https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

Biogeosciences Discussions

Diversity and distribution of Nitrogen Fixation Genes in the Oxygen Minimum Zones of the World Oceans

[1]Amal Jayakumar and [1]Bess B. Ward

[1]Department of Geosciences
Princeton University
Princeton, NJ 08544

Abstract

Diversity and community composition of nitrogen fixing microbes in the three main oxygen minimum zones (OMZs) of the world ocean were investigated using operational taxonomic unit (OTU) analysis of *nifH* clone libraries. Representatives of the all four main clusters of *nifH* genes were detected. Cluster I sequences were most diverse in the surface waters and the most abundant OTUs were affiliated with Alpha- and Gammaproteobacteria. Cluster II, III, IV assemblages were most diverse at oxygen depleted depths and none of the sequences were closely related to sequences from cultivated organisms. The OTUs were biogeographically distinct for the most part – there was little overlap among regions, between depths or between cDNA and DNA. Only a few cyanobacterial sequences were detected. The prevalence and diversity of microbes that harbour *nifH* genes in the OMZ regions, where low rates of N fixation are reported, remains an enigma.

Introduction

Nitrogen fixation is the biological process that introduces new biologically available nitrogen into the ocean, and thus constrains the overall productivity of large regions of the ocean where N is limiting to primary production. The most abundant and most important diazotrophs in the ocean are members of the filamentous genus *Trichodesmium* and several unicellular

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

genera, including *Chrocosphaera sp*. and the symbiotic genus *Candidatus* Atelocyanobacterium thalassa (UCYN-A). Although these cyanobacterial species are wide spread and have different biogeographical distributions (Moisander et al. 2010), they are restricted to surface waters, mainly in tropical or subtropical regions.

Because diazotrophs have an ecological advantage in N depleted waters, and because those conditions occur in the vicinity of oxygen minimum zones, due to the loss of fixed N by denitrification, it has been proposed that N fixation should be favoured in regions of the ocean influenced by OMZs (Deutsch et al. 2007). The search for non cyanobacterial diazotrophs has resulted in discovery of diverse *nifH* genes, but they have not been associated with significant rates of N fixation (Moisander et al. 2017). It has also been suggested that the energetic constraints on N fixation might be partially alleviated under reducing, i.e., anoxic, conditions (Großkopf and LaRoche 2012). In response to these ideas, the search for organisms with the capacity to fix nitrogen has been focused recently in regions of the ocean that contain OMZs. That search usually takes the form of characterizing and quantifying one of the genes involved in the fixation reaction, *nifH*, which encodes the dinitrogenase reductase enzyme. Here we report on the distribution and diversity of *nifH* genes in all three of the world ocean's major OMZs, including samples from both surface and anoxic depths, and both DNA and cDNA (i.e., both presence and expression of the *nifH* genes).

Materials and Methods:

Samples analysed for this study were collected from the three major OMZ regions of the world oceans (Table1) from surface, oxycline and oxygen depleted zone (ODZ) depths. Particulate material from water samples (5 – 10 L), collected using Niskin samplers, mounted on a CTD (Conductivity-Temperature-Depth) rosette system (Sea-Bird Electronics), was filtered onto Sterivex

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

capsules (0.2 µm filter, Millipore, Inc., Bedford, MA) immediately after collection using peristaltic pumps. The filters were flash frozen in liquid nitrogen and stored at -80°C until DNA and RNA could be extracted. For samples from the Arabian Sea, DNA extraction was carried out using the PUREGENE™ Genomic DNA Isolation Kit (Qiagen, Germantown, MD) and the RNA was extracted using the ALLPrep DNA/RNA Mini Kit (Qiagen, Germantown, MD). For samples collected from ETNP and ETSP DNA and RNA were simultaneously extracted using the ALLPrep DNA/RNA Mini Kit (Qiagen, Germantown, MD). SuperScript III First Strand Synthesis System (Invitrogen, Carlsbad, CA, USA) was used to synthesise cDNA immediately after extraction following purification of RNA using the procedure described by the manufacturer, including RT controls. DNA was quantified using PicoGreen fluorescence (Molecular Probes, Eugene, OR) calibrated with several dilutions of phage lambda standards.

PCR amplification of *nifH* genes from environmental sample DNA and cDNA was done on an MJ100 Thermal Cycler (MJ Research) using Promega PCR kit following the nested reaction (Zehr et al. 1998), with slight modification as in Jayakumar et al. (2017). Briefly, 25µl PCR reactions containing 50 pmoles each of outer primer and 20-25ng of template DNA, were amplified for 30 cycles (1 min at 98°C, 1 min at 57°C, 1 min at 72°C), followed by amplification with the inner PCR primers 50 pmoles each (Zehr and McReynolds 1989). Water for negative controls and PCR was freshly autoclaved and UV-irradiated every day. Negative controls were run with every PCR experiment, to minimize the possibility of amplifying contaminants (Zehr et al. 2003). The PCR preparation station was also UV irradiated for 1 hour before use each day and the number of amplification cycles was limited to 30 for each reaction. Each reagent was tested separately for amplification in negative controls. *nifH* bands were excised from PCR products after electrophoresis on 1.2% agarose gel, and were cleaned using a QIAquick Nucleotide Removal Kit (Qiagen). Clean

