Supplementary Information

Towards quantitative microbiome community profiling using internal standards

Yajuan Lin*, Scott Gifford, Hugh Ducklow, Oscar Schofield,
and Nicolas Cassar*

* Corresponding authors. Email: yajuan.lin@duke.edu (Y.L.) and Nicolas.Cassar@duke.edu (N.C.)

This PDF file includes:

Material and Methods

Figures S1 and S2

Tables S1, S2 and S3

References

Other supplementary material includes:

Tables S4 and S5 (Excel format)

Materials and Methods

Site description and environmental sampling

The Palmer Antarctica Long Term Ecosystem Research (PAL) program was initiated in 1991 to study how annual sea ice variability structures the ecology of the West Antarctic Peninsula (WAP). A major component of PAL is an annual ship-survey conducted each austral summer in the month of January. The cruise conducts a series of cross shelf transects that span a north-south gradient. Discrete water samples were collected by Niskin bottles attached to a profiling Conductivity-Temperature-Depth (CTD) rosette system. All core datasets collected by the program are freely available through the Palmer LTER data system (http://pal.lternet.edu/data).

Each station was sampled by a rosette equipped with a SeaBird Conductivity-Temperature-Depth (CTD) system and collecting discrete water samples with Niskin bottles. Seawater collected at different depths was filtered (1-2 L) onto 25 mm GF/F filters, wrapped in foil, and frozen at -80ºC for fluorometric phytoplankton chlorophyll-*a* (chl-*a*) analysis (mg chl-*a* m^{-3}) or flash frozen in liquid nitrogen and stored at -80ºC for HPLC pigment/phytoplankton composition analysis (mg pigment m^{-3}). Filters were extracted in 90% HPLC grade acetone for 24 hours after being macerated with either Teflon pestle and being ultrasonicated. Procedures and gradients for the HPLC procedures are described in (1). The taxonomic composition of the phytoplankton assemblages was derived quantitatively from a High Performance Liquid Chromatography (HPLC) analysis of pigment data using the program CHEMTAX (V195) with initial pigment ratios derived from WAP phytoplankton by (1). Starting from the initial pigment ratio matrix and the observed pigment concentration data set, CHEMTAX quantitatively estimates the phytoplankton abundances (in chl-a

concentrations) using factor analysis and a "steepest descent" algorithm to optimize the pigment ratios (2).

Phytoplankton carbon-fixation rates were measured by the uptake of radioactive sodium bicarbonate. In borosilicate flasks, 200 ml aliquots of the sea-water, were inoculated with nominally 1 μCi of $NaH^{14}CO_3$ per bottle. The borosilicate bottles were incubated for 24 hours with bottles screened to *in situ* light levels and incubated in an outdoor deck incubator. After incubation, samples were filtered onto GF/F filters, washed with 10% HCl, dried and counted in a scintillation counter.

Bacterial productivity rates were derived from rates of ^{3}H-leucine incorporation measured. The leucine assays followed a procedure modified from the protocol originally proposed by (3). Briefly, triplicate 1.5 ml samples were incubated for ~3 h with $_{3}$H-leucine (MP Biomedical, Santa Ana, CA; >100 Ci/mmol, 20-25 nM final concentration) in 2.0 ml microcentrifuge tubes (Axygen SCT-200, Union City, CA). Incubations were maintained within 0.5°C of the *in situ* temperature in refrigerated circulator baths and terminated by the addition of 0.1 ml of 100% trichloroacetic acid (TCA). Samples were concentrated by centrifugation, rinsed with 5% TCA and 70% ethanol and air-dried overnight prior to radioassay by liquid scintillation counting in Ultima Gold cocktail (Perkin-Elmer, Waltham, MA). Blank values of TCA-killed samples were subtracted from the average of the triplicates for each discrete depth sample.

For DNA and RNA, 4 L of surface seawater from the ship's flow-through underway system was gently vacuum-filtered onto a 47 mm 0.45 μm pore size Supor filter for years 2012 and 2013, or 47 mm 0.2 μm Supor filters for year 2015. For some coastal high biomass stations, filtration was stopped when the filter was clogged (< 4 L) and the total filtrate volume was recorded. 1 ml of RNA later was added to each filter, and the filter was stored in -

80C freezer immediately until further analysis. Surface seawater for flow cytometry were collected from the Niskin bottles and analyzed on board as described in (4)(5).

