

Marine Ultrasmall Prokaryotes Likely Affect the Cycling of Carbon, Methane, Nitrogen, and Sulfur

Romain Lannes[†] , Louise Cavaud, Philippe Lopez[†] , and Eric Bapteste*

Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Muséum National d'Histoire Naturelle, EPHE, Université des Antilles, Paris, France

*Corresponding author: E-mail: eric.bapteste@upmc.fr.

Accepted: 11 December 2020

[†]These authors contributed equally to this work.

Abstract

Recently, we uncovered the genetic components from six carbon fixation autotrophic pathways in cleaned ultrasmall size fractions from marine samples (<0.22 μm) gathered worldwide by the *Tara Oceans* Expedition. This first finding suggested that prokaryotic nanoorganisms, phylogenetically distantly related to the known CPR and DPANN groups, could collectively impact carbon cycling and carbon fixation across the world's ocean. To extend our mining of the functional and taxonomic microbial dark matter from the ultrasmall size fraction from the *Tara Oceans* Expedition, we investigated the distribution of 28 metabolic pathways associated with the cycling of carbon, methane, nitrogen, and sulfur. For all of these pathways, we report the existence not only of novel metabolic homologs in the ultrasmall size fraction of the oceanic microbiome, associated with nanoorganisms belonging to the CPR and DPANN lineages, but also of metabolic homologs exclusively found in marine host taxa belonging to other (still unassigned) microbial lineages. Therefore, we conclude that marine nanoorganisms contribute to a greater diversity of key biogeochemical cycles than currently appreciated. In particular, we suggest that oceanic nanoorganisms may be involved in a metabolic loop around Acetyl-CoA, have an underappreciated genetic potential to degrade methane, contribute to sustaining redox-reactions by producing Coenzyme F420, and affect sulfur cycling, notably as they harbor a complete suite of homologs of enzymes of the SOX system.

Key words: metagenomics, marine ultrasmall organisms, metabolism, microbial dark matter.

Significance

The contribution of marine microorganisms to biogeochemical cycles, involving sulfur, nitrogen, methane, and carbon, and especially that of ultrasmall oceanic prokaryotes is receiving increasing attention, with the recent discovery of ultrasmall prokaryotic taxa that encode metabolic genes, for instance involved in carbon fixation in the world's ocean. However, the extent of this potential contribution remains likely underappreciated. We report novel, massive evidence that, in addition to known ultrasmall bacteria and archaea, uncultivated prokaryotes thriving in the oceans, belonging to microbial dark matter, encode a rich phylogenetic diversity of homologs of metabolic genes from pathways associated with nitrogen, sulfur, carbon, and methane metabolisms. Therefore, uncultivated oceanic ultrasmall prokaryotes likely take part to elemental cycling in many more ways than previously suspected.

Elemental cycling is crucial for supporting Life on Earth, and is impacted by microbes (Anantharaman et al. 2016; Escalas et al. 2019; Tanentzap et al. 2019; Logares et al. 2020). Consistently, genes involved in the sulfur, nitrogen, carbon, and methane biogeochemical cycles are of increasing interest,

and dozens of associated metabolic pathways have been documented in the KEGG database. Because genes associated with these KEGG pathways mostly come from cultured organisms, the diversity and the geographical distribution of the taxa contributing to these metabolisms in nature is likely