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

72 *nifH* products were inserted into a pCR®2.1-TOPO® vector using One Shot® TOP10 Chemically
73 Competent *E. coli*, TOPO TA Cloning® Kit (Invitrogen) according to manufacturer's specifications.
74 Inserted fragments were amplified with M13 Forward (-20) and M13 Reverse primers from
75 randomly picked clones. PCR products were sequenced at Macrogen DNA Analysis Facility using
76 Big DyeTM terminator chemistry (Applied Biosystems, Carlsbad, CA, USA). Sequences were
77 edited using FinchTV ver. 1.4.0 (Geospiza Inc.), and checked for identity using BLAST. Consensus
78 *nifH* sequences (359 bp) were translated to amino acid (aa) sequences (108 aa after trimming the
79 primer region) and aligned using ClustalX (Thompson et al. 1997) along with published *nifH*
80 sequences from the NCBI database. Neighbor-joining trees were produced from the alignment using
81 distance matrix methods (PAUP 4.0, Sinauer Associates). Bootstrap analysis was used to estimate
82 the reliability of phylogenetic reconstruction (1000 iterations). The *nifH* sequence from
83 *Methanosarcina lacustris* (AAL02156) was used as an outgroup. The accession numbers from
84 GenBank for the *nifH* sequences in this study are Arabian Sea DNA sequences JF429940‑ JF429973
85 and cDNA sequences accession numbers JQ358610‑JQ358707, ETNP DNA sequences KY967751-
86 KY967929 and cDNA sequence KY967930-KY968089, and ETSP DNA sequences MK408165‑
87 MK408307 and cDNA sequences MK408308‑MK408422.
88
89 Standardization and verification of specificity for Q-PCR assays was performed as described
90 previously (Jayakumar et al. 2009). Primers nifHfw and nifHrv (Mehta et al. 2003, Dang et al. 2013)
91 forward 5-GGHAARGGHGGHATHGGNAARTC-3 and reverse 5-
92 GGCATNGCRAANCCVCCRCANAC-3, which correspond to the amino acid positions 10 to 17
93 (GKGGIGKS) and 132 to 139 (VCGGFAMP) of *Klebsiella pneumoniae* numbering (Mehta et al.
94 2003), were used (100 pmoles per 25 mL reaction) to amplify a ~400 bp region of the *nifH* gene for

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

nifH quantification. Assays were carried out with Qiagen master mix (Qiagen Sciences, Maryland, USA) at an annealing temperature of 56 °C. Amplification conditions were chosen based on amplification efficiency and reproducible results with a single product, after test assays on a Stratagene MX3000P (Agilent Technologies, La Jolla, CA, USA). The amplified products were visualized after electrophoresis in a 1.0% agarose gels stained with ethidium bromide. Standards for quantification were prepared by amplifying a constructed plasmid containing the *nifH* gene fragment, followed by quantification and serial dilution. Assays for all depths were carried out within a single assay plate (Smith et al. 2006). Each assay included triplicates of the no template controls (NTC), no primer control (NPR), four or more standards, and 20-25ng of template DNA of the environmental DNA samples. A subset of samples from the previous run was included in subsequent assays, as well as a new dilution series for standard curves on every assay. These new dilution series were produced immediately following re-quantification of plasmid DNA concentrations to verify gene abundance (because concentrations declined upon storage and freeze–thaw cycles). Automatic analysis settings were used to determine the threshold cycle (Ct) values. The copy numbers were calculated according to: *Copy number = (ng * number/mole) / (bp * ng/g * g/ mole of bp)* and then converted to copy number per ml seawater filtered, assuming 100% extraction efficiency.

The *nifH* nucleotide alignment (of 787 sequences) was used to define operational taxonomic units (OTUs) on the basis of DNA sequence identity. Distance matrices based on this nucleotide alignment were generated in MOTHUR (Schloss and Handlesman 2009). The relative *nifH* richness within each clone library was evaluated using rarefaction analysis. OTUs were defined as sequences which differed by ≤3% using the furthest neighbor method in the

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

MOTHUR program (Schloss and Handlesman 2009). The 3% OTU definition is similar to the level at which species are conventionally defined using 16S rDNA sequences, so it may overestimate the meaningful diversity of the functional gene.

Results and Discussion:

DNA and cDNA sequences (787 in total) derived from the OMZ regions of the Arabian Sea (AS), Eastern Tropical North Pacific (ETNP) and Eastern Tropical South Pacific (ETSP) were subjected to OTU and phylogenetic analyses to compare the diversity and community composition, biogeography and gene expression, of *nifH* possessing microbes among the three OMZ regions. Phylogenetic analysis of the sequences from the AS, ETNP and ETSP were reported previously (Jayakumar et al. 2012, Jayakumar et al. 2017, Chang et al. 2019), but the sequences have been combined for additional analyses here. We compared the threshold OTU definitions at 3 and 10% and found that the number of OTUs decreased, as expected, as the resolution decreased. Even at the 3% threshold, however, OTUs tended to separate by depth and location, indicating a functionally useful distinction at this level. Thresholds of 3 – 5% as the OTU definition correspond to within and between species level distinctions for *nifH* (Gaby et al. 2018). The sequences from the OMZ regions represented all four sequence clusters (I, II, III, IV) described by Zehr et al. (1998).

Cluster I *nifH* OTU distributions: Diversity analysis of the *nifH* cluster 1 sequences for the three OMZs based on OTUs using MOTHUR identified 41 OTUs at a distance threshold of 3% (Supplemental Table 1A and B). The number of sequences and the number of OTUs varied widely among depths and stations, so the results are grouped by region (AS, ETNP,

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

ETSP) or depth horizon (surface or OMZ, including upper oxycline depths) or cDNA vs DNA (Table 1).

For all regions and depths combined, the number of OTUs detected (41) was less than the sum of OTUs detected when each region was analyzed separately (45), indicating that there was some overlap of OTUs among regions. The overlap was not large, however. Only three of the 12 most abundant OTUs contained sequences from more than one region and none contained sequences from all three regions (Figure 1A). When sequences for all three regions were combined, only four of the 12 most abundant OTUs contained sequences from both depth horizons (Figure 1B). Most OTUs represented a single depth, and many a single sample.

The Arabian Sea was strikingly less diverse than other regions and sample subsets (Figure 2). For example, when all DNA and cDNA sequences for all depths are grouped together, the Arabian Sea (OTUs = 14, Chao = 21) contains less species richness than the combined surface samples from all three regions (OTUs = 25, Chao = 52), despite having a similar number of total sequences (178 for the Arabian Sea, 198 for all surface samples combined). This lack of diversity in the AS data may be partly due to the preponderance of cDNA sequences, which generally contained less diversity than a similar number of DNA sequences (see below).

Although similar numbers of sequences were obtained for cDNA (255) vs DNA (257), the OTU "density", i.e., number of OTUs per number of sequences analyzed, was higher for DNA (0.136 for DNA, 0.094 for cDNA). The Chao statistic verified this observation for the combined data from each region in predicting higher total numbers of OTUs for DNA (Chao = 42) than for cDNA (Chao = 24). This difference could indicate that some of the *nifH* genes present were not expressed at the time of sampling, but the cDNA sequences were not simply a

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

subset of the DNA community. Half of the 12 most abundant OTUs contained either cDNA or DNA (Figure 1C), meaning that some genes were never expressed and some expressed genes could not be detected in the DNA.

