	0	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
AZCS-MAG_40_1_Rhodobacteraceae	89	1	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
BB16-MAG_32_2_*Amylibacter sp.*	96	1	Producer	CobF, CobH, CobK, CobL, CobM, CbiD, CbiG, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
AZCS-MAG_19_2_*Ascidiaceihabitans sp.*	99	3	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE

MAG/bin name	% completion	% redundancy	Group	Supporting genes
AZOF-MAG_39_2_Rhodobacteraceae	85	3	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
BB16-MAG_86_Rhodobacteraceae	92	3	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
BB16-MAG_51_Rhodobacteraceae	80	4	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
AZOF-MAG_37_1_Rhodobacteraceae	77	6	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CobO, BtuR, CobD, CobQ, CobT, CobU, MetH, MCM, MetE
AZOS-MAG_22_1_Rhodobacteraceae	73	7	Producer	CobF, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobT, MetH, MCM, MetE
AZCS-MAG_22_2_Rhodobacteraceae	100	8	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
AZOS-MAG_16_1_*Planktomarina sp.*	100	8	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
AZCS-Bin_14_3_Rhodobacteraceae	69	10	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
AZCS-Bin_46_*Amylibacter sp.*	82	17	Producer	CobB, CobF, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobT, CobU, MetH, MCM, MetE, BtuB
AZCS-MAG_19_1_Rhodobacteraceae	77	3	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
BB16-MAG_75_2_Puniceispirillaceae	100	4	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH, MCM, MetE
AZOF-MAG_68_2_*Synechococcus sp.*	97	3	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CobN, CbiD, CobO, BtuR, CobD, CobQ, CobS, CobT, CobU, MetH

MAG/bin name	% completion	% redundancy	Group	Supporting genes
BB16-MAG_49_*Synechococcus sp.*	80	6	Producer	CobB, CobJ, CobK, CobL, CobM, CobN, CbiD, BtuR, CobD, CobS, CobT, CobU, MetH
AZOF-MAG_92_*Prochlorococcus sp.*	83	6	Producer	CobF, CobJ, CobK, CobM, CobN, CbiD, BtuR, CobQ, CobS, MetH
AZCF-MAG_46_2_Pseudomonadales	86	7	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CbiG, BtuR, CobD, CobQ, CobS, CobU, MetH, MetE
AZCS-Bin_67_Oceanospirillaceae	79	14	Producer	CobB, CobF, CobH, CobJ, CobK, CobL, CobM, CbiD, CbiG, BtuR, CobD, CobQ, CobS, CobU, MetH, MetE

Table S2.4. Percent of cobalamin-assigned reads that could and could not be taxonomically classified. SE: surface euphotic.

Location	Percent classified	Percent unclassified
Spring-Coast_SE	5	95
Spring-Coast_Euphotic	6	94
Spring-Coast_Deep	5	95
Spring_Offshore_SE	6	94
Spring_Offshore_Euphotic	10	90
Spring_Offshore_Deep	5	95
Fall_Coast_SE	7	93
Fall_Coast_Euphotic	7	93
Fall_Coast_Deep	6	94
Fall_Offshore_SE	28	72
Fall_Offshore_Euphotic	10	90
Fall_Offshore_Deep	9	91

Table S3.1. Synthetic ocean water (SOW) media composition.

	Compound	Concentration (M)
Artificial seawater	NaCl	4.20×10^{-1}
	Na_2SO_4	2.88×10^{-2}
	KCl	9.39×10^{-3}
	$NaHCO_3$	3.00×10^{-3}
	KBr	8.40×10^{-4}
	H_3BO_3	4.85×10^{-4}
	NaF	7.14×10^{-5}
	$MgCl_2 \cdot 6H_2O$	5.46×10^{-2}
	$CaCl_2 \cdot 2H_2O$	1.05×10^{-2}
	$SrCl_2 \cdot 6H_2O$	6.38×10^{-5}
Nutrients	$NaNO_3$	2.00×10^{-3}
	$NaH_2PO_4 \cdot H_2O$	7.70×10^{-5}
	Na_2CO_3	9.40×10^{-5}
Trace metal mix	$ZnSO_4 \cdot 7H_2O$	7.72×10^{-8}
	$MnCl_2 \cdot 4H2O$	7.07×10^{-8}
	$CoCl_2 \cdot 6H_2O$	8.59×10^{-8}
	$Na_2MoO_4 \cdot 2H_2O$	1.16×10^{-6}
	$FeCl_3$	7.40×10^{-7}
	Na_2SeO_3	5.00×10^{-8}
	$CuCl_2$	5.00×10^{-8}
	$NiCl_2$	5.00×10^{-8}
Trace metal buffer	$EDTA \cdot 2H_2O$	1.35×10^{-5}

Table S3.2. *Synechococcus* peptides SRM transitions settings to quantify protein abundance.