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

underestimated. In particular, the metabolic potential of uncultivated microbes, referred to as microbial dark matter, is still largely underexplored (Rinke et al. 2013; Brown et al. 2015; Castelle et al. 2015; Parks et al. 2017), especially within the ultrasmall size fraction (below 200 nm). Indeed, this size fraction has long been considered not to encompass any microbial cells, only viruses. The recent discovery of ultrasmall bacteria (the CPR group) and ultrasmall archaea (within the DPANN group) has challenged this view and revealed novel ecological and evolutionary players. For example, CPR and DPANN lineages were recently proposed to contribute to carbon fixation (Wieder et al. 2015; Dykma et al. 2016; Guidi et al. 2016; La Cono et al. 2018; Jaffe et al. 2019). Moreover, Anantharaman et al. detected the presence of key enzymes involved in carbon, nitrogen, sulfur, and hydrogen cycles in CPR from aquifers (Colorado, USA), where Rubisco type II/III genes were also found (Wrighton et al. 2016). More recently, our broad scale study of the gene content from marine ultrasmall size fractions samples reinforced this conclusion, and proposed that phylogenetically diverse marine ultrasmall prokaryotes, hereafter called nanoorganisms, could play an important role in carbon fixation (Lannes et al. 2019; McGrath 2019). Here, we extended our former analysis, using the same methodology as in Lannes et al. (2019), to retrieve putative additional metabolic homologs of the sulfur, nitrogen, carbon, and methane biogeochemical cycles. We used enzymes of KEGG pathways (as of October 2019) associated with sulfur, nitrogen, carbon, and methane as seeds to detect homologs in nanoorganisms from 65 sampling sites from the *TARA Oceans* metagenome data sets (Sunagawa et al. 2015). We also analyzed their geographical distribution and phylogenetic diversity, identifying metabolic homologs exclusive to the marine environment and hosted by lineages distinct from known CPR/DPANN lineages and different from any of the complete genomes present in NCBI, likely encoded within the genomes of the ultrasmall fraction of microbial dark matter.

For this, we took advantage of the sorting and decontamination (i.e., the removal of sequences from known viruses and regular-sized prokaryotes) realized by Lannes et al. (2019) for the publicly available, predicted ORFs from the ultrasmall size fraction (<0.22 μm) of the *TARA Oceans* metagenomic database. In brief, sequences found in the ultrasmall size fraction but not exclusively in that size fraction define a "Potentially ultra-small" (PU) data set, comprising 6,119,497 sequences. We then filtered these sequences into two nested data sets: the UO data set (for "Ultra-small Only," 4,300,092 sequences) comprises sequences exclusively found in ultrasmall size fractions, and the more stringent WUO data set ("Widespread Ultra-small Only," 1,128,306 sequences) contains sequences exclusively found in the ultrasmall size fraction at more than one sampling sites across the world's ocean. We also took advantage of the fact that Lannes et al. had further assigned these microbial sequences to two levels of microbial dark matter, based on their

percentage of identity (%ID) to their best hits in nr (NCBI database) and also against an exhaustive CPR/DPANN database, using DIAMOND (Buchfink et al. 2015). We classified as "light dark matter," the sequences that showed a mutual coverage of >80% and a %ID <90% with their best hit in nr, that is, sequences that were <90% similar to potential homologs in known taxa. Strongly divergent sequences, that is, with >80% mutual coverage but <70% %ID with their best hit in nr, were classified as "dark matter." Such a distinction allowed to contrast two thresholds of %ID while also counterbalancing the risk of focusing on very divergent sequences only. We distributed across both "dark matter" and "light dark matter" the sequences that were taxonomically assigned to DPANN, CPR, unclassified bacteria, unclassified archaea, candidate or "root: unassigned," because these taxa likely correspond to bona fide ultrasmall, yet uncultivated prokaryotes.

A list of KEGG Orthology (KO) terms corresponding to 28 metabolic pathways associated with Methane, Sulfur, Carbon, and Nitrogen cycling (12, 3, 6, 7 pathways, resp.) was defined using the KEGG database (Ogata et al. 1999) (supplementary table 1, Supplementary Material online). As in Lannes et al., all proteins involved in those pathways were retrieved using the Uniprot mapping tool (<http://www.uniprot.org/mapping/>) or the KEGG API service (March 2017). Homologs to each of these KEGG proteins were then identified in the PU, UO, and WUO data sets using NCBI BLAST (version 2.6.0) (Camacho et al. 2009), using the following criteria, which allow for potentially distant homologs: %ID > 25%, *E*-value < 1e-5, and mutual alignment coverage > 80%. KEGG pathways being composed of enzymes, we estimated the completeness of a given pathway as the percentage of the required enzymes for which homologs were found in our data sets (Lannes et al. 2019).