For all regions combined, similar numbers of OTUs were detected in surface waters (OTUs = 25) and in OMZ samples (OTUs = 23), although a larger number of sequences was analyzed for the OMZ environment (198 vs. 314 sequences for surface and OMZ depths, respectively). It might be expected that the presence of phototrophic diazotrophs in the surface water would lead to greater diversity there, but only one OTU representing a known cyanobacterial phototroph (OTU-12 = *Katagmynene spiralis* or *Trichodesmium*) was identified, so most of the additional diversity must be present in heterotrophic or unknown sequences.

Rarefaction curves (Figure 2) indicate that sampling did not approach saturation either for region or depth. The Chao statistic also indicated that much diversity remains to be explored, despite the great uncertainty in these estimates. The total number of OTUs detected, the shape of the rarefaction curve and the diversity indicators (Figure 2, Table 1) all indicate that the greatest *nifH* diversity occurred in surface waters, and much of that diversity was in singletons, i.e., not represented in the 12 most abundant OTUs, which represented 441 (86 %) of the total 512 *nifH* Cluster 1 sequences analyzed. Most of that diversity was contained in the ETNP, not solely a function of number of sequences analyzed (Figure 2).

Cluster I *nifH* Phylogeny: Phylogenetic affiliations at both DNA and protein level are shown for the 12 most abundant OTUs in Table 2. The most abundant OTU (129 sequences), OTU-1, contained Gammaproteobacterial DNA and cDNA sequences from both surface and OMZ depths of the ETNP and cDNA sequences from oxycline and OMZ depths in the Arabian Sea (Figure 3). Although very similar to each other, none of these sequences had higher than

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

91% identity at the DNA level (96% at AA level) with cultivated strains and were most closely related to *Pseudomonas stutzeri*. *P. stutzeri* is a commonly isolated marine denitrifier, but it is also known to possess the capacity for N fixation (Krotzky and Werner 1987). OTU-4, OTU-6 and OTU-8 also contained Gammaproteobacterial sequences. All had high identity with cultivated strains at the protein level but none were >91% identical to cultivated strains at the DNA level.

Gammaproteobacterial sequences with very close identities to *Azotobacter vinelandii* have been reported from the Arabian Sea ODZ and also from the ETSP (Turk-Kubo et al. 2014). This group of *nifH* sequences with close identities to *A. vinelandii* was also retrieved from the English Channel, Himalayan soil, South Pacific gyre, Gulf of Mexico, mangrove soil and many other environments (Figure 3). *Azotobacter*- like sequences were included in OTU-6 but were not closest identity at the DNA level. Although a large number of clones were analyzed here, no sequence that was closely associated with *A. vinelandii* was retrieved from the three regions. None of the g-244774A11 sequences, Gammaproteobacterial relatives that were abundant in the South Pacific (Moisander et al. 2014), were detected in this study.

OTUs-2, 3, 5, 10, and 11 all represented Alphaproteobacterial sequences, with closest identities to various *Bradyrhizobium, Sphingomonas* and *Methylosinus* species. Thus, Alphaproteobacterial sequences (206 sequences) were the most abundant in the clone library. OTU-2 contained almost exclusively ETSP ODZ DNA and cDNA sequences (plus one AS ODZ DNA sequence). OTU-3 contained DNA sequences from ETNP surface waters. OTU-5 contained exclusively Arabian Sea DNA sequences from Station 3, while OTU-10 contained only surface samples from the ETNP. An OTU threshold of 11% grouped all (179 sequences in five OTUs) of

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

these Alphaproteobacterial sequences together, but the 3% threshold is consistent with the phylogenetic tree, which shows small scale biogeographical separation of sequence groups.

OTUs-7 and -9 were identified as Betaproteobacteria with closest identities to *Rubrivivax gelatinosum* and *Burkholderia*, 91 and 90% respectively at the DNA level. However, at the AA level, these sequences were 99 and 100% identical to *Novosphingobium malaysiense* and *S. azotifigens,* both Alphaproteobacteria, and again were biogeographically distinct. OTU-7 contained 25 DNA sequences from the ODZ depths in the Arabian Sea, and OTU-9 contained 17 *Burkholderia*-like sequences from the oxycline at Station 1 in the Arabian Sea. No Betaproteobacterial *nifH* sequences were detected in the ETNP or ETSP, but sequences similar to *Burkholderia phymatum*, *Cupriavidus sp.* and *Sinorhizobium meliloti* were reported from ETSP previously (Fernandez et al. 2015). Consistent with our previous report, however, there is no clear separation between the alfa and the beta groups in *nifH* phylogeny (Jayakumar et al 2017).

Most of the Cluster I ETSP sequences from this study were contained in two OTUs (2 and 4). OTU-2 contained 89 Alphaproteobacterial sequences with >98% identity to *nifH* sequences from *Bradyrhizobium* sp. Uncultured bacterial sequences retrieved from the South China Sea, English Channel, mangrove sediment, wastewater treatment and grassland soil were related to these ETSP sequences. OTU-4 contained 29 Gammaproteobacterial sequences retrieved from both surface and ODZ depths. Four of the remaining ETSP Cluster I sequences were grouped together as OTU-17 (Alphaproteobacteria, 89 and 96% identities with *Methyloceanibacter sp.* and *Bradyrhizobium sp.* at the DNA and AA level respectively), three were in OTU-23 (*Bradyrhizobium* 100% identity) and two were singletons. One of the singletons was most closely related to uncultured soil and sediment sequences and to *Azorhizobium sp*. (86%) and one had 97% identity with *Bradyrhizobium denitrificans* and many sequences from marine sediments.