Peptide (transitions)	Protein	Precursor (m/z)	Product (m/z)	Collision Energy (V)	Representative retention time (min)
VLESAALGWVR(+2)	PSHCP	600.8	988.5	20.7	25.6
VLESAALGWVR(+2)		600.8	859.5	21.8	25.6
VLESAALGWVR(+2)		600.8	701.4	19.3	25.6
VLESAALGWVR(+2)		600.8	517.3	18.2	25.6
VLESAALGWVR (heavy)(+2)		604.3	995.5	20.7	25.6
VLESAALGWVR (heavy)(+2)		604.3	866.5	21.8	25.6
VLESAALGWVR (heavy)(+2)		604.3	708.4	19.3	25.6
VLESAALGWVR (heavy)(+2)		604.3	517.3	18.2	25.6
YSFGYPAC[+57]PNVADSR(+2)	MetH	852.4	1086.5	26.7	22.4
YSFGYPAC[+57]PNVADSR(+2)		852.4	918.4	28.4	22.4
YSFGYPAC[+57]PNVADSR(+2)		852.4	758.4	28.5	22.4
YSFGYPAC[+57PNVADSR(+2)		852.4	618.3	22.6	22.4
YSFGYPAC[+57]PNVADSR (heavy)(+2)		855.4	1092.5	26.7	22.4
YSFGYPAC[+57]PNVADSR (heavy)(+2)		855.4	924.4	28.4	22.4
YSFGYPAC[+57]PNVADSR (heavy)(+2)		855.4	764.4	28.5	22.4
YSFGYPAC[+57PNVADSR (heavy)(+2)		855.4	618.3	22.6	22.4
GLVLVFTGQGK(+2)	CobO	559.8	849.5	17.8	25.4
GLVLVFTGQGK(+2)		559.8	736.4	18.3	25.4
GLVLVFTGQGK(+2)		559.8	637.3	17.5	25.4
GLVLVFTGQGK(+2)		559.8	389.2	17.2	25.4
GLVLVFTGQGK (heavy)(+2)		562.8	855.5	17.8	25.4
GLVLVFTGQGK (heavy)(+2)		562.8	742.4	18.3	25.4
GLVLVFTGQGK (heavy)(+2)		562.8	637.3	17.5	25.4
GLVLVFTGQGK (heavy)(+2)		562.8	389.2	17.2	25.4
IVALLPFVNPEG(+2)	S18	634.9	872.5	15.2	38.5
IVALLPFVNPEG(+2)		634.9	759.4	15.1	38.5
IVALLPFVNPEG(+2)		634.9	302.1	20.4	38.5
IVALLPFVNPEG(+2)		634.9	967.6	15.6	38.5
IVALLPFVNPEG (heavy)(+2)		637.9	878.5	15.2	38.5
IVALLPFVNPEG (heavy)(+2)		637.9	765.4	15.1	38.5
IVALLPFVNPEG (heavy)(+2)		637.9	302.1	20.4	38.5
IVALLPFVNPEG (heavy)(+2)		637.9	973.6	15.6	38.5

Table S3.3. Metabolite SRM transitions settings to quantify particulate pseudocobalamin.