Amplicon library construction

16S rRNA gene were amplified by PCR using V4 primer set 515F (5' – GTGYCAGCMGCCGCGGTAA – 3') (6) and 805R (5' – GACTACNVGGGTATCTAAT – 3') modified from (7) and (8). To decrease the degeneracy of the reverse primer the 13th base pair is 'A' (7, 9) not 'W' as in the 806R V4 primer (8). With this single base difference, the total degeneracy of the reverse primer decreases from 24 to 12, while it still yields good *in silico* primer coverage using SILVA TestProbe (10) with zero mismatch hits as below:

Bacteria 91.8%, including Bacterioidetes 93.3%, Cyanobacteria 91.0% (mostly low hits on choloroplast 16S rRNA gene), alphaproteobacteria 95.2%, betaprotebacteria 94.9%, deltaproteobacteria 93.5%, and gammaproteobacteria 94.5%.

Archaea 91.1%, including Crenarchaeota 90.9%, and Euryarchaeota 92.8%.

18S rRNA genes were amplified by PCR using V4 primer set EukF (5' – CCAGCASCYGCGGTAATTCC – 3') (11) and EukR (5' – ACTTTCGTTCTTGAT – 3') modified from (11) as described in (12). The zero mismatch using SILVA TestProbe is as below,

Original (11) reverse primer: Cryptophyceae 94.9%, Haptophyta 2.4%, Alveolata 80.8% (including Dinoflagellata 90.2%), Stramenopiles 92.6% (including Diatomea 93.7%).

Modified reverse primer: Cryptophyceae 94.9%, Haptophyta 97.1%, Alveolata 81.1% (including Dinoflagellata 90.7%), Stramenopiles 93.0% (including Diatomea 94%).

