Microbial diversity of the Arabian Sea in the Oxygen minimum zones by metagenomics approach

Short Title: Microbial diversity of the Arabian Sea

Mandar S. Paingankar[#], Kedar Ahire, Pawan Mishra, Shriram Rajpathak, Deepti D. Deobagkar*

Molecular Biology Research Laboratory, Centre for Advanced Studies, Department of Zoology, Savitribai Phule Pune University, Pune - 411 007, India.

[#]Current Address - Department of Zoology, Government Science College, Chamorshi Road, Gadchiroli -442605, Maharashtra, India.

*Correspondence

Prof. Deepti D. Deobagkar,
ISRO Chair Professor, ISRO Science and technology Cell &
Department of Zoology, Centre for Advance Studies,
Savitribai Phule Pune University, Pune 411007, India.
Tel.: +91992118487
deepti.deobagkar@gmail.com

23 **ABSTRACT**

24 Large oxygen depleted areas known as oxygen minimum zones (OMZ) have been
25 observed in the Arabian Sea and recent reports indicate that these areas are expanding at an
26 alarming rate. In marine waters, oxygen depletion may also be related to global warming
27 and the temperature rise, acidification and deoxygenation can lead to major consequences
28 wherein the plants, fish and other biota will struggle to survive in the ecosystem.

29 The current study has identified the microbial community structure using NGS
30 based metagenomics analysis in the water samples collected at different depth from the
31 oxygen depleted and non-OMZ areas of Arabian Sea. Environmental variables such as
32 depth, site of collection and oxygen concentration appeared to influence the species
33 richness and evenness among microbial communities in these locations. Our observations
34 clearly indicate that population dynamics of microbes consisting of nitrate reducers
35 accompanied by sulphate reducers and sulphur oxidizers participate in the interconnected
36 geochemical cycles of the OMZ areas. In addition to providing baseline data related to the
37 diversity and microbial community dynamics in oxygen-depleted water in the OMZ; such
38 analysis can provide insight into processes regulating productivity and ecological
39 community structure of the ocean.

40 **Keywords:** Arabian Sea, bacterial diversity, oxygen minimum zones, metagenomics,
41 sulphur and nitrogen metabolism, Goa, Mangalore, Calicut.

42

43 INTRODUCTION

44 The Oxygen minimum zone (OMZ) in the Arabian-Sea is the second-most intense OMZ
 45 amongst the tropical oceans in the world^{1,2} with a near-total depletion of oxygen at depths
 46 from 200 to 1000m³. In these locations, suboxic levels ($\leq 5 \mu\text{mol O}_2/\text{kg}$) of oxygen are
 47 seen over vast areas at different depths and denitrification occurs in its upper portion⁴.
 48 Geochemical observations indicate that oxygen minimum zones have expanded over the
 49 past decades⁵ and could expand further in response to the ocean warming and increased
 50 stratification associated with climate change^{6,7}. It has been suggested that the biological
 51 consumption of oxygen is most intense below the region of highest productivity in the
 52 western Arabian Sea⁸⁻¹⁰. The total volume of the OMZ in the ocean is growing at an
 53 alarming rate, their upper boundaries are vertically shoaling, and the degree of anoxia is
 54 intensifying within the cores of the OMZs^{5,11}. The expansion of the oxygen minimum
 55 zones (OMZs) in the Arabian Sea has become the major concern because of its impact on
 56 the marine ecosystems. The expansion of the OMZs due to climate change and its impacts
 57 on the ecosystems and the atmosphere is multi-dimensional and requires intense study.

58 OMZ is characterized by high nitrite accumulations and very low or undetectable
 59 oxygen concentrations¹². The nitrous oxide (N_2O) concentration in the OMZ has been
 60 reported to vary inversely with nitrite concentration¹³. Often as the oxygen levels diminish
 61 the ecosystem cannot sustain normal biotic inhabitants and macrofauna. As a result,
 62 OMZs are often associated with coastal and equatorial upwelling regions and the increased
 63 primary production rates determine the high levels of altered microbial metabolism^{11,14}.
 64 Importantly, Nitrogen (N) cycling plays crucial role in nitrate reduction to N_2
 65 (denitrification) and anaerobic ammonia oxidation (anammox) along with nitrate reduction
 66 to ammonia¹⁵. Moreover, nitrification has been shown to be an important source of
 67 oxidized N at the OMZ boundaries¹⁶⁻¹⁸.