Compound (transitions)	Precursor (m/z)	Product (m/z)	Collision Energy (V)	Retention time (min)
OH-pseudocobalamin	659.8	136.1	30	4.1
	659.8	348.1	30	4.1
OH-cobalamin	665	147.1	37.4	4.4
	665	359.1	24.4	4.4
Me-pseudocobalamin	668.3	136.1	30	5.7
	668.3	348.1	30	5.7
Me-cobalamin	673.5	147.1	40.8	6.1
	673.5	359.1	26	6.1
Ado-pseudocobalamin	785.3	136.1	30	4.7
	785.3	348.1	30	4.7
Ado-cobalamin	790.9	147.1	43.3	5.5
	790.9	359.1	29.7	5.5
CN-cobalamin-heavy	681.9	154.1	38	4.9
	681.9	366.1	23.7	4.9
	681.9	997.5	23	4.9
	681.9	1210.5	22.7	4.9
	681.9	639.4	21.2	4.9
	681.9	912.5	35	4.9
B1-heavy	268.1	147.1	12.2	2.2
	268.1	122.2	12.8	2.2
B2-heavy	383.2	249.1	21.1	5.2
	383.2	202.1	35.3	5.2
	383.2	175.1	35.5	5.2
	383.2	204.1	28.3	5.2

Table S3.4. One-way ANOVA and Post-hoc Tuckey's Test for the effects of low temperature and low N:P ratio in growth rate and carbon and protein cellular content in *Synechococcus sp.* WH 8102. Growth rate at stationary phase is also being included in this analysis for comparison with other treatments. Only significant pairwise comparisons based on the post-hoc Tukey's Test are shown.

Measurement	Source of variation	Df	Sum Sq	Mean Sq	F value	Pr(>F)
nmol C/cell	Treatmets	2	6.29E-28	3.14E-28	115.5	1.62E-5
	Residuals	6	1.63E-29	2.72E-30		
ug protein/cell	Treatmets	2	8.40E-14	4.20E-14	27.32	0.0009
	Residuals	6	9.23E-15	1.54E-15		
Growth rate (d^{-1})	Treatmets	3	0.009533	0.003178	34.67	6.22E-5
	Residuals	8	0.000733	0.000092		

Post-hoc Tuckey's Test

Measurement	Comparison	diff	lwr	upr	p adj
nmol C/cell	Low T ºC-Control	1.74E-14	1.32E-14	2.15E-14	3.29E-5
	Low T ºC-Low N:P	1.81E-14	1.39E-14	2.22E-14	2.61E-5
ug protein/cell	Low T ºC-Control	1.94E-07	9.54E-08	2.92E-07	0.0022
	Low T ºC-Low N:P	2.15E-07	1.16E-07	3.13E-07	0.00130
Growth rate (d^{-1})	Low T ºC-Control	-0.05	-0.07503	-0.02497	0.0009
	Low T ºC-Low N:P	-0.07333	-0.09837	-0.0483	6.34E-5
	Low T ºC-Exponential phase	-0.06333	-0.08837	-0.0383	0.0002

Table S3.5. Two-way ANOVA and Post-hoc Tuckey's Test for the effects of growth phase and diel cycle in carbon and protein cellular content in *Synechococcus sp.* WH 8102. Only significant pairwise comparisons based on the post-hoc Tukey's Test are shown.

Measurement	Source of variation	Df	Sum Sq	Mean Sq	F value	Pr(>F)
nmol C/cell	Time	2	2.14E-29	1.07E-29	32.19	1.50E-5
	Growth phase	1	1.01E-28	1.01E-28	303.98	6.88E-10
	Time*Growth phase	2	6.80E-30	3.40E-30	10.26	0.0025
	Residuals	12	3.98E-30	3.30E-31		
ug protein/cell	Time	2	7.81E-15	3.91E-15	1.48	0.267
	Growth phase	1	3.76E-15	3.76E-15	1.42	0.256
	Time*Growth phase	2	5.21E-15	2.61E-15	0.99	0.402
	Residuals	12.00	3.17E-14	2.64E-15		

