

Exploring vitamin B1 cycling and its connections to the microbial community in the north Atlantic ocean

Suffridge C.P., Bolaños L.M., Bergauer K., Worden A.Z., Morré J.,
Behrenfeld M.J., and Giovannoni S.J.

Luis M. Bolaños

June 2020

Contents

1	Introduction	1
2	Samples, data generation and raw data availability	2
3	Get NAAMES4 16S rRNA datasets that were coupled to vitamin measurements	2
4	Figures	3
4.1	Figure2: MAP. NAAMES 4 station locations	4
4.2	Figure7 Canonical Analysis of Principle Coordinates (CAP) of 16S rRNA ASV profiles constrained by physical and chemical environmental parameters, including TRC measurements	4
4.3	Figure8 Contrasting oceanographic and microbial dynamics at stations 1 and 2	4
5	Provided Files	4

1 Introduction

This document was created to provide the data and scripts necessary to reproduce the the community composition analysis presented in the manuscript "Exploring vitamin B1 cycling and its connections to the microbial community in the north Atlantic ocean". This project is part of the The North Atlantic Aerosols and Marine Ecosystems Study (NAAMES). Data and analysis hosted in this github directory is part of the larger analysis that can be found

in https://github.com/lbolanos32/NAAMES_2020
(Seasonality of the microbial community composition in the western North Atlantic)

2 Samples, data generation and raw data availability

16S rRNA amplicon sequencing was performed on libraries made of amplicons using 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 338 RPL (5'-GCWGCCWCCCGTAGGWT-3') primer set. For more information: <https://doi.org/10.1038/s41396-020-0636-0> and https://github.com/lbolanos32/Phyto_NAAMES_2019. Raw 16S rRNA datasets are publicly available at:

- **NCBI SRA.** BioProject: PRJNA627189 (SRA accession numbers SRR11596858 to SRR11597110)
- **NASA SeaWiFS Bio-optical Archive and Storage System (SeaBASS).** Amplicon Sequences download Instructions
 - 1) Go to <https://seabass.gsfc.nasa.gov/investigator/Giovannoni,%20Stephen>
 - 2) Click the blue [search] button
 - 3) Check the "Include all associated files" box.
 - 4) Click download all

Questions? please contact

- Stephen Giovannoni (Steve.giovannoni@oregonstate.edu) or
- Luis M Bolanos (bolanosl@oregonstate.edu, lbolanos@lcg.unam.mx)

Citation:

- If curated phytoplankton fraction from "https://github.com/lbolanos32/Phyto_NAAMES_2019" is used, please cite: "Small Phytoplankton Dominate Western North Atlantic Biomass" (<https://doi.org/10.1038/s41396-020-0636-0>).
- If data sets are used in any other way, you followed this pipeline or used the attached scripts, please cite: "Seasonality of the microbial community composition in the western North Atlantic".

3 Get NAAMES4 16S rRNA datasets that were coupled to vitamin measurements

Allcat files can be found in https://github.com/lbolanos32/NAAMES_2020/tree/master/Tables

Listing 1: Extract 16S rRNA with vitamin measurements

```
#Load libraries
library("phyloseq")
count_tab <- read.table("Allcat.otu.txt", header=T, row.names=1,
  check.names=F)
sample_info_tab <- read.table("Allcat.env.txt", header=T, row.
  names=1, check.names=F, sep = "\t")
tax_tab <- as.matrix(read.table("Allcat.tax", header=T, row.names
  =1, check.names=F, na.strings="", sep="\t"))

sample_info_tab$depth <- factor(sample_info_tab$depth, levels = c(
  "5", "25", "50", "75", "100", "150", "200", "300"))

OTU = otu_table(count_tab, taxa_are_rows = TRUE)
TAX = tax_table(tax_tab)
SAM = sample_data(sample_info_tab)

physeqall = phyloseq(OTU, TAX, SAM)

vits = get_variable(physeqall, "vits") %in% "yes"
sample_data(physeqall)$vits <- factor(vits)
physeqvits<-subset_samples(physeqall, vits %in% TRUE)
#physeqvits

physeqvits_prune<-filter_taxa(physeqvits, function(x) sum(x)>0,
  TRUE)

write.table(tax_table(physeqvits_prune), file= "/Users/luisbolanos
  /Documents/OSU_postdoc/NAAMES/Vits_Chris/vitsf.tax", quote=
  FALSE, sep = "\t")
write.table(otu_table(physeqvits_prune), file= "/Users/luisbolanos
  /Documents/OSU_postdoc/NAAMES/Vits_Chris/vitfs.otu", quote=
  FALSE, sep = "\t")

#The generated files can be found in this directory. Below is the
  list of Rscripts used to generate the figures in the
  manuscript.
```

4 Figures

Cutadapt is used to chop the primers from the raw sequences. In this SOP, we are using a fixed number of bp to trim each paired-end. The fixed number match the primer length: 27F (20bp) and 338RPL (18bp) (**trimf.sh** and **trimrev.sh**) to chop the fixed number of bp from the raw sequences

Task 4.1 Figure2: MAP. NAAMES 4 station locations

This MAP was achieved using the file CoordsStV2.txt provided in lbolanos32/NAAMES_2020/Tables/ directory of this github and overlapped with the MDT map

```
library(ggplot2)
library(ggmap)
library(maps)
library(mapdata)

samps <- read.table("/Users/luisbolanos/Documents/MisDrafts/
  InProgress/NAAMES_annual/CoordsStV2.txt", header=T, sep="\t")

#Get the row number
N4<-which(samps$cruise=="NAAMES4")
#NAAMES4
svg("N4map.svg")
image(x=-75:-20, y = 30:60, z = outer(0, 0), xlab = "lon", ylab =
  "lat")
map("world", add = TRUE, fill=TRUE,bg='light blue')
points(samps[N4,3], samps[N4,2], pch=19, col="black", cex=1.5,
  type="p")
dev.off()
```

Task 4.2 Figure7 Canonical Analysis of Principle Coordinates (CAP) of 16S rRNA ASV profiles constrained by physical and chemical environmental parameters, including TRC measurements

Rscript provided along this document CAPFIG7Vits.R

Task 4.3 Figure8 Contrasting oceanographic and microbial dynamics at stations 1 and 2

Rscript provided along this document HMFIG8Vits.R

5 Provided Files

- vitsf.tax
- vits.otu
- vits.env