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

232 OTU-22 represents the Deltaproteobacterial group. This novel group was reported
233 previously from the ETNP (Jayakumar et al. 2017) and has three sequences from Arabian sea (OTU-
234 22) and two singletons from ETNP surface waters. *nifH* possessing Deltaproteobacteria have been
235 reported not only from all the three ODZs but also in several other marine environments including
236 Chesapeake Bay water column, microbial mats from intertidal sandy beach in a Dutch barrier island,
237 Jiaozhou Bay sediment, Rongcheng Bay sediment, Bohai Sea, Mediterranean Sea, Narragansett Bay,
238 and the south Pacific gyre.
239 Although *Trichodesmium* like clones have been retrieved from the surface waters of the
240 Arabian Sea and the ETNP OMZs, only ten clones (OTU-12) in the combined clone library analyzed
241 here were related to *Trichodesmium* (98% identity), including both cDNA and DNA from the
242 Arabian Sea and cDNA from the ETNP. These sequences were actually 100% identical to
243 *Katagnymene spiralis*, a close relative of *Trichodesmium* isolated from the South Pacific Ocean.
244 Turk-Kubo et al. (2014) also retrieved only a few cyanobacterial sequences from the ETSP. No other
245 cyanobacterial *nifH* sequences were identified.
246 **Clusters II, III, IV *nifH* OTU distributions:** The other three *nifH* clusters were combined
247 for OTU analysis due to the limited number of sequences and OTUs obtained. A total of 18 OTUs
248 were identified in the combined set of 275 sequences with a 3% distance threshold (Table 2). Most
249 of the Cluster II, III, IV sequences were from the ETNP and ETSP. As with the Cluster I sequences,
250 there was very little geographic and depth overlap among these OTUs (Figure 4A, 4B). Only OTU-
251 1 contained sequences from more than one site, the ETNP and the ETSP. OTU-2 contained only
252 cDNA sequences representing ODZ depths at both ETNP stations. OTU-3 contained exclusively
253 ETSP DNA sequences from surface and cDNA sequences from ODZ depths. Only 10 of the Cluster
254 II, III, IV sequences were from the Arabian Sea, and they formed three separate OTUs, a greater

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

"OTU density" than was present at either of the Pacific sites. As observed for Cluster I, most of the OTUs that were detected in the DNA were not being expressed, and those that were expressed were not detected in the DNA (Figure 4C).

Rarefaction curves (Figure 5) indicate that sampling for Cluster II, III, IV did not approach saturation. The Chao statistic also indicated that much diversity remains to be explored, despite the great uncertainty in these estimates. Unlike the Cluster I analysis, there were relatively few singletons in the Cluster II, III, IV data and the assemblages were dominated by a few types.

Clusters II, III, IV *nifH* phylogeny: Three large OTUs (OTU-1, -4 and -6) in Clusters II, III, IV belonged to *nifH* Cluster IV and Alphaproteobacteria/Spirochaeta and Deltaproteobacteria were the dominant phylogenies (Table 2, Figure 6). The largest OTU, OTU-1, contained 88 DNA sequences from the ETNP ODZ depths from both stations and from both depths in the ETSP. This OTU had no similarity to any cultured microbe. OTU-4 contained 30 sequences from the ETSP, all cDNA from one surface station, in *nifH* Cluster IV.

OTU-2 (75 sequences) in Cluster II contained only cDNA sequences, all from ODZ samples in the ETNP (both stations), and had no close relatives among cultivated species. Turk-Kubo et al. (2014) also retrieved a few clones identified as belonging to Cluster II from the euphotic zone of the ETSP. OTU-3 contained 35 sequences in Cluster III and was dominated by DNA sequences from surface depths of the ETSP. OTU-5 represented Deltaproteobacteria in *nifH* Cluster III and contained 18 identical DNA sequences from 90 m at Station BB1 in the ETNP. Thus, of the five most common OTUs (89% of the total Cluster II, III, IV sequences analyzed), only one could be identified to a closely related genus (i.e., OTU-4 with 90% identity

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

277 with *R. palustris*) and there was no overlap between DNA and cDNA OTUs from the same
278 depths.

279 The other 13 OTUs in the Cluster II, III, IV sequences represented either Cluster III or IV.
280 None of these were very closely related to any cultivated sequences. OTU-6 contained both DNA
281 and cDNA from the OMZ at one ETSP station. OTU-7 contained four sequences from ETNP
282 surface waters with close identities with a sequence retrieved from Bohai sea. OTU-11, had one
283 DNA and one cDNA sequences from the ETSP. All of the other sequences were less than 84%
284 identical to any sequence in the database and could only be loosely identified as Firmicutes or
285 Proteobacteria.

286

287 **Conclusions**

288 The OMZ regions of the world ocean contain substantial *nifH* diversity, both in surface
289 waters and oxygen depleted intermediate depths. Surface waters contained greater diversity for
290 Cluster I, but the ODZ held the highest diversity for Clusters II, III, IV. Cyanobacterial sequences
291 were rare and were not detected in the ETSP. The ETSP contained the least diversity of Cluster I
292 sequences, while Cluster II, III, IV were least abundant and least diverse in the Arabian Sea. Most
293 of the sequences in all four Clusters of the conventional *nifH* phylogeny were not closely related to
294 any sequences from cultivated Bacteria or Archaea. The most abundant OTUs in Cluster I and in
295 Clusters II, III, and IV could be assigned to the Alphaproteobacteria, followed by the
296 Gammaproteobacteria for Cluster I and Deltaproteobacteria accounted for Clusters II, III, IV
297 sequences. Most of the OTUs were not shared among regions, depths or DNA vs cDNA and
298 sometimes were restricted to individual samples. Some Cluster I sequences had high identity to
299 known species (e.g., *Bradyrhizobium*, *Trichodesmium*) but most of the Cluster II, III, IV sequences

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

were only distantly related to any cultured species. While measurements of N_2 fixation rates are not reported here, the abundance of cDNA sequences suggests that the cells harboring these genes are active. Low, but analytically significant, rates have been detected in ODZ depths in the ETNP (Jayakumar et al. 2017) and ETSP (Chang et al. 2019), which suggests that non-cyanobacterial N fixation could make a minor contribution to the nitrogen budget of the ocean. It is therefore important in future work to determine how the diversity described here actually contributes to biogeochemically significant reactions and what environmental and biotic factors might influence or control the activity of diazotrophs in the dark ocean.

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

Figure Legends

Figure 1. Histogram of the 12 most common OTUs from Cluster I *nifH* clone libraries from the three OMZ regions. OTUs were considered common if the total number of sequences in an OTU was ≥2% of the total number of *nifH* clones analyzed (The common OTUs contained 441 of the 512 Cluster I sequences). OTUs were defined according to 3% nucleotide sequence difference using the furthest neighbor method. OTU designation is from most common (OTU-1) to least. A) OTU distribution among regions. B) OTU distribution between OMZ (including core of the ODZ and the upper oxycline depths) and surface depths (oxygenated water). C) OTU distribution of cDNA vs DNA clones.