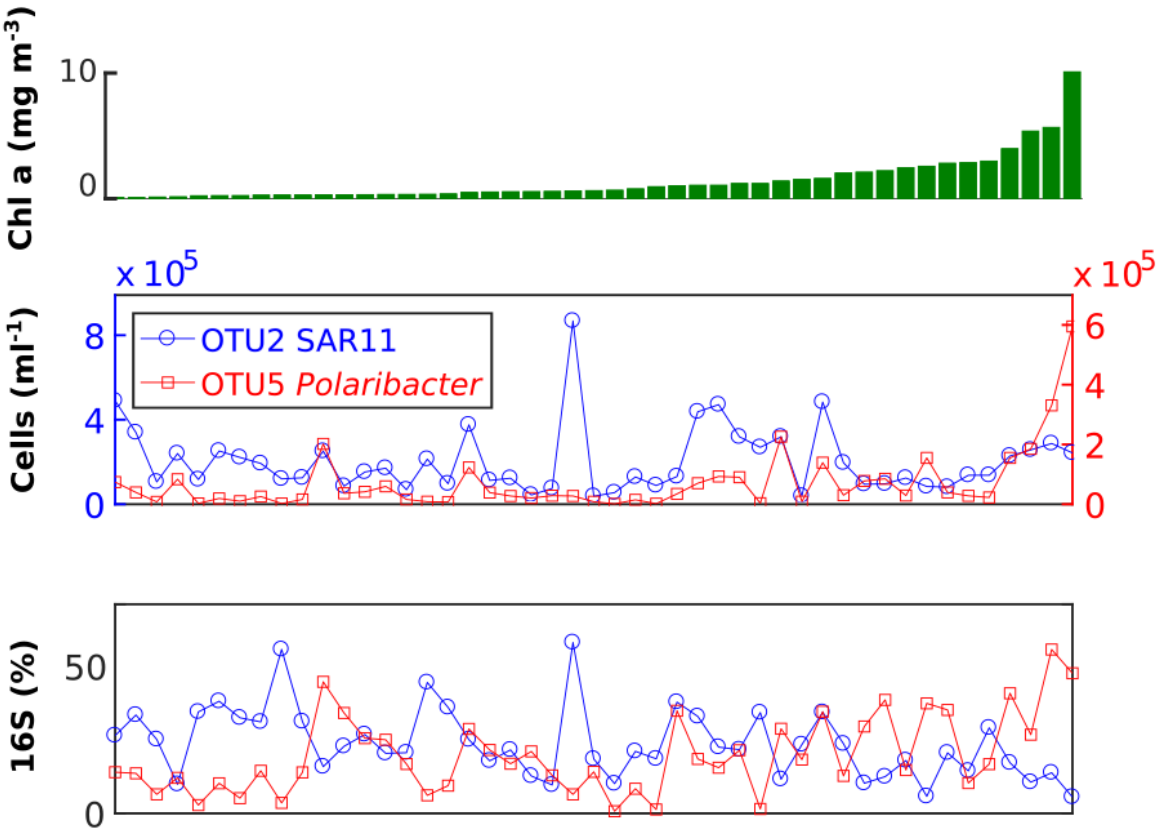
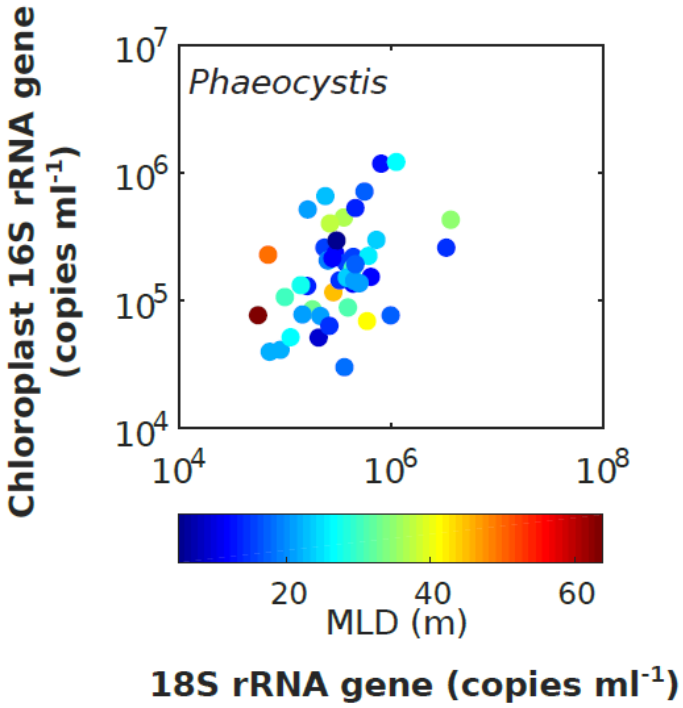


Figure S1. QMP and RMP for two representative OTUs ranked by ascending Chl a concentrations. QMP (top) in cells ml^{-1} and RMP (bottom) in % for OTU2 (SAR11) in blue circles and OTU5 (*Polaribacter*) in red squares.



95

96 **Figure S2.** *Phaeocystis* abundances chloroplast 16S vs. genomic18S rRNA gene. Data points

97 are color-coded by mixed layer depth ($\Delta\sigma_\theta$ = 0.125 kg m^{-3}).

98

Table S1. The CV (%) of taxa abundances estimates from duplicate samples Coastal 2A and Costal 2B in Figure 1B. Internal standards at two different concentrations (high and low) were spiked into each sample.

	CV(high)	CV(low)
Cryptophytes	0.5	0.1
Haptophytes	2.6	4.6
Alveolata	2.5	4.1
Rhizaria	12.3	1.8
Stramenopiles	1.6	2.6
Others	0.5	1.0

Table S2. A two-component example of the unknown rrn effect on the FCM normalization method. Estimates (the bottom row) with more than 100% off the true values are colored in red. The rrn for *Oceanosirillaceae* is estimated as the genus averaged rrn from rrnDB. Note that this calculation does not account for multiple genome per cell which can also affect the FCM normalization results.