The completeness of each of metabolic pathway was represented using heatmaps, for the UO (fig. 1) and for the WUO (fig. 2) data sets, which contain sequences exclusively associated with the ultrasmall size fraction of the oceanic microbiome. Moreover, we classified homologs of reference metabolic genes in a gradable way to improve the description of the oceanic genetic microbial dark matter (i.e., gene that can be assigned to known families, although they clearly diverged from known representatives). Identifying homologs in the "light dark matter" hints at environmental variants of metabolic genes that are closer to known genes, in some sense, exciting but less surprising than the homolog variants associated with the "dark" matter. Contrasting these results shows the variability of these gene families in the oceans. Each heatmap shows the global geographic and stratigraphic distribution of the 28 tested metabolic pathways, sorted by their contribution to the metabolisms of sulfur, nitrogen, methane and carbon. This broad view of the metabolisms encoded by the genetic material from the ultrasmall oceanic microbiome reveals a heterogeneous distribution of these metabolisms

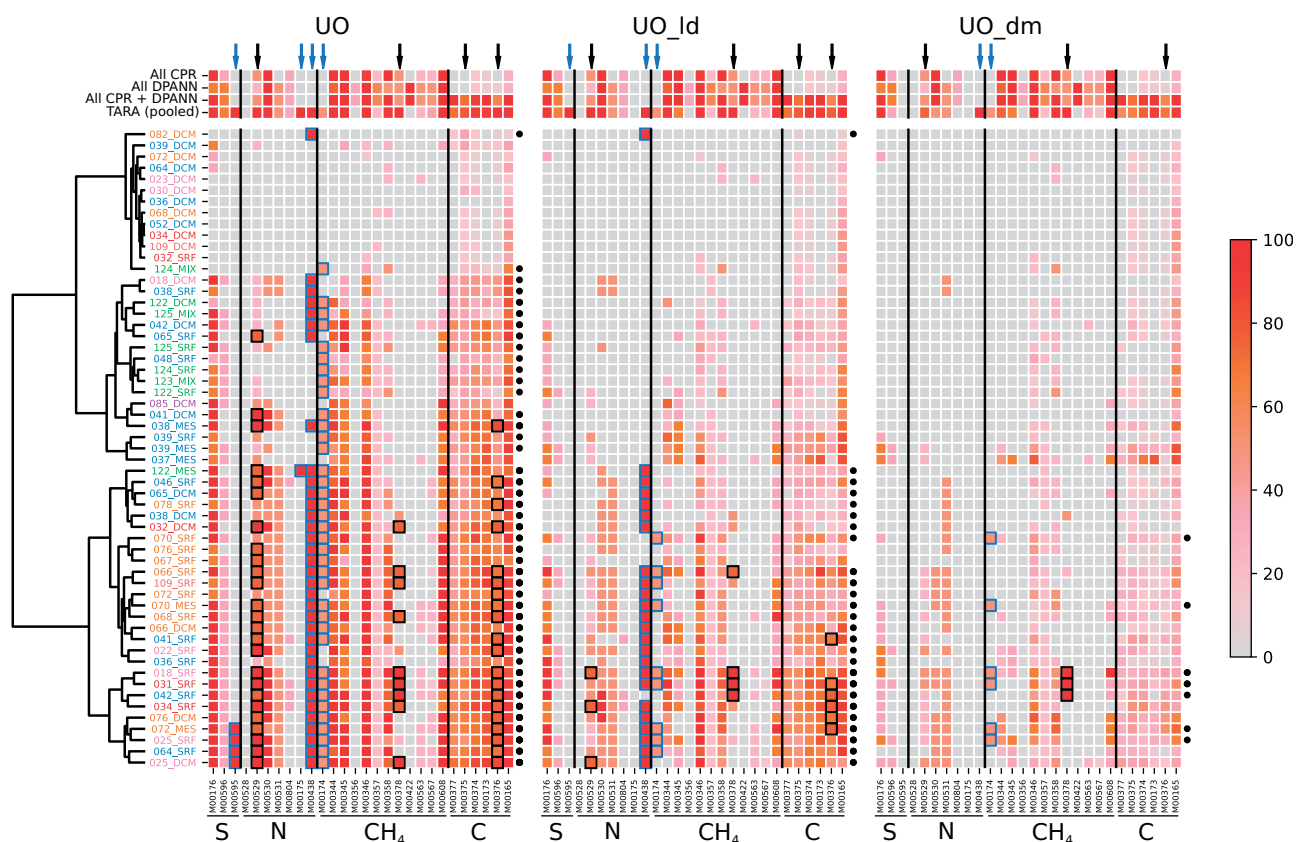


Fig. 1.—Heatmap showing the general distribution of 28 metabolic pathways in the World Ocean, for the UO data set. Top rows correspond to pooled data sets of known ultrasmall prokaryotic taxa and of the TARA OCEANS data set under study. Clustered rows (using `scipy.cluster.hierarchy.linkage` [“ward” method], dendrogram shown to the left of the heatmap) correspond to sampling sites in the format: TARA sampling site id (three digits) _ depth. Depths are: SRF (Surface), DCM (Deep Chlorophyll Maximum), MES (Mesopelagic), and MIX (mixed). Row label colors represent oceanic regions: brown for North Pacific Ocean, green for South Pacific Ocean, purple for Southern Ocean, orange for South Atlantic Ocean, dark blue for Indian Ocean, red for Red Sea, pink for Mediterranean Sea. Columns correspond to KEGG pathways, grouped by the biogeochemical cycle to which they contribute (S for Sulfur, N for Nitrogen, CH₄ for methane, and C for carbon). Cell colors reflect the % of pathway completeness at each site, as indicated by the colored bar on the right. Black squares indicate marine samples in which pathway