Figure 2. Rarefaction curve displaying observed OTU richness versus the number of clones sequenced for Cluster I *nifH* sequences (cDNA and DNA). OTUs were defined and designated as in Figure 1. Chao estimators (individual symbols) are shown for each of the same subsets represented in the rarefaction curves.

Figure 3. Phylogenetic tree of Cluster 1 based on amino acid sequences. Positions of the OTUs are shown relative to their nearest neighbors from the database. Individual sequence identities comprising each OTU are listed in Supplemental Table 2.

Figure 4. Histogram of the 6 most common OTUs from Cluster I *nifH* clone libraries from the three OMZ regions. OTUs were considered common if the total number of sequences in an OTU was ≥2% of the total number of *nifH* clones analyzed (the common OTUs contained 252 of

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

the 275 Cluster II, III, IV sequences). OTUs were defined according to 3% nucleotide sequence difference using the furthest neighbor method. OTU designation is from most common (OTU-1) to least. A) OTU distribution among regions. B) OTU distribution between OMZ (including core of the ODZ and the upper oxycline depths) and surface depths (oxygenated water). C) OTU distribution of cDNA vs DNA clones.

Figure 5. Rarefaction curve displaying observed OTU richness versus the number of clones sequenced for Cluster II, III, IV *nifH* sequences (cDNA and DNA). OTUs were defined and designated as in Figure 4. Chao estimators (individual symbols) are shown for each of the same subsets represented in the rarefaction curves.

Figure 6. Phylogenetic tree of Clusters II, III, IV based on amino acid sequences. Positions of the OTUs are shown relative to their nearest neighbors from the database. Individual sequence identities comprising each OTU are listed in Supplemental Table 2.

Tables

Table 1. OTU summary for both clusters

Richness and diversity statistics for *nifH* clone libraries from three OMZ regions. ACE and Chao are non-parametric estimators that predict the total number of OTUs in the original sample.

Table 2. OTU identities for both clusters

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

355 Cultivated species with closest nucleotide identity to the OTUs identified in the *nifH* clone
356 libraries from three OMZ regions. Only the 12 most common OTUs (out of 41 total) are listed
357 for Cluster 1 sequences, and the six most common (out of 18 total) for the Clusters II, III, IV
358 libraries.
359
360 Supplemental
361
362 S Table 1A and B. List of sequences in each OTU for both clusters
363 S Table 2
364
365
366
367

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

References

1. Chang, B. X., Jayakumar, A., Widner, B., Bernhardt, P., Mordy, C. M., Mulholland, M. R. and Ward, B. B., 2019. Low rates of dinitrogen fixation in the eastern tropical South Pacific. Limnology And Oceanography. 64, 1913-1923.
2. Dang, H. Y., Yang, J. Y., Li, J., Luan, X. W., Zhang, Y. B., Gu, G. Z., Xue, R. R., Zong, M. Y. and Klotz, M. G., 2013. Environment-Dependent Distribution of the Sediment nifH-Harboring Microbiota in the Northern South China Sea. Applied and Environmental Microbiology. 79, 121-132.
3. Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N. and Dunne, J. P., 2007. Spatial coupling of nitrogen inputs and losses in the ocean. Nature. 445, 163-167.
4. Fernandez, C., Lorena Gonzalez, M., Munoz, C., Molina, V. and Farias, L., 2015. Temporal and spatial variability of biological nitrogen fixation off the upwelling system of central Chile (35-38.5 degrees S). Journal of Geophysical Research-Oceans. 120, 3330-3349.
5. Gaby, J. C., Rishishwar, L., Valderrama-Aguirre, L. C., Green, S. J., Valderrama-Aguirre, A., Jordan, I. L. and Kostka, J. E., 2018. Diazotroph community characterization via a high-throughput nifH amplicon sequencing and analysis pipeline. Applied And Environmental Microbiology. 84, eO1512-01517.
6. Großkopf, T. and LaRoche, J., 2012. Direct and indirect costs of dinitrogen fixation in Crocosphaera watsonii WH8501 and possible implications for the nitrogen cycle. Frontiers in Microbiology. 3.
7. Jayakumar, A., Al-Rshaidat, M. M. D., Ward, B. B. and Mulholland, M. R., 2012. Diversity, distribution, and expression of diazotroph nifH genes in oxygen-deficient waters of the Arabian Sea. Fems Microbiology Ecology. 82, 597-606.
8. Jayakumar, A., Chang, B. N. X., Widner, B., Bernhardt, P., Mulholland, M. R. and Ward, B. B., 2017. Biological nitrogen fixation in the oxygen-minimum region of the eastern tropical North Pacific ocean. Isme Journal. 11, 2356-2367.
9. Jayakumar, A., Naqvi, S. W. A. and Ward, B. B., 2009. Distribution and relative quantification of key genes involved in fixed nitrogen loss from the Arabian Sea oxygen minimum zone. Indian Ocean Biogeochemical Processes and Ecological Variability. (J. D. Wiggert and R. R. Hood). American Geophysical Union, Washington, D. C.: 187-203.
10. Krotzky, A. and Werner, D., 1987. NITROGEN-FIXATION IN PSEUDOMONAS-STUTZERI. Archives of Microbiology. 147, 48-57.
11. Mehta, M. P., Butterfield, D. A. and Baross, J. A., 2003. Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca ridge. Applied and Environmental Microbiology. 69, 960-970.
12. Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P. and Zehr, J. P., 2010. Unicellular Cyanobacterial Distributions Broaden the Oceanic N-2 Fixation Domain. Science. 327, 1512-1514.
13. Moisander, P. H., Benavides, M., Bonnet, S., Berman-Frank, I., White, A. E. and Riemann, L., 2017. Chasing after Non-cyanobacterial Nitrogen Fixation in Marine Pelagic Environments. Frontiers in Microbiology. 8.