Post-hoc Tuckey's Test

Measurement	Comparison	diff	lwr	upr	p adj
nmol C/cell	22:00-18:00	2.37E-15	1.48E-15	3.25E-15	3.36E-5
	9:30-22:00	-2.25E-15	-3.14E-15	-1.36E-15	5.50E-5
	Stationary-Exponential	4.73E-15	4.14E-15	5.32E-15	0
	18:00:Stationary-18:00:Exponential	3.03E-15	1.45E-15	4.61E-15	0.0004
	22:00:Stationary-18:00:Exponential	6.83E-15	5.25E-15	8.41E-15	1E-7
	9:30:Stationary-18:00:Exponential	4.27E-15	2.69E-15	5.85E-15	1.18E-5
	9:30:Exponential-22:00:Exponential	-1.93E-15	-3.51E-15	-3.54E-16	0.0141
	18:00:Stationary-22:00:Exponential	2.10E-15	5.21E-16	3.68E-15	0.0078
	22:00:Stationary-22:00:Exponential	5.90E-15	4.32E-15	7.48E-15	3E-7
	9:30:Stationary-22:00:Exponential	3.33E-15	1.75E-15	4.91E-15	1.43E-4
	18:00:Stationary-9:30:Exponential	4.03E-15	2.45E-15	5.61E-15	2.12E-5
	22:00:Stationary-9:30:Exponential	7.83E-15	6.25E-15	9.41E-15	0
	9:30:Stationary-9:30:Exponential	5.27E-15	3.69E-15	6.85E-15	1.2E-6
	22:00:Stationary-18:00:Stationary	3.80E-15	2.22E-15	5.38E-15	3.90E-5
	9:30:Stationary-22:00:Stationary	-2.57E-15	-4.15E-15	-9.87E-16	0.0016

Table S3.6. One-way ANOVA and Post-hoc Tuckey's Test for the effects of low temperature and low N:P ratio in protein abundance and pseudocobalamin content in *Synechococcus sp.* WH 8102. Only significant pairwise comparisons based on the post-hoc Tukey's Test are shown.

		Source of variation	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Protein (fmol/µg protein	CobO	Treatments	2	0.45	0.22	0.085	0.920
		Residuals	6	15.78	2.63		
	MetH	Treatments	2	74.63	37.32	46.65	2.21E-4
		Residuals	6	4.80	0.80		
	PSHCP	Treatments	2	159.70	79.86	1.37	0.325
		Residuals	6	350.90	58.49		
	S18	Treatments	2	1897	948.70	23.62	0.001
		Residuals	6	241	40.20		
Metabolite (nmol/mol C)	Me-pseudocobalamin	Treatments	2	9447425	4723712	6.96	0.0273
		Residuals	6	4071719	678620		
	OH-pseudocobalamin	Treatments	2	414553	207276	7.95	0.0206
		Residuals	6	156447	26075		
	Ado-pseudocobalamin	Treatments	2	872170	436085	22.28	0.00167
		Residuals	6	117437	19573		
	Total-pseudocobalamin	Treatments	2	20839338	10419669	13.33	0.0062
		Residuals	6	4689895	781649		

Post-hoc Tuckey's Test

		Comparison	diff	lwr	upr	p adj
Protein (fmol/µg protein	MetH	Low T ºC-Control	6.21	3.97	8.25	0.0004
		Low T ºC-Low N:P	6.01	3.77	8.25	0.0004
	S18	Low T ºC-Control	32.83	16.95	48.71	0.0017
		Low T ºC-Low N:P	28.26	12.38	44.14	0.0038
Metabolite (nmol/mol C)	Me-pseudocobalamin	Low T ºC-Control	-2506.33	-4570.11	-442.56	0.0228
	OH-pseudocobalamin	Low T ºC-Control	-504.33	-908.87	-99.80	0.0204
	Ado-pseudocobalamin	Low T ºC-Control	-707.33	-1057.82	-356.84	0.00198
		Low T ºC-Low N:P	-600.33	-950.82	-249.84	0.00459
	Total-pseudocobalamin	Low T ºC-Control	-3717.86	-5932.77	-1502.96	0.00508

Table S3.7. Two-way ANOVA and Post-hoc Tuckey's Test for the effects of growth phase and diel cycle in protein abundance and pseudocobalamin content in *Synechococcus sp.* WH 8102. Only significant pairwise comparisons based on the post-hoc Tukey's Test are shown.

