**Supplementary Material**

**Coastal upwelling systems as dynamic mosaics of bacterioplankton functional specialization**

Erick Delgadillo-Nuño1, Eva Teira1, Benjamin Pontiller2,4, Daniel Lundin2, Vanessa Joglar1, Carlos Pedrós-Alió3, Emilio Fernández1, Jarone Pinhassi2, Sandra Martínez-García1\*

1Centro de Investigación Mariña da Universidade de Vigo (CIM-UVigo), Departamento de Ecoloxía e Bioloxía Animal, Vigo, Spain.

2Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, Kalmar, Sweden.

3Centro Nacional de Biotecnología (CNB-CSIC), Campus de Cantoblanco, Madrid, Spain.

4 GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.

**Supplementary methods**

**Study site**

The upwelling system of the Iberian Peninsula (NW Spain) is characterized by the inflow of subsurface oceanic Eastern North Atlantic Central Water (ENACW) from March to September (when northerly winds dominate) [1]. During autumn and winter, downwelling conditions prevail due to the predominance of south and southwest winds [2]. During spring and summer, coinciding with the input of upwelled nutrients, seawater temperature, and irradiance levels increase.

**Sampling**

Seawater for this study was collected during the three ENVISION 10-day cruises conducted (CTM2014-59031-P, PI: Eva Teira [3]) on board the R/V *Ramón Margalef* in February (ENVISION I; winter), April (ENVISION II; spring), and August (ENVISION III; summer) 2016 to cover a wide range of initial hydrographic and ecological conditions. The first cruise was carried out from 17 to February 26, the second cruise was carried out from 16 to April 25, and the last cruise was carried out from 5 to August 14. Coinciding with the spring bloom and the early and late summer upwelling. Samples were taken at two locations in the eastern Atlantic Ocean, in the upwelling system near the Ría de Vigo (coastal station; 88 m depth, 42º N, 8.88º W, and offshore station; 260 m depth, 42º N, 9.06º W). Water was collected with 20L Niskin metal-free bottles mounted on a rosette sampler (equipped with CTD sensors) from water from 5 m. The pooled water was pre-filtered through a 200 µm filter net.

**Hydrographic survey**

Vertical profiles of temperature (°C), salinity (PSU), turbidity (NTU), total chlorophyll fluorescence, and photosynthetically active radiation (PAR) were obtained using a Seabird CTD rosette down to 60 m in the coastal station and 200 m in the offshore station. Aliquots of 50 ml for nitrite (NO₂⁻), nitrate (NO3-), ammonium (NH₄⁺), phosphate (PO₄³⁻), and silicate (SiO2) were collected directly from the Niskin bottle in polyethylene bottles using contamination-free plastic gloves before all other variables to avoid contamination. Samples were analyzed with a Bran + Luebbe segmented flow analyzer [33]. The detection limit was 0.1 µmol L−1 for NO3-, 0.02 µmol L−1 for NO2- and PO₄³⁻, and 0.05 µmol L−1 for NH₄⁺ and SiO2- µmol L−1.

**Chlorophyll a**

Chlorophyll a (Chl-*a*) concentration was measured as a phytoplankton biomass proxy. A volume of 300 mL of water samples was filtered through 0.2 µm polycarbonate filters, and the filters were immediately frozen at −20°C until further analysis. Chl-*a* was extracted with 90 % acetone and kept in darkness at 4°C overnight. Fluorescence was determined with a TD-700 Turner Designs fluorometer calibrated with pure Chl-*a* standard solution (absorption coefficient at 663 nm = 87.7 [4].

**Prokaryotic abundance and biomass**

Seawater samples (2 mL) were preserved by adding 1% paraformaldehyde +0.05% glutaraldehyde, incubated for 20 min, quick-frozen by submerged in liquid nitrogen, and stored at −80°C. Prokaryotic abundance was calculated by measuring the signature of side scatter (SSC) and green fluorescence of the stained cells (SybrGreen staining) as described by Gasol and Del Giorgio [5]using a Becton Dickinson FACSCalibur flow cytometer (488‐nm light produced by argon‐ion laser). Prokaryotic biomass was estimated by measuring the SSC and cell diameter to calculate cell biovolume (BV) as described by Calvo-Díaz and Moran [6]. BV was converted into biomass using the allometric factor of Norland (1993: fg C cell−1 = 120×BV0.72) for the coastal experiments and the open-ocean conversion factor for the oceanic experiments (fg C cell−1 = 350×BV).