68 Interestingly, various metagenomic studies on OMZ have revealed that complex
69 communities (such as nitrifiers) play an important role in N cycle in the OMZ¹⁸. Members
70 of the Planctomycetes, Thaumarchaeota and Nitrospinae phyla have been observed to
71 perform the majority of anammox, ammonia oxidation and nitrite oxidation and play
72 important role in the OMZ dynamics^{12,18-24}. Although some reports exist, the
73 denitrification²⁵⁻²⁷ and heterotrophic denitrification via a complete sequential reduction of
74 nitrate (NO₃) to N₂ has not been fully explored in the OMZ areas^{28,29}. A few studies have
75 been carried out to understand the microbial diversity in the OMZ areas of Arabian Sea³⁰⁻
76 ³⁴. The special growth requirements of these microbes and abundance of uncultured
77 organisms (over 99%) make NGS based metagenomic the method of choice in order to
78 unravel the complexities of microbial communities, their dynamics and ecological
79 significance.

80 In the current study, water samples collected from different depths of sea (100 to
81 1000 meters across transect) from Goa, Mangalore and Calicut (Supplementary
82 Information Table SI1) were processed for high throughput next generation sequencing
83 based metagenomics (based on 16S rRNA gene sequencing). The microbial diversity and
84 predicted metabolic activities associated with these microbial communities in OMZ and
85 non-OMZ areas in Arabian Sea of India provide valuable insight into the nature of
86 biogeochemical processes.

87

88 MATERIAL AND METHODS

89 *Sample Collection*

90 Water samples at different depths were collected during the Sagar Sampada cruise (Sagar
91 Sampada Cruse Number 340, 16 May 2015 to 08 June 2015) from Goa (GAS1, GAS2,
92 GAS3 and GAS4; distance from coast ranging from 51 km to 90 km), Mangalore (MGS5,

93 MGS6, MGS7 and MGS8; distance from coast ranging from 52 km to 84 km) and Calicut
 94 (CLS9, CLS10 and CLS11; distance from coast ranging from 66 km to 109 km)
 95 (Supplementary Information Table SI1; Fig. 1) A conductivity–temperature–depth (CTD)
 96 system equipped with attached oxygen and turbidity measurement sensors was deployed to
 97 record the physical properties of the water (Supplementary Information Table SI2) and the
 98 samples were grouped into OMZ and non OMZ .

99 *DNA extraction*

100 1000 ml water was collected from each sampling sites and organisms collected by filtering
 101 water through 0.22 µm filter (Millex, Merck Millipore, USA) were utilised for DNA
 102 isolation using Power water DNA isolation kit (MoBio laboratories Inc. Carlsbad, CA).
 103 DNA isolation was carried out on the ship to avoid degradation of DNA. DNA
 104 concentration was measured using the Quantus fluorimeter (Promega, USA).

105 *Amplification primers and Sequence analysis*

106 16s rRNA (corresponding to V3 and V4 regions) was amplified from total genomic DNA

107 isolated	(16S	Amplicon	PCR	Forward	Primer
--------------	------	----------	-----	---------	--------

108 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG3';

109 16S	Amplicon	PCR	Reverse	Primer
---------	----------	-----	---------	--------

110 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAA

111 TCC3') with appropriate sample bar coding index sequences and Illumina adapters.

112 AMPure XP beads were employed to remove unused primers and other unwanted nucleic

113 acid fragments and the purified PCR amplicons were quantified, normalized and an

114 equimolar pool of all the samples was made. This multiplexed library was further subjected

115 to QC using an Agilent Bioanalyzer DNA Chip. The sequencing libraries generated from

116 V3 and V4 amplicons from all the samples were sequenced using an Illumina paired end

117 overlapping sequencing. Sequence reads were binned according to index sequences and

118 QC of the raw sequence data was performed by custom scripts. Low quality reads were
119 filtered out and trimmed based on observed quality pattern in the data set. Read pairs with
120 high sequence quality and overlapping regions were fused together to obtain a single read
121 traversing full length of V3 and V4 region.