		Sample S1	Sample S2	Ratio S1/S2
True Cell number	**SAR11**	1	10	0.1
	Oceanospirillaceae	10	1	10
FCM counts		11	11	
16S rRNA gene per cell	**SAR11**	1	1	
	Oceanospirillaceae	5.4	5.4	
16S rRNA gene counts	**SAR11**	1	10	
	Oceanospirillaceae	54	5.4	
16S rRNA gene percentage	**SAR11**	2%	65%	
	Oceanospirillaceae	98%	35%	
Calculated cell number (FCM × 16S rRNA%)	**SAR11**	**0.2**	7.1	**0.028**
	Oceanospirillaceae	10.8	**3.9**	**2.8**

Table S3. Chloroplast 16S and genomic 18S rRNA gene linear correlations (type-II least-square fit): correlation coefficient R^2, slope with standard deviation (SD) and y-intercept with SD for Cryptophytes, *Fragilariopsis* (diatom), *Corethron* (diatom), *Proboscia* (diatom), and *Phaeocystis*.

	Cryptophytes	***Fragilariopsis***	***Corethron***	***Proboscia***	***Phaeocystis***
R^2	0.87	0.55	0.72	0.40	0.064
Slope	1.1	0.67	1.3	1.3	0.36
SD	0.058	0.072	0.11	0.17	0.066
y-intercept	9.10×10^4	-3.02×10^4	4.23×10^4	-7.70	5.77×10^4
SD	1.36×10^5	3.03×10^4	2.07×10^4	6.12×10^2	5.60×10^4

The chloroplast 16S rRNA counts and the genomic 18S rRNA counts are within the same order of magnitude. Phytoplankton could have much less chloroplasts per cell than genomic 18S rRNA gene copies per cell. For example, pennate diatoms generally have 1 or 2 large chloroplasts per cell (19) and hundreds of 18S rRNA copies per cell (20). Each chloroplast contains multiple chloroplast DNA (cpDNA) and some phytoplankton could have up to 650 copies of cpDNA per chloroplast (21). Assuming a cpDNA number around 10^2 (which is typical in higher plants and as reported in a phytoplankton study (21)), we would expect chloroplast 16S and genomic 18S rRNA gene copies per cell in the same order of magnitude.

References

1. Kozlowski WA, Deutschman D, Garibotti I, Trees C, Vernet M. 2011. An evaluation of the application of CHEMTAX to Antarctic coastal pigment data. Deep Sea Res Part I Oceanogr Res Pap 58:350–364.

2. Mackey MD, Mackey DJ, Higgins HW, Wright SW. 1996. CHEMTAX—a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. Mar Ecol Prog Ser 265–283.

3. Smith DC, Azam F. 1992. A simple, economical method for measuring bacterial protein synthesis rates in seawater using 3H-leucine. Mar Microb Food Webs 6:107–114.

4. Luria CM, Amaral-Zettler LA, Ducklow HW, Rich JJ. 2016. Seasonal Succession of Free-Living Bacterial Communities in Coastal Waters of the Western Antarctic Peninsula . Front Microbiol .

5. Bowman JS, Amaral-Zettler LA, J Rich J, M Luria C, Ducklow HW. 2017. Bacterial community segmentation facilitates the prediction of ecosystem function along the coast of the western Antarctic Peninsula. Isme J 11:1460.

6. Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 18:1403–1414.

7. Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. 2011. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 27:1159–1161.

8. Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU

rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton.

9. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci 108:4516–4522.

10. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596.

11. Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, BREINER H, Richards TA. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol 19:21–31.

12. Lin Y, Cassar N, Marchetti A, Moreno C, Ducklow H, Li Z. 2017. Specific eukaryotic plankton are good predictors of net community production in the Western Antarctic Peninsula. Sci Rep 1.

13. Bystrykh L V. 2012. Generalized DNA barcode design based on Hamming codes. PLoS One 7:e36852.

14. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

15. Edgar RC, Flyvbjerg H. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476–3482.

16. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

17. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK,

Fierer N, Pena AG, Goodrich JK, Gordon JI. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336.

18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

19. Round FE, Crawford RM, Mann DG. 1990. Diatoms: biology and morphology of the genera. Cambridge university press.

20. Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, Karunasagar I. 2008. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Appl Environ Microbiol 74:7174–7182.

21. Ersland DR, Aldrich J, Cattolico RA. 1981. Kinetic complexity, homogeneity, and copy number of chloroplast DNA from the marine alga Olisthodiscus luteus. Plant Physiol 68:1468–1473.