**Microbial community composition**

A total of 24 DNA samples were taken at the surface in the coastal and oceanic stations on day 1, day 3, day 5, and day 7 of each cruise. Community composition was assessed by sequencing the 16S rRNA gene (16S rDNA) for prokaryotes and the 18S rRNA gene (18S rDNA) for eukaryotes. Water samples were sequentially filtered through 3 µm pore size polycarbonate filters (Millipore, Italy) and a 0.2 µm pore size Sterivex (Millipore, Italy) filter and immediately frozen in liquid nitrogen and conserved at −80°C. DNA retained in the 3 and 0.2 µm filters were extracted using the PowerSoil DNA isolation kit (MoBio Laboratories Inc., CA, USA) and the PowerWater DNA isolation kit (MoBio Laboratories Inc., CA, USA), respectively, according to the manufacturer’s instructions. Prokaryotic DNA from 0.2 µm filters was amplified using the universal primers 515F and 926R [7] and eukaryotic DNA from 3 and 0.2 µm filters, using the primers TAReuk454FWD1 and TAReukREV3 [8]. Amplified regions were sequenced in an Illumina Miseq platform, with a sequencing depth of 10 million reads, at the Research and Testing Laboratory (Lubbock, TX, USA). The sequences obtained were analyzed with the Software DADA2 for amplicon sequence variants (ASVs) [9] using the SILVA reference database for taxonomic assignment of 16S [10], and the databases PR2 [11] and the marine protist from the BioMarKs project [12] for taxonomic assignment of 18S ASVs as described in [3]. The raw ASV tables of prokaryotes and eukaryotes were subsampled to the number of reads present in the sample with the lowest number of reads, 2080 and 1286, for 16S rDNA and 18S rDNA, respectively. The sequence abundances of the subsampled ASV tables were transformed using the centred log ratio (clr) [13, 14]. Before the clr transformation, zeros were replaced by the minimum value larger than 0 divided by 2 [15–17].

**Metatranscriptomic analyses: prokaryotic community gene expression**

A total of 24 RNA samples were taken at the surface in the coastal and oceanic stations on day 1, day 3, day 5, and day 7 of each cruise at the surface in the coastal and oceanic stations. Samples for community gene expression analysis were taken on days 1, 3, 5, and 7 of each cruise by sequentially filtering the water through 3 µm pore size polycarbonate filters (Millipore, Italy) and a 0.2 µm pore size Sterivex filter (Millipore, Italy) and immediately preserved in RNAlater (Invitrogen), flash-frozen in liquid nitrogen and stored at −80°C until further processing in the laboratory. The total RNA of the free-living size fraction (<3 µm >0.22 µm) was extracted using the Qiagen RNeasy Mini Kit (Qiagen). RNA was then sequenced on a Illumina HiSeq 2500 platform, with a sequencing depth of 70 million reads, at the Science for Life Laboratory in Stockholm in 2016 (SciLifeLab; [www.scilifelab.se](http://www.scilifelab.se)) in rapid mode and with v3 chemistry to obtain 2 × 125 bp paired-end reads. The sequences obtained were quality checked with FastQC [18] and MultiQC [19], adapter-primer sequences were removed with Cutadapt (v1.13) [20], and reads were trimmed with Sickle using default settings (v1.33) [21]. The stable RNA reads were filtered out by aligning all reads to an in-house database of stable RNA sequences with ERNE (v2.1.1) [22]. Remaining quality reads were then de-novo assembled with MEGAHIT (v1.1.2) [23]. Open reading frames (ORFs) were determined with Prodigal (v2.6.3) [24] in single mode, and the nucleotide sequences from all assemblies were clustered at a 99% level with VSEARCH (v2.20.1) [25]. To quantify ORFs, quality reads were mapped to the ORF sequences with Bowtie2 (v2.3.5.1) [26]. Taxonomic annotations were assigned to ORFs by alignment with Diamond (v0.9.24) [27] to NCBI’s RefSeq protein database (release date: December 20, 2018) [28] and subsequent post-processing with MEGAN (v6.7.3) [22]. Subsequently, 200 protein families were detected and classified among the ORFs with HMMER3 using HMM profiles to select ecologically relevant genes (Pfams and Tigrfams, Supp. Table 1). To make comparisons between samples, considering that the transcripts have different lengths and the libraries differ in size, we used a Transcript Per Million (TPM) normalization. We observed that the TIGR01151 profile of the photosystem II protein (psbA) was assigned to incorrect taxonomic groups. Therefore, non-Cyanobacteria taxa were filtered out for this gen.