completeness is greater than in the known CPR+DPANN data set. Blue squares indicate samples for which the pathway is exclusively found in the oceanic samples and not in CPR/DPANN. Black and blue arrows stress the column with boxed cells, whereas black dots indicate the corresponding samples. From left to right: Pathway completeness in the UO data set for all sequences, for sequences assigned to the light dark matter and for sequences assigned to the dark matter.

across sampling sites, both for the UO and the WUO data sets. Remarkably, many sites harbor near complete metabolic pathways (completeness over 80%) that are encoded by the ultrasmall microbial size fraction, whereas other sites lack such a metabolic potential. In those metabolically competent sites, the contribution of nanoorganisms thus appears to deserve further investigation to quantify their actual contribution to elemental cycling. Furthermore, boxed cells and black dots in our heatmaps (figs. 1 and 2) indicate sites at which pathway completeness in ultrasmall microbes was greater in the oceanic data set than reported in the known CPR and DPANN genomes. Therefore, such oceanic sites suggest novel, underappreciated metabolic contributions from environmental ultrasmall prokaryotes, with respect to the metabolic potential of known CPR and DPANN lineages, to the cycling of nitrogen, sulfur, carbon, and methane. For example, the Nitrate/

nitrite transport system pathway and M00378 (F420 biosynthesis) pathway constitute striking examples of the possible role of genes from ultrasmall microbial dark matter, since divergent homologs of all their constitutive enzymes were found at multiple individual sampling sites. Interestingly, all the sites with ≥ 5 pathways with more complete pathways in ultrasmall oceanic microbes than in known CPR and DPANN are either from the Surface or from the Deep Chlorophyll Maximum layer (DCM). In particular, the Tara station 025_DCM (Black Sea province, DCM layer) is the richest in metabolic pathways encoded by ultrasmall marine prokaryotes, and is thus a prime candidate for further genomic investigation of their impact on biogeochemical cycling.

Overall, for the UO and the WUO data sets, seven metabolic pathways (M00529 [Denitrification, nitrate \Rightarrow nitro-], M00595 [Thiosulfate oxidation by SOX complex,