		Source of variation	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Protein (fmol/µg protein)	Cobo	Time	2	0.89	0.45	0.24	0.7940
		Growth phase	1	2.53	2.53	1.34	0.2700
		Time*Growth phase	2	0.25	0.12	0.07	0.9370
		Residuals	12	22.67	1.89		
	MetH	Time	2	5.50	2.75	4.58	0.0333
		Growth phase	1	3.16	3.16	5.27	0.0406
		Time*Growth phase	2	2.56	1.28	2.13	0.1612
		Residuals	12	7.21	0.6		
	PSHCP	Time	2	32.70	16.4	1.90	0.1910
		Growth phase	1	368.00	368	42.83	2.75E-5
		Time*Growth phase	2	22.20	11.1	1.29	0.3100
		Residuals	12	103.10	8.6		
	S18	Time	2	269.20	134.6	1.74	0.2170
		Growth phase	1	815.00	815	10.54	0.0070
		Time*Growth phase	2	165.90	83	1.07	0.3730
		Residuals	12	927.80	77.3		
Metabolite (nmol/mol C)	Me-pseudocobalamin	Time	2	5.42E+7	2.71E+7	0.164	0.8507
		Growth phase	1	1.94E+9	1.94E+9	11.752	0.0056
		Time*Growth phase	2	1.64E+7	8.22E+6	0.05	0.9517
		Residuals	11	1.82E+9	1.65E+8		
	OH-pseudocobalamin	Time	2	4826928	2413464	6.583	0.0132
		Growth phase	1	88886	88886	0.242	0.6321
		Time*Growth phase	2	5142244	2571122	7.013	0.0109
		Residuals	11	4033023	366638		
	Ado-pseudocobalamin	Time	2	1123664	561832	111.65	4.94E-8
		Growth phase	1	236479	236479	46.99	2.75E-5
		Time*Growth phase	2	148509	74254	14.76	0.0008
		Residuals	11	55353	5032		
	Total-pseudocobalamin	Time	2	3.98E+7	1.99E+7	0.116	0.8919
		Growth phase	1	1.93E+9	1.93E+9	11.183	0.0066
		Time*Growth phase	2	3.92E+7	1.96E+7	0.114	0.8935
		Residuals	11	1.89E+9	1.72E+8		

Post-hoc Tuckey's Test						
		Comparison	**diff**	**lwr**	**upr**	**p adj**
Protein (fmol/µg protein)	MetH	9:30-22:00	1.35	0.16	2.55	0.0266
		Stationary-Exponential	-0.84	-1.63	-0.04	0.0406
		22:00:Stationary-9:30:Exponential	-2.69	-4.81	-0.56	0.0112
	PSHCP	Stationary-Exponential	-9.04	-12.05	-6.03	0.0000
		18:00:Stationary-22:00:Exponential	-8.68	-16.71	-0.64	0.0320
		18:00:Stationary-9:30:Exponential	-12.92	-20.96	-4.88	0.0017
		22:00:Stationary-9:30:Exponential	-12.06	-20.10	-4.02	0.0030
		9:30:Stationary-9:30:Exponential	-12.16	-20.20	-4.12	0.0028
	S18	Stationary-Exponential	-13.46	-22.49	-4.43	0.0070
		22:00:Stationary-9:30:Exponential	-24.20	-48.32	-0.09	0.0490
Metabolite (nmol/mol C)	Me-pseudocobalamin	Stationary-Exponential	21321.68	7574	35068	0.0058
	OH-pseudocobalamin	9:30-18:00	-1303	-2293	-312	0.0116
		22:00:Exponential-18:00:Exponential	-1844	-3530	-157	0.0299
		9:30:Exponential-18:00:Exponential	-2699	-4584	-813	0.0049
	Ado-pseudocobalamin	22:00-18:00	-514.33	-624	-403	2E-7
		9:30-18:00	-562.97	-678	-446	1E-7
		Stationary-Exponential	-235	-311	-159	2.9E-5
		22:00:Exponential-18:00:Exponential	-681.67	-879	-484	1.6E-6
		9:30:Exponential-18:00:Exponential	-779.67	-1000	-558	1.3E-6
		18:00:Stationary-18:00:Exponential	-483	-680	-285	4.9E-5
		22:00:Stationary-18:00:Exponential	-830.00	-1027	-632	2E-7
		9:30:Stationary-18:00:Exponential	-821.00	-1018	-623	2E-7
		18:00:Stationary-22:00:Exponential	198.67	1.14	396	0.0484
		18:00:Stationary-9:30:Exponential	296.67	75.82	517	0.0077
		22:00:Stationary-18:00:Stationary	-347.00	-544	-149	0.0010
	Total-pseudocobalamin	Stationary-Exponential	21230.52	7198.85	35262	0.0067