122 *Bioinformatics analysis*

123 The sequences which were less than 300 bps and sequences with less than average quality
124 score (25 or less) were removed from the library. The taxonomic assignment of
125 unassembled clean metagenomic sequences was performed using Ez-Taxone database³⁵
126 and BLASTX. Information related to the metagenomics reads of the samples is depicted in
127 Table 1.

128 *Statistical analysis*

129 Dominance, Simpson, Shannon, Evenness, Brillouin, Menhinick, Margalef, Equitability,
130 Fisher_alpha, Berger-Parker, Chao-1, Whittaker, Harrison, Cody, Routledge, Wilson-
131 Shmida, Mourelle, Harrison 2 and Williams indices of clonal and beta diversity were
132 estimated using the PAST3 programs available from the University of Oslo website link
133 Relationship between chemical composition and (i) species diversity unifracs distances, (ii)
134 species alpha diversity indices and (iii) species Beta diversity indices were determined by
135 Mantel tests. P values were calculated using 9999 permutations on rows and columns of
136 dissimilarity matrices. Principal coordinate analysis (PCO), canonical correlation analysis
137 (CCor), permutational analysis of variance (PERMANOVA) and analysis of similarity
138 (ANOSIM) was performed using the Past 3 software. For the predictive functional
139 analyses, the PICRUSt software package³⁶ was used to identify predicted gene families and
140 associated pathways.

141 *Analysis of predicated functional profiles for the identified microbial communities* The
142 16S rRNA sequencing data sets were analysed by PICRUSt script-

(normalize_by_copy_number.py script) for copy number normalization³⁶. Functional predictions were assigned up to KO tier 3 and categories including metabolism, genetic information processing, environmental information processing, and cellular processes were analysed further. KEGG Pathway analysis was carried out by employing functions .py PICRUSt scripts followed by STAMP (Statistical Analysis of Metagenomic Profiles) software³⁷, with Welch's t-test and P value cut-off of 0.05 was considered to reject null hypotheses. This identification of functional features of the genes and metabolic pathways has relevance in understanding metabolic processes in the context of the ecosystem.

151 *Identification of bacterial markers by LDA Effect Size (LEfSe) analysis*

152 Linear Discriminant Analysis (LDA) Effect Size (LEfSe) analysis was utilised for
153 identification of unique microbial communities present in different samples³⁸. The LEfSe
154 analysis with LDA score threshold of 2 using online Galaxy version 1.0 was used to
155 identify variations in bacterial diversity at specific locations and depths.

156

157 **3. RESULTS**

158 **3.1 Species diversity in Arabian Sea**

159 Water samples (total 11) collected from Arabian Sea at specific locations and depths
160 (Supplementary Information Table SI1, SI3) were subjected to metagenomic analysis using
161 next generation sequencing technology of amplified rDNA libraries. A total of 498062
162 (45278.36 ±5369.15 per sample) high-quality sequences with 3551 (1311.72±186.45 per
163 sample) distinct bacterial species were recorded (Supplementary Information Table 3) were
164 identified. This data was analysed extensively to identify if there were differences in the
165 OMZ and the non OMZ regions. In all three sampling sites in OMZ 1371 species were
166 common. while 777 species were found common at all depths (100m, 200m, 500m and
167 1000m) across different sampling sites (Fig. 2A, 2B).