**Statistics, normalizations, and visualization**

Redundancy Analysis (RDA) was performed using all ORFs (1 611 755) with the *rda* function. RDA plots were constructed with *ggplot2* (v 3.4.1) [29] and *ggord* (v 1.1.7) [30] packages. Zero values of raw counts were replaced with 0.5. Subsequently, centred log ratios (clr) were computed with the *codaSeq.clr* function from the *CoDaSeq* (v 0.99.6) package [31] and Euclidean distances calculated with the *vegdist* function of *vegan* (v 2.6.4) package [32]. Environmental variables were selected based on PERMANOVA analysis to evaluate significant differences between seasons and sampling stations. The significance of the explanatory variables was examined by permutation analysis (permutations = 999) [33]. We performed heatmap clustering analysis based on Euclidean distances to visualize strong correlations (p=0.01) between environmental variables, the taxonomic composition of the microbial community, and procaryotic activity. Heatmaps were constructed using the ComplexHeatmap package (v 2.14.0) [34]. Spearman correlations were calculated using the cor function of the *Hmisc* (v 4.7.2) package [35]. To visualize the short-term (daily) changes in the most important or abundant genes, we represented in bubble plots the TPM values as a ratio of each gene over the sampling days (1, 3, 5, and 7) divided by the mean of all days. Allowing us to simultaneously visualize the relative abundance of specific prokaryotic genes and the daily changes patterns.

**Supplementary references**

1. Álvarez-Salgado XA, Figueiras FG, Fernández-Reiriz MJ, Labarta U, Peteiro L, Piedracoba S. Control of lipophilic shellfish poisoning outbreaks by seasonal upwelling and continental runoff. *Harmful Algae* 2011; **10**: 121–129.

2. Figueiras FG, Labarta U, Fernández Reiriz MJ. Coastal upwelling, primary production and mussel growth in the Rías Baixas of Galicia. *Hydrobiologia* 2002; **484**: 121–131.

3. Joglar V, Prieto A, Barber-Lluch E, Hernández-Ruiz M, Fernández E, Teira E. Spatial and temporal variability in the response of phytoplankton and bacterioplankton to B-vitamin amendments in an upwelling system. *Biogeosciences Discuss* 2020; 1–43.

4. Lorenzen CJ, Newton Downs J. The specific absorption coefficients of chlorophyllide a and pheophorbide a in 90% acetone, and comments on the fluorometric determination of chlorophyll and pheopigments. *Limnol Oceanogr* 1986; **31**: 449–452.

5. Gasol JM, Del Giorgio PA. Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. *Sci Mar* 2000; **64**: 197–224.

6. Calvo-Díaz A, Morán XAG. Seasonal dynamics of picoplankton in shelf waters of the southern Bay of Biscay. *Aquat Microb Ecol* 2006; **42**: 159–174.

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

8. Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, et al. Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* 2014; **24**: 813–821.

9. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13**: 581–583.

10. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 2013; **41**: 590–596.

11. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 2013; **41**.

12. Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, et al. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ Microbiol* 2015; **17**: 4035–4049.

13. Fernandes, D. A, Reid, J., Macklaim, M. J, et al. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2014; **2**: 1–13.

14. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: And this is not optional. *Front Microbiol* 2017; **8**: 1–6.

15. Aitchison J. The Statistical Analysis of Compositional Data. *J R Stat Soc Ser B* 1982; **44**: 139–160.

16. Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Math Geol* 2003; **35**: 253–278.

17. Joglar V, Prieto A, Barber-Lluch E, Hernández-ruiz M, Fernández E, Teira E. Spatial and temporal variability in the response of phytoplankton and bacterioplankton to B-vitamin amendments in an upwelling system. *Biogeosciences Discuss* 2020; 1–43.

18. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010.

19. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; **32**: 3047–3048.

20. Marcel M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011; **17**: 10–12.

21. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. 2011. , [Software]

22. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An extensive evaluation of read trimming effects on illumina NGS data analysis. *PLoS One* 2013; **8**: 1–13.

23. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015; **31**: 1674–1676.

24. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Integrated nr Database in Protein Annotation System and Its Localization. *Nat Commun* 2010; **6**: 1–8.

25. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* 2016; **2016**: 1–22.

26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**: 357–359.

27. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2014; **12**: 59–60.

28. O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016; **44**: D733–D745.

29. Wickham H, Chang W. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* 2014; **2**: 1–189.

30. Moshinsky M. ggord: Ordination Plots with ggplot2. 2022.

31. Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It’s all relative: analyzing microbiome data as compositions. *Ann Epidemiol* 2016; **26**: 322–329.

32. Oksanen r J, Simpson GL, Blanchet FG, Solymos P, Stevens MHH, Szoecs E, et al. vegan: Community Ecology Package. *Vegan Community Ecol Packag* 2022.

33. Borcard D, Gillet F, Legendre P. Numerical ecology with R. 2011.

34. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016; **32**: 2847–2849.

35. Miscellaneous TH, Yes L. Hmisc: Harrell Miscellaneous. 2023.