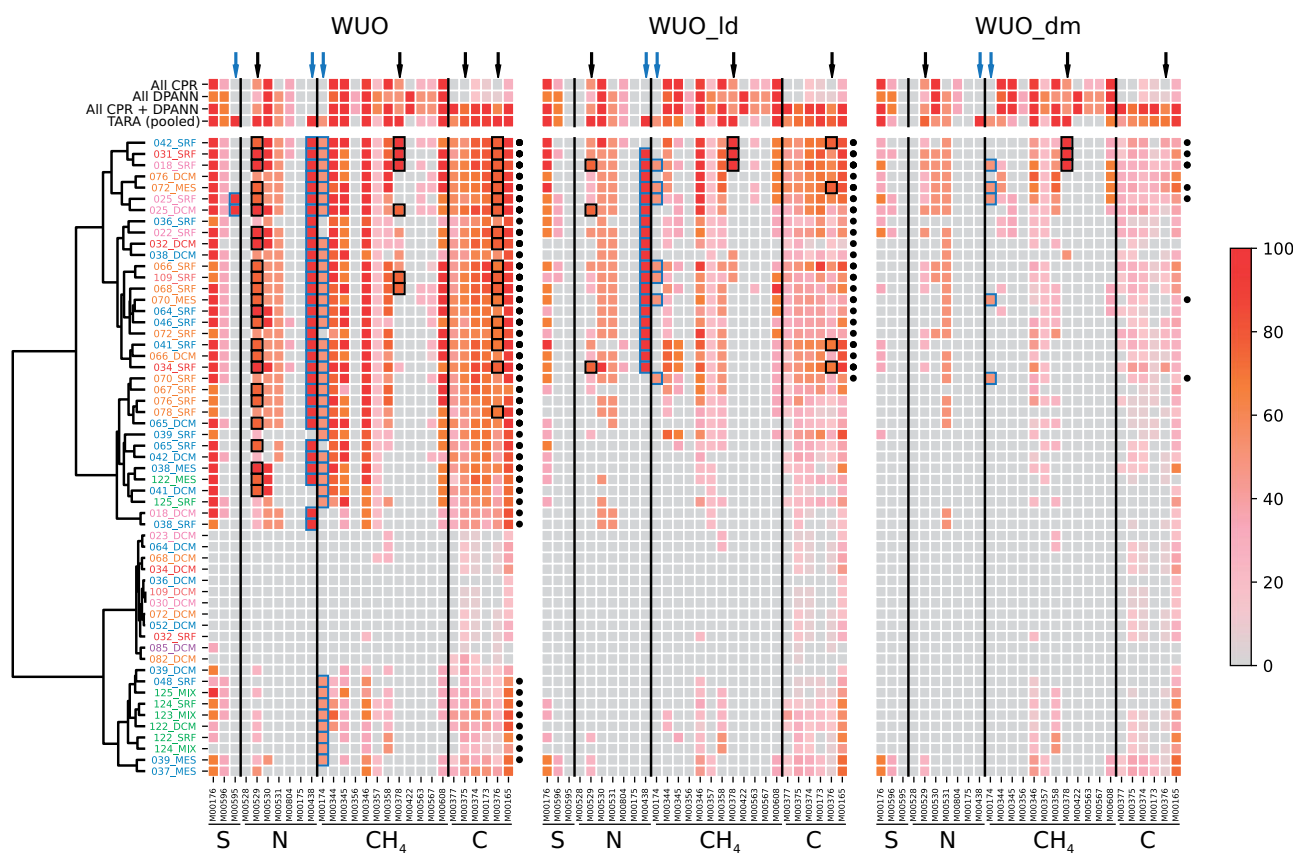


FIG. 2.—Heatmap showing the general distribution of 28 metabolic pathways in the World Ocean, for the WUO data set. Legend is the same as above. From left to right: Pathway completeness in the WUO data set for all sequences, for sequences assigned to the light dark matter and for sequences assigned to the dark matter.

thiosulfate => sulfate], M00376 [3-Hydroxypropionate bi-cycle], M00378 [F420 biosynthesis], M00438 [Nitrate/nitrite transport system], M00174 [Methane oxidation, methanotroph, methane => formaldehyde], M00175 [Nitrogen fixation, nitrogen => ammonia]) are more frequently found in ultrasmall oceanic microbes than in all known CPR/DPANN. Interestingly, these include three pathways (M00595, M00438, M00174) that are only found in oceanic samples for the WUO data set (and also in an additional exclusively marine pathway, M00175, from the UO data set). Our results thus highlight the novel and underappreciated metabolic potential of organisms captured in ultrasmall size fractions from the world's ocean. Collectively, oceanic nanoorganisms in both the UO and WUO data sets encode genes for all the enzymes associated to 16 pathways. Therefore, oceanic nanoorganisms, including those associated with two levels of dark matter, are predicted to interact with the S, N, CH₄, and C cycles.