168 *Community composition of the Arabian Sea*

169 A large number of uncultured and novel microbes were abundant at these locations
 170 (Supplementary Information SI3). Proteobacteria and SAR406 were common while
 171 Firmicutes, Spirochaetes, Chloroflexi and Verrucomicrobia were present in relatively
 172 lower numbers. Alpha proteobacteria (20.43-35.51%) (Fig. 2C) Deltaproteobacteria
 173 (11.03-15.93%) and Gamma proteobacteria (9.98-32.18%) were abundant in significant
 174 numbers in all the OMZ samples analysed. At the family level, SAR11-2_f (5.38-12.49%),
 175 Bacteria_uc_f (5.84-19.38%), Ruthia_f (0.69-7.56%), Arenicella_f (1.44-4.38%),
 176 Nitrospinaceae (1.86-5.89%), Erythrobacteraceae (0.45-4.80%) were present across all
 177 samples (Fig. 2B). Bacterial orders such as SAR11 (10.73-23.81%), Bacteria_uc_o (5.84-
 178 19.39%), Ruthia (0.70-7.91%), Alteromonadales (0.34-14.17%), Nitrospinaceae (1.96-
 179 6.11%) showed high abundance in all the samples. At 1000m depth SAR324_f (7.90-
 180 10.67%), Bacteria_uc_f (6.69-19.34%) and Erythrobacteraceae (0.94-4.80%) were
 181 predominant. Bacterial families such as Homogoneae and Thoreales were affiliated only
 182 with GAS4 sample whereas Synarophyceae, Ceramiales, Euglenida and Cloacamonas were
 183 exclusively present in CLS11 sample. Vaucheriales, Crenarchaeota, Pedinophyceae,
 184 Zetaproteobacteria and synergista were specific to MGS5 sample. Nitrospireae,
 185 Methanomicrobia, Bryospida were exclusive to 200m depth. SAR11-2_f (7.81-12.30%)
 186 and SAR11-1_f (6.31-12.32%) were predominant at 100m and 200m depth while
 187 Prochlorococcaceae (1.81-3.19%) was predominantly present at 100m depth.
 188 SAR406_o_uc (1.17-2.69%) was abundant at 200m depth.

189 Genera such as *Bacteria_uc_g* (5.85-19.38%), *Pelagibacter* (3.44-9.89%),
 190 *SAR324_g* (2.60-7.20%) were ubiquitous. *Croceicoccus* (1.25-2.14%) was predominantly
 191 present in samples from Goa as compared to other samples. Correspondence analysis
 192 revealed that in the depth of 1000 meters, *Methylopila*, *Mycoplasma*, *Asticcacaulis*,

193 *Cellulomonas*, *Phalanopsis* were exclusive to MGS7 sample whereas *Spirochaeta*,
 194 *Chroococcidiopsis*, *Thysira*, *Leeuwenhoekiella* were selectively present in CLS9 sample.
 195 Water sample at 1000 meters depth from Mangalore (MGS5) revealed presence of
 196 *Terasakiella*, *Chlorodendrales*, *Vaucheria*, *Congregibacter*, *Planktotelea*,
 197 *Pseudoflaviniifactor* whereas *Spirobacillus*, *Moraxella*, *Tiobacter*, *Roseburia*,
 198 *Marinoscillum*, *Thiohalophilus*, *Akkermansia*, *Caedobacter*, *Oceanicaulis*, *Epibacterium*,
 199 *Ditylium* were exclusively seen in Goa (GAS4). A more detailed analysis of data based at
 200 the species level revealed that *Bacteria_uc_s*, *SAR406_f_uc_s*, *Ruthia_f_uc_s*,
 201 *Arenicella_f_uc_s*, *Nitrospinaceae_uc_s*, *Oceanospirillaceae_uc_s* and
 202 *Rhodospirillaceae_uc_s* were present in high numbers in all samples that were analysed in
 203 our study (Fig. 2D, Supplementary Information Table SI3).