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

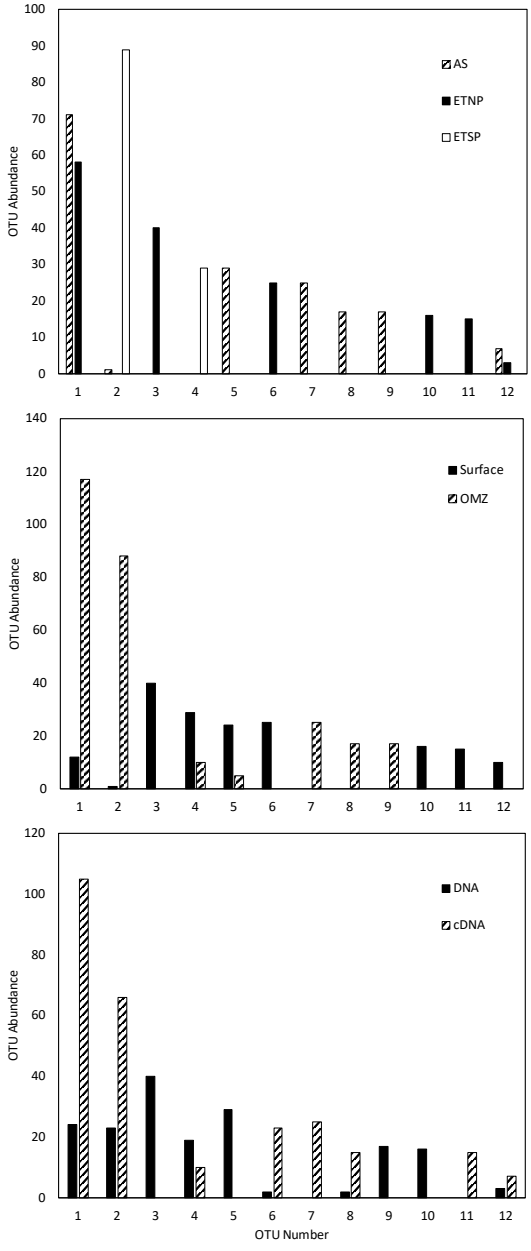
410 14. Moisander, P. H., Serros, T., Paerl, R. W., Beinart, R. A. and Zehr, J. P., 2014.
411 Gammaproteobacterial diazotrophs and nifH gene expression in surface waters of the
412 South Pacific Ocean. The ISME Journal 8, 1962–1973.
413 15. Schloss, P. D. and Handlesman, J., 2009. Introducing DOTUR, a computer program for
414 defining operational taxonomic units and estimating species richness. Applied and
415 Environmental Microbiology. 71, 1501-1506.
416 16. Smith, C. J., Nedwell, D. B., Dong, L. F. and Osborn, A. M., 2006. Evaluation of quantitative
417 polymerase chain reaction-based approaches for determining gene copy and gene
418 transcript numbers in environmental samples. Environmental Microbiology. 8, 804-815.
419 17. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G., 1997. The
420 CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided
421 by quality analysis tools. Nucleic Acids Research. 25, 4876-4882.
422 18. Turk-Kubo, K. A., Karamchandani, M., Capone, D. G. and Zehr, J. P., 2014. The paradox of
423 marine heterotrophic nitrogen fixation: abundances of heterotrophic diazotrophs do not
424 account for nitrogen fixation rates in the Eastern Tropical South Pacific. Environmental
425 Microbiology. 16, 3095-3114.
426 19. Zehr, J. P., Crumbliss, L. L., Church, M. J., Omoregie, E. O. and Jenkins, B. D., 2003.
427 Nitrogenase genes in PCR and RT-PCR reagents: implications for studies of diversity of
428 functional genes. Biotechniques. 35, 996-1005.
429 20. Zehr, J. P. and McReynolds, L. A., 1989. Use of degenerate oligonucleotides for amplification
430 of the *nifH* gene from the marine cyanobacterium *Trichodesmium theiebautii*. Applied
431 and Environmental Microbiology. 55, 2522-2526.
432 21. Zehr, J. P., Mellon, M. T. and Zani, S., 1998. New nitrogen-fixing microorganisms detected in
433 oligotrophic oceans by amplification of nitrogenase (nifH) genes. Applied and
434 Environmental Microbiology. 6, 3444-3450.
435