Specifically, we recovered homologs for most known enzymes involved in carbon fixation in ultrasmall size fractions and at multiple sites (figs. 3 and 4). This more detailed view confirms the likely involvement of nanoorganisms in carbon

fixation, and expand the analysis of Lannes et al. (2019) to a broader set of CPR/DPANN lineages. Regarding their comparative distribution across nanoorganisms, genes encoding six enzymes involved in carbon fixation in reference taxa are exclusively found in known CPR/DPANN, whereas eight enzymes (6.4.1.1, 1.2.1.75, 1.1.1.298, 6.2.1.36, 4.2.1.116, 1.3.1.84, 4.2.1.148, 4.2.1.153) are only found in marine samples. These seven latter enzymes are all involved in the conversion of a derived product of Acetyl-CoA, Malonyl-CoA, to Propionyl-CoA, and in the functionally related conversion of a derived of Propionyl-CoA, B-Methylmalyl-CoA, in Acetyl-CoA. This prevalence of exclusively marine enzymes associated with the ultrasmall size fraction suggests a possible role of marine nanoorganisms in a metabolic loop around Acetyl-CoA in the oceans.

Our analyses also recovered homologs for most known enzymes connected to methane metabolism. In particular, genes encoding two enzymes (1.14.1325, 1.14.183) catalyzing the transition from methane to methanol were exclusively found in the marine ultrasmall size fraction, in the PU and UO data sets, respectively. Moreover, two out of three enzymes (1.1.2.7, 1.1.3.13) involved in the transition from methanol to

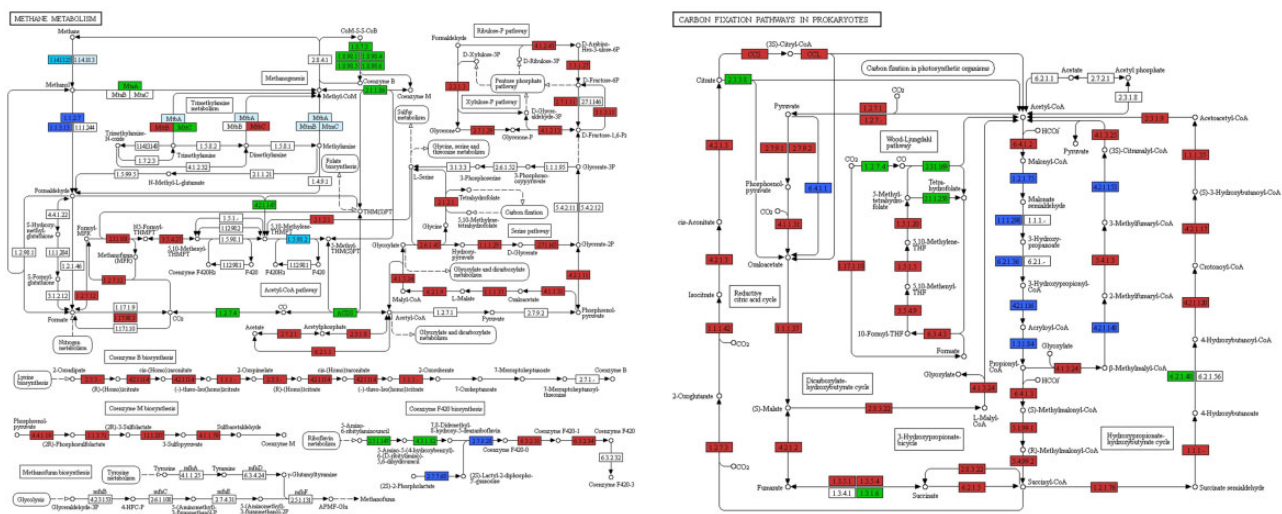


Fig. 3.—Methane Metabolism and Carbon fixation KEGG maps describing the distribution of homologs of metabolic enzymes across samples. The name of the considered metabolisms is indicated in an upper box. Enzymes are color coded based on their distribution. Red: present in the WUO data set and in CPR/DPANN; light blue: absent in CPR/DPANN yet present in the PU data set; blue: absent in CPR/DPANN yet present in the UO data set; dark blue: blue: absent in CPR/DPANN yet present in the WUO data set; Green: exclusively present in CPR/DPANN.