204 *Linear Discriminant Analysis (LDA) Effect Size (LEfSe) analysis*

205 In order to determine the unique and predominant bacteria present at a particular location,
 206 a comparative assessment of the biodiversity LefSe was carried out. This resulted in the
 207 identification of specific marker families for different locations as well as for the various
 208 sampling depths (Supplementary information SI2; SI3). Bacterial families including
 209 *Erysipelotrichi_uc_f*, *SAR11_uc*, *EU335161_o_uc*, *Pseudoalteromonadaceae* and
 210 *Alteromonadales_uc* were specific to Calicut while family *FJ444691_c_uc_f* was seen in
 211 Mangalore. Water samples from Goa *Salinisphaeraceae*, *EU686587_f*, and
 212 *Dehalococcoidales_uc* showed significant enrichment. LefSe analysis with respect to the
 213 depth showed enrichment of bacterial families (total 66) such as *Brumimicrobiaceae*,
 214 *Bacteriovoracaceae*, *Dinophysaceae*, *Spirochaetaceae*, and *Chaetocerotaceae* at shallower
 215 depth (100m) while bacterial families (In total 22) such as *Methylobacteriaceae*,
 216 *Halomonadaceae* and *Rhizobiaceae* were found to be enriched at the depth of 500-1000m.

217 *Alpha and Beta diversity of samples*

Alpha diversity analysis highlighted the rich taxonomic diversity in the sea samples (Supplementary Information Table SI4). Simpson index of all samples close to 1 for all samples indicated the presence of highly diverse microbial communities in samples. Shannon's index varied from 4.29 to 5.21 indicating the high species richness in bacterial diversity in these sea samples. Evenness index ranged from 0.093 to 0.138 while Margalef richness index was also high emphasising the richness of bacterial species in the sea area. Chao-1 analysis predicted the number of bacterial species in each sample to be between 1106-2122 (Supplementary Information Table SI4). No significant difference was observed in alpha diversity indices when pair wise comparison was carried out between Goa, Mangalore and Calicut sampling sites (ANOVA $P > 0.05$; Mann Whitney U test $P > 0.05$ for each comparison). Beta diversity indices of these sea samples are depicted in Supplementary Information Table SI5. At species level, high beta diversity was observed in all the sea samples (Supplementary Information Table SI5). This extensive analysis documented not only the rich and diverse micro flora present in each sample but also emphasised the differences in the microbial communities in the Arabian Sea.

Depth and geochemical parameter influences the community structure Principal component analysis (PCoA) led to the identification of depth as an important determinant which influences characteristic and typical community structures of a given niche. (Fig. 3) Samples from similar depth clustered together, indicating that the communities in these locations are very similar to each other. The correlations between environmental factors and alpha diversity indices were accessed by Canonical correlation analysis (CCorA) (Fig. 4A). Depth, turbidity and density were seen to influence the dominance of certain species while temperature and conductivity correlated with the richness and evenness in samples (Fig. 4 A). Beta diversity showed a correlation with geochemical characters of sample ($r = -0.262$; p value = 0.05) while alpha diversity ($r = -0.0004$; p value = 0.744) and unifrac

distances among the sampling sites ($r = -0.09$; $p = 0.517$) were not affected by the geochemical characters of sample (Fig. 4B). The sampling site did not influence the community composition while depth was a major factor (PERMANOVA ($F = 4.036$ $P = 0.0009$ ANOSIM $R = 0.7222$, $P = 0.0008$) (Table 2).

OMZ vs Non-OMZ samples

A comparative analysis and assessment of all samples showed presence of 2718 species in OMZ areas and 2223 species in non OMZ. It was seen that 1690 operational taxonomic units (OTUs) were common in OMZ and non OMZ while 1328 and 533 are unique to OMZ and non-OMZ respectively (Fig. 5A). This clearly documents that although several common inhabitants are seen in the ocean the depletion of oxygen is changing the species pattern. Differential abundance was clearly visible when the top 50 families present at these locations were compared (Supplementary Information Table SI3). In case of OMZ, SAR324, Ruthia, Arenicella, Zunongwangia, Rhodospirallaceae, Nirtitreductor, Oleiobacter, Hyphomonas, Methylophaga, Clamydiales, Xanthomonadeaceae, Sphingopya, Pararhodobacter, Anoxybacillus, Gemella, Phenyllobacterium and Legionella families were present while in non-OMZ Prochlorobacteriaceae, SAR11, Dianophyceae, pelagomonadeaceae, Delatproteobacteria, Firmicutes, Cytophagales, Flavobacteriales, Chroococcales, Dongicola, Planctomycetacia, Sphingobacteria, Draconibacterium families were unique. Out of total 203 families which showed differential abundance, 86 were present in OMZ whereas 117 in Non-OMZ respectively (Fig. 5B). The principal component analysis (PCoA) revealed that non OMZ samples from similar depths clustered closely together; indicating that the communities in these locations are similar to each other and the depth typically influences characteristic and typical community structures.