436

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

437

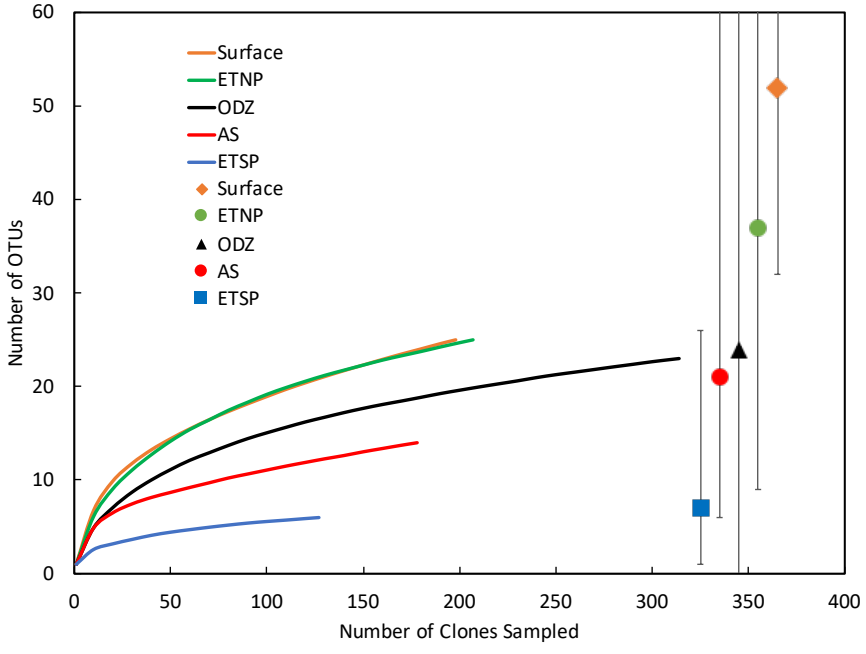
438 Figure.1



439

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

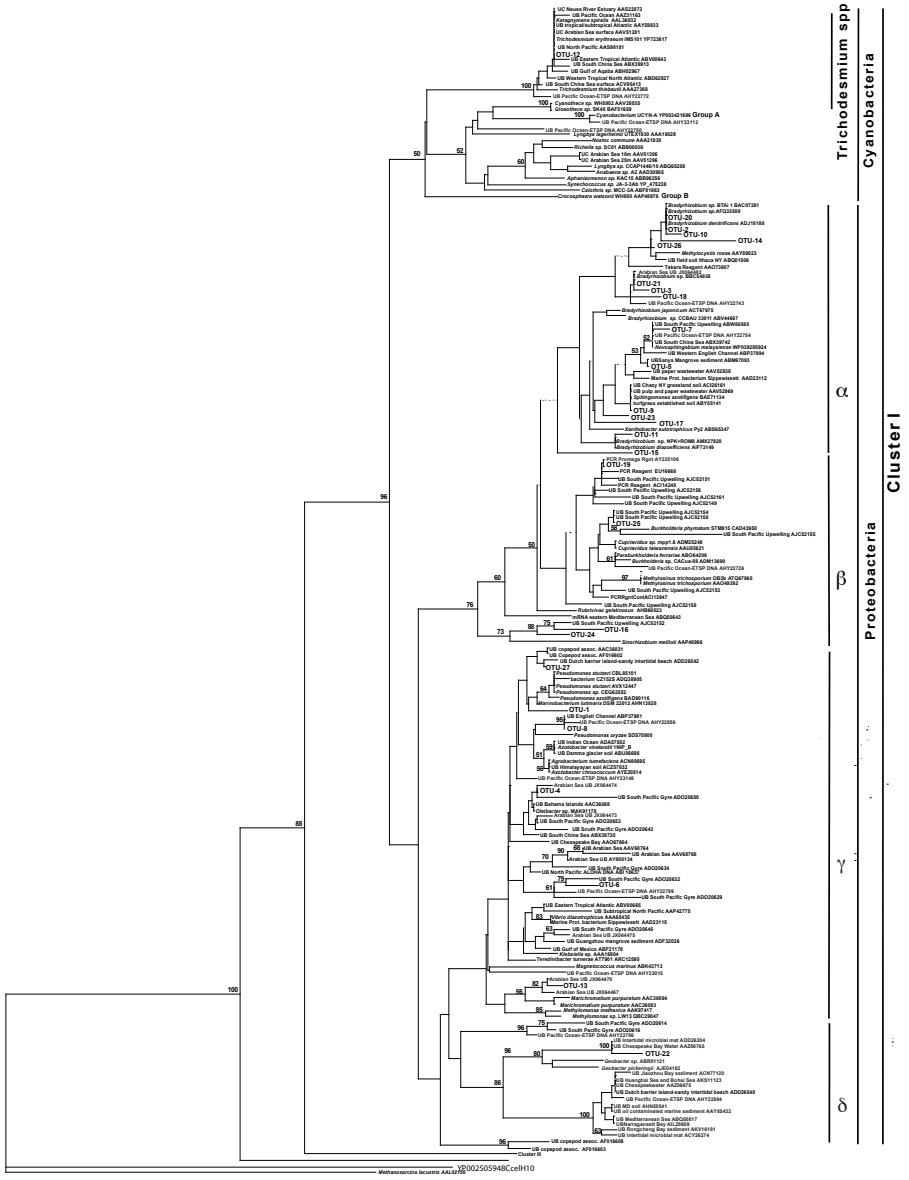
440 Figure. 2



441
442

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

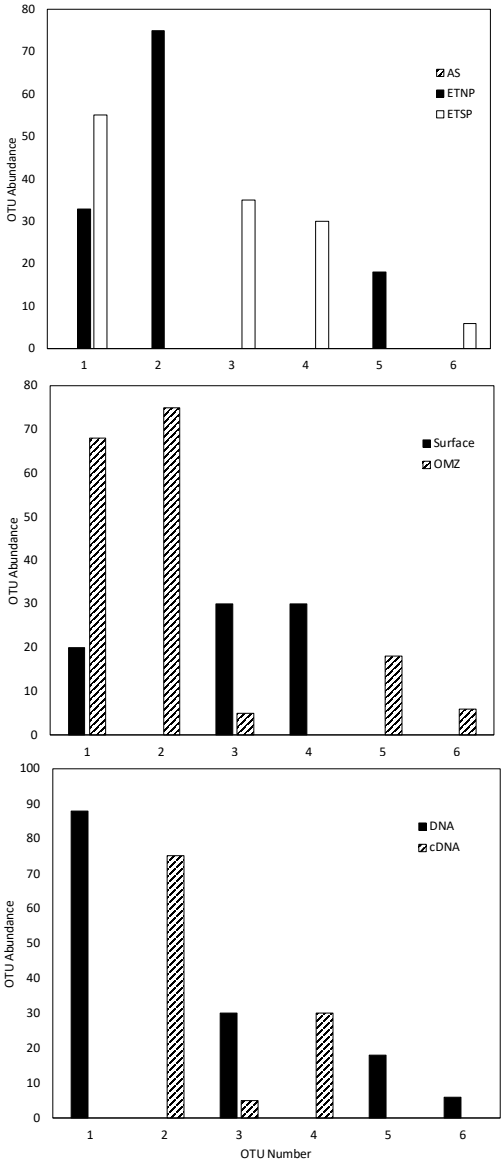
443 Figure. 3



444
445

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

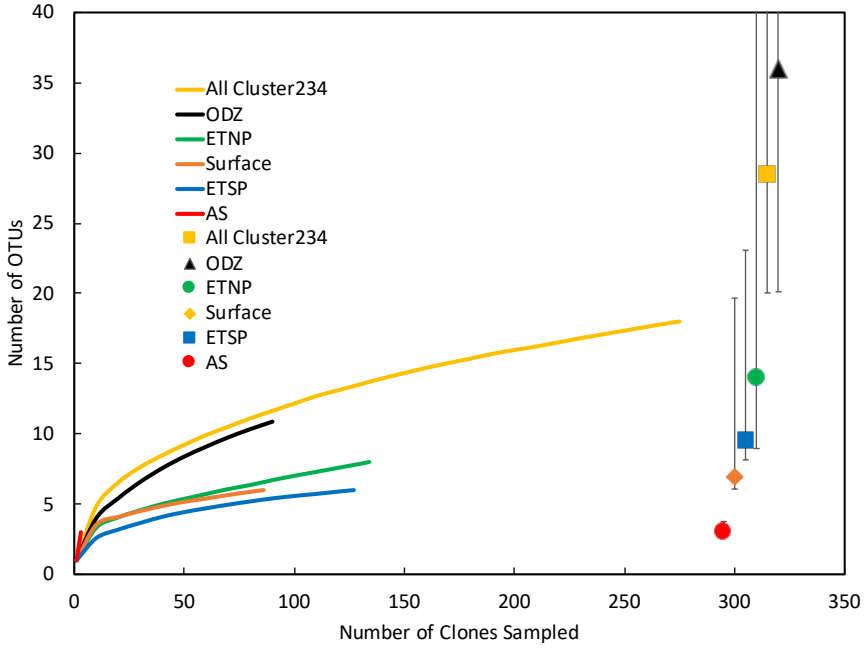
446 Figure. 4



447
448

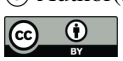
https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