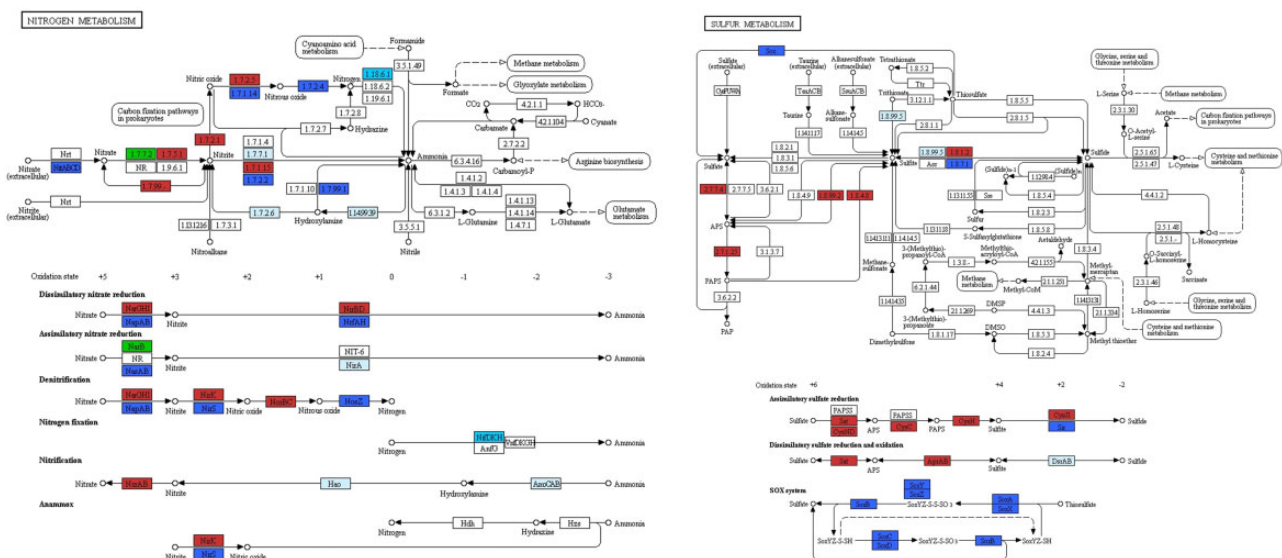


Fig. 4.—Nitrogen and sulfur metabolisms KEGG maps describing the distribution of homologs of metabolic enzymes across samples.

formaldehyde were also found in the WUO data set. Together, these results suggest that marine nanoorganisms have an underappreciated genetic potential to degrade methane. Two enzymes exclusively found in the UO data sets (2.7.8.28, 2.7.7.68) were also involved in the biosynthesis of Coenzyme F420-0, and a third enzyme also found in the UO data set (1.5.98.2) catalyzed the transition from 5,10-Methylene THMPT to 5-Methyl THM(S)PT, mediated by Coenzyme F420-H2, which also releases Coenzyme F420. These results suggest that marine nanoorganisms contribute to sustain this form of coenzyme, and therefore redox-reactions in the ocean. Of note, we also found homologs of

six enzymes in known CPR/DPANN, predicted to contribute to the synthesis of the coenzymes B and M. Overall, these analyses support an extended role for nanoorganisms in methane cycling.

Similarly, we report that nanoorganisms encode for multiple enzymes involved in nitrogen cycling. For example, homologs of enzymes involved in dissimilatory nitrate reduction (NarGH1, NirBD), in nitrification (NxrAB), and in denitrification (NarGH1, NirK, NorBC) are shared by both CPR/DPANN and marine nanoorganisms. In addition, other homologs involved in dissimilatory (NapAB, NrfAH) and assimilatory (NasAB), nitrate reduction, denitrification (NapAB, NirS, NosZ), and the

integration of extracellular nitrate (NrtABCD) are exclusively associated with oceanic ultrasmall microbial size fractions, and are found at multiple sites. Oceanic nanoorganisms may also affect sulfur metabolism. Genes encoding homologs of enzymes shared by known CPR/DPANN and by candidate marine nanoorganisms were found for Assimilatory Sulfate reduction, and Dissimilatory Sulfate reduction and oxidation. Furthermore, we identified a remarkably complete suite of homologs of enzymes of the SOX system (SoxA, SoxB, SoxC, SoxD, SoxX, SoxY, SoxZ) (Friedrich et al. 2005), exclusively found in marine nanoorganisms in our data set, and at multiple sites.