Functions associated with microbial communities

268 PICRUST analysis is a bioinformatics software package designed to predict functional
 269 content from microbial community identification carried out by 16S rRNA based
 270 metagenomic analysis. The percent OTUs associated with different metabolic functions are
 271 xenobiotics biodegradation (2-3%), glycan biosynthesis (2.3-3 %), energy metabolism (7-
 272 7.5%) and (3-45%) for lipid metabolism. Three terms under Environmental Information
 273 Processing contain membrane transport (9-13%), signal transduction (1-2%) and signalling
 274 molecules and interaction (0.04-0.06%). DNA sequences encoding proteins such as
 275 Nitrogenase FeMo-cofactor scaffold and assembly protein NifN, QscR quorum-sensing
 276 control repressor, Cobalt-zinc-cadmium resistance protein CzcA; Cation efflux system
 277 protein CusA, tellurite resistance protein-related protein, Nitrite reductase [NAD(P)H]
 278 large subunit (EC 1.7.1.4), Nitrogenase FeMo-cofactor synthesis FeS core scaffold, Sulfur
 279 deprivation response regulator proteins and assembly protein NifB were found
 280 predominantly in samples from Goa as compared to other samples. In sea water samples
 281 from Mangalore, gene sequences encoding activities such as Cobalt-zinc-cadmium
 282 resistance protein CzcD, Sirohydrochlorin cobaltochelate CbiK (EC 4.99.1.3) /
 283 Sirohydrochlorin ferrochelate (EC 4.99.1.4), type IV fimbrial biogenesis protein PilY1,
 284 predicted L-rhamnose ABC transporter, methyl-accepting chemotaxis protein III, phage
 285 tail protein were enriched as compared to other activities. Predominance of genes for 5-
 286 O-(4-coumaroyl)-D-quinic/3-hydroxylase (EC 1.14.13.36), beta-glucanase
 287 precursor (EC 3.2.1.73) (Endo-beta-1,3-1,4 glucanase) (1,3-1,4-beta-D-glucan 4-
 288 glucanohydrolase), glutathione S-transferase C terminus, Shikimate/quinic 5-
 289 dehydrogenase I beta (EC 1.1.1.282), Tlr0729 protein, two component system response
 290 regulator, putative Fe-S containing oxidoreductase, possible polygalacturonase (EC
 291 3.2.1.15), microbial collagenase (EC 3.4.24.3), chitinase, aspartate ammonia-lyase (EC

292 4.3.1.1), FtsK/SpoIIIE family protein, putative EssC component of type VII secretion
293 system were dominant in microbial communities from Goa and Calicut.

294 PICRUST and STAMP analysis have identified OTUs associated with few of these
295 KO terms differ significantly ($P < 0.05$) between samples collected from different locations.
296 On comparison of samples from Goa and Mangalore, Dioxin degradation and translation
297 proteins differed significantly, while processes related to polycyclic aromatic hydrocarbon
298 degradation enriched differentially in Calicut and Mangalore samples. In Calicut and Goa
299 samples bacterial OTUs associated with Phenylpropanoid biosynthesis, polycyclic
300 aromatic hydrocarbon degradation, Dioxin degradation showed differential enrichment.