449 Figure. 5.



450
451

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020



Biogeosciences
Discussions
Open Access
EGU

452 Figure 6.

453
454

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

455 Table 1 OTU Summary

Sample subset	Depths, regions included	No. of Sequences	No. of Unique Sequences	No. of OTUs (cutoff ~3)	OTU /seq	Shannon	Simpson	Chao	Ace
Cluster I									
AS	Arabian Sea, all depths	178	36	14	0.079	1.8	0.22	21	45
ETNP	ETNP, all depths	207	80	25	0.121	2.37	0.14	37	34
ETSP	ETSP, OMZ depths	127	51	6	0.047	0.87	0.53	7	8
All ClusterI	Three regions, all depths	512	165	41	0.080	2.7	0.11	59	67
All ClusterI DNA	Three regions, all depths	257	97	35	0.136	2.8	0.08	42	45
All ClusterI cDNA	Three regions, all depths	255	75	24	0.094	1.7	0.25	24	27
All ClusterI Surface	Three regions, surface depths	198	73	25	0.126	2.5	0.10	52	75
All ClusterI OMZ	Three regions, all depths	314	98	23	0.073	0.9	0.23	30	37
Clusters II, III, IV									
AS	Arabian Sea, all depths	10	6	3	0.300	1.09	0.27	3	3
ETNP	ETNP, all depths	134	49	8	0.060	1.19	0.39	14	38
ETSP	ETSP, all depths	131	64	8	0.061	1.37	0.30	9	19
All Clusters II,III,IV	Three regions, all depths	275	117	18	0.065	1.88	0.21	28	26
All Clusters II,III,IV DNA	Three regions, all depths	155	65	12	0.077	1.20	0.37	22	17
All Clusters II,III,IV cDNA	Three regions, all depths	120	56	9	0.075	1.11	0.45	12	15
All Clusters II,III,IV Surface	Three regions, surface depths	86	46	6	0.070	1.32	0.29	7	13
All Clusters II,III IV OMZ	Three regions, OMZ depths	189	76	15	0.079	1.57	0.29	46	24

456
457
458
459

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

Biogeosciences Discussions Open Access EGU

460 Table 2
461

	No. of Sequences	Phylogenetic Affiliation	Closest cultured relative (DNA)	Identity DNA %	Coverage %	Closest cultured relative (Protein)	Identity AA %	Coverage %
Cluster I								
OTU-1	129	Gamma	*Psuedomonas stutzeri*	91	98	*Pseudomonas stutzeri strain SGAir0442*	95.8	99
OTU-2	89	Alpha	*Bradyrhizobium sp*	99	100	*Bradyrhizobium denitrificans strain LMG 8443*	99	99
OTU-3	40	Alpha	*Bradyrhizobium sp. TM124*	94	98	*Bradyrhizobium sp. MAFF 210318*	99	98
OTU-4	29	Gamma	*Marinobacterium lutimaris*	87	100	*Oleibacter sp*	100	99
OTU-5	29	Alpha	*Methylosinus trichosporium*	92	99	*Sphingomonas azotifigens*	99	100
OTU-6	25	Gamma	*Azotobacter chroococcum* strain B3	81	99	*Psuedomonas stutzeri*	94	99
OTU-7	25	Beta/Alpha	*Rubrivivax gelatinosus*	91	99	*Novosphingobium malasiense*	99	100
OTU-8	17	Gamma	*Psuedomonas stutzeri*	91	98	*Azotobacter chroococcum* strain B3	97	100
OTU-9	17	Beta/Alfa	*Burkholderia*	90	100	*Sphingomonas azotifigens*	100	100
OTU-10	16	Alpha	*Bradyrhizobium*	97	98	*Bradyrhizobium sp. ORS 285*	99	99
OTU-11	15	Alpha	*Bradyrhizobium*	97	98	*Bradyrhizobium diazoefficiens*	98	99
OTU-12	10	Cyanobacterium	*Katagnymene spiralis*	100	99	*Trichodesmium erythraeum*	100	99
Clusters II, III IV								
OTU-1	88	Alpha/Spirochaetaceae	*Rhizobium sp*	74	59	*Treponema primitia ZAS-1]*	55	98
OTU-2	75	Delta/Firmicutes	*Geobacter*	73	43	*Desulfitobacterium hafniense*	98	61
OTU-3	35	Verrumicrobia	*Opitutaceae bacterium*	82	99	*Coraliomargarita akajimensis*	95	99
OTU-4	30	Alpha	*Rhodopseudomonas palustris*	90	98	*Rhodoplanes elegans*	96	99
OTU-5	18	Delta/Chlorobi	*Desulfovibrio piezophilus*	79	99	*Prosthecochloris sp. V1, Chloroherpeton thalassium, Chloroherpeton thalassium*	92	99
OTU-6	6	Beta/Delta	*Azoarcus communis*	70	88	*Enhygromyxa salina*	70	74
OTU-7	4	Delta	*Desulfovibrio carbinolicus strain DSM 3852*	81	99	*Desulfovibrio inopinatus*	90	99
OTU-8	4	Delta/Firmicutes	*Desulfovibrio desulfuricans strain IC1*	77	100	*Sporobacter termitidis*	99	99

https://doi.org/10.5194/bg-2019-445
Preprint. Discussion started: 13 January 2020

	OTU-9	3	Delta/Lentisphaerae	*Desulfovibrio magneticus RS-1 DNA*	84	100	*Lentisphaerae bacterium GWF2_57_35, Desulfatitalea tepidiphila, Desulfobacteraceae bacterium*	84	100
	OTU-10	3	Delta/Methanococci	*Desulfovibrio desulfuricans strain IC1*	77	100	*Methanocaldococcus villosus*	65	99
	OTU-11	2	Verrucomicrobia	Verrucomicrobia bacterium S94	87	100	Verrucomicrobia bacterium S94	97	99

462
463
464