The phylogenetic position of all metabolic homologs was inferred by Maximum Likelihood analysis (using IQ-Tree [Trifinopoulos et al. 2016], LG+G model, and 1000 ultrafast bootstraps replicates [Minh et al. 2013]), for trimmed alignments (using MAFFT [Katoh et al. 2002] and trimAl [Capella-Gutierrez et al. 2009] with default parameters) of these metabolic homologs, which also include reference sequences from complete prokaryotic genomes and from all published CPR and DPANN genomes. Phylogenetic analyses of these metabolic genes further confirmed the rich phylogenetic diversity of hosts of metabolic homologs in the ultrasmall fraction, beyond known CPR/DPANN (available at <https://itol.embl.de/shared/TeamAire>, under Lannes_Cavaud_et al directory, last accessed December 23, 2020), in agreement with our former study (Lannes et al. 2019), as can be seen for instance in the trees of K05884 (N), K13039 (CH₄), and K08691 (CH₄) (supplementary fig. 1, Supplementary Material online). Moreover, some of the environmental sequences that qualified as “light dark matter” and as “dark matter” formed clades in these phylogenies, enhancing the described phylogenetic diversity for these metabolic enzymes and hinting at undescribed ultrasmall lineages within known major prokaryotic groups, which could take part in N, S, C, and CH₄ metabolisms.

Overall, the detection of genes associated with N, S, C, and CH₄ metabolisms, exclusive to the ultrasmall size fraction generated from the *TARA Oceans* project, encourages future single cell genome analyses and meta-transcriptomic studies to further characterize the precise mechanisms by which these formerly undetected sequences from nanoorganisms contribute to elemental cycling.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank G. Bernard and C. Bowler for critical comments. This work was granted access to the HPC resources of the

institute for computing and data sciences (ISCD) at University Pierre et Marie Curie. R.L., E.B., and L.C. were supported by FP7/2007-2013 Grant Agreement No. 615274.

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Literature Cited

- Anantharaman K, et al. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun.* 7(1):13219.
- Brown CT, et al. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523(7559):208–211.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 12(1):59–60.
- Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
- Castelle CJ, et al. 2015. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol.* 25(6):690–701.
- Dykstra S, et al. 2016. Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. *ISME J.* 10(8):1939–1953.
- Escalas A, et al. 2019. Microbial functional diversity: from concepts to applications. *Ecol Evol.* 9(20):12000–12016.
- Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J. 2005. Prokaryotic sulfur oxidation. *Curr Opin Microbiol.* 8(3):253–259.
- Guidi L, et al. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532(7600):465–470.
- Jaffe AL, Castelle CJ, Dupont CL, Banfield JF. 2019. Lateral gene transfer shapes the distribution of RuBisCO among candidate phyla radiation bacteria and DPANN archaea. *Mol Biol Evol.* 36(3):435–446.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30(14):3059–3066.
- La Cono V, et al. 2018. Contribution of bicarbonate assimilation to carbon pool dynamics in the deep Mediterranean Sea and cultivation of actively nitrifying and CO₂-fixing bathypelagic prokaryotic consortia. *Front Microbiol.* 9:3.
- Lannes R, Olsson-Francis K, Lopez P, Baptiste E. 2019. Carbon fixation by marine ultrasmall prokaryotes. *Genome Biol Evol.* 11(4):1166–1177.
- Logares R, et al. 2020. Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* 8(1):55.
- McGrath C. 2019. Highlight: contribution of ultra-small microbes to global carbon cycles. *Genome Biol Evol.* 11(5):1431–1431.
- Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 30(5):1188–1195.
- Ogata H, et al. 1999. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27(1):29–34.
- Parks DH, et al. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol.* 2(11):1533–1542.
- Rinke C, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459):431–437.
- Sunagawa S, et al. 2015. Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348(6237):1261359–1261359.

- Tanentzap AJ, et al. 2019. Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proc Natl Acad Sci USA*. 116(49):24689–24695.
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 44(W1):W232–235.
- Wieder WR, Cleveland CC, Lawrence DM, Bonan GB. 2015. Effects of model structural uncertainty on carbon cycle projections: biological nitrogen fixation as a case study. *Environ Res Lett*. 10(4):044016.
- Wrighton KC, et al. 2016. RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J*. 10(11):2702–2714.

Associate editor: Eyre-Walker, Adam