301

302 **DISCUSSION**

303 Arabian Sea is typically characterised by presence of vast areas of OMZs and these are
304 expanding further. Depletion of oxygen in the habitats changes the microbial composition
305 and leads to alterations in the nutrient as well as elemental cycles. Analysis of the
306 correlation between the geochemical parameters and bacterial diversity is important in the
307 understanding the dynamics of microbial communities in the OMZs. This study
308 emphasises that Arabian Sea has high species richness with a complex community
309 structure across oxygen gradients and between the depths of sea (Fig. 2). Chao-1 analysis
310 highlighted presence of diverse assemblage of indigenous microbial species that remain
311 completely uncharacterized at present. Our analysis indicated that relationships between
312 environmental variables, conductivity, temperature and oxygen concentration have
313 significant role in increasing the species richness and evenness in microbial communities
314 (Fig. 4). OMZ samples in the Arabian sea displayed rich taxonomic diversity which
315 typically showed a depth specific variation.

316 Nitrate reducing bacteria were present at all collection sites in OMZ and non OMZ
 317 areas of Arabian Sea. Reports from the suboxic zone of the Black Sea have identified
 318 single clade of nitrifying Crenarchaeota which is closely related to *Nitrosopumilus*
 319 *maritimus*³⁹. Global Ocean Sampling (GOS) database across diverse physiochemical
 320 habitats and geographic locations has 1.2% *N. Maritimus*⁴⁰. Interestingly, *N. Maritimus* is a
 321 cultured nitrifier isolated from a marine aquarium⁴¹. It has been shown that *N. maritimus*
 322 typically dominates low depth samples²¹. However, in the current study, *N. maritimus* was
 323 underrepresented in low depth samples.

324 Based on the metagenomic profiles of microbial assemblage gene repertoire and
 325 predicted functions were assessed. Genes encoding nitrite/nitrate sensor proteins, nitrilase,
 326 nitrate reductase, nitrate reductase associated proteins were predominant in the datasets,
 327 emphasising that nitrate/nitrite metabolism plays a key role in the dynamics of microbial
 328 communities in the OMZ areas and play important role in nitrogen cycle in OMZ (Fig. 6).
 329 Naqvi *et al.*⁴² have reported the presence of the nepheloid layer with significant amounts
 330 of suspended matter caused by bacteria in Arabian Sea while an increase in nitrifying
 331 bacteria (both ammonium and nitrite oxidizers) has been suggested to be the cause for
 332 such nepheloid layer⁴³. Recent taxonomic, metagenomic, and metatranscriptomic analysis
 333 of many OMZs has shown that diverse sulphur-oxidizing microbial community are
 334 abundant and these communities are particularly enriched in γ -proteobacteria.
 335 Interestingly, the sulphate reducing bacteria (SRB) were present through-out the water
 336 column at all collection sites in our analysis. The presence of SRB has been shown not
 337 only at the bottom sediments but also in aerobic surface waters and beach sediments⁴⁴. It
 338 has been shown that SRB populations increase from the surface waters up to the oxic–
 339 anoxic boundary. Colourless sulphur-oxidizing bacteria have earlier been reported from the
 340 Arabian Sea and these bacteria are known to mediate nitrogen cycle reductively even under

341 autotrophic conditions⁴⁵. SRBs are also known to participate in nitrate reduction.
 342 Jayakumar *et al.*⁴⁶ and Ward *et al.*⁴⁷ reported the dominance of denitrifying bacteria in the
 343 biomass of the OMZ and suggested that the denitrifying bacteria in this zone could be in a
 344 viable but non culturable state.

345 Our analysis revealed that microorganisms involved in activities associated with
 346 sulphate metabolism were predominant with Sulphate permeases and reductase being
 347 predominant in the OMZ areas of Arabian Sea. Additionally, together with other recent
 348 analyses⁴⁸⁻⁵⁰, our results indicate the presence of an active sulphur oxidizing community in
 349 the Arabian sea OMZ. It is likely that sulphur cycle carried out by these SRB fuels nitrate
 350 reduction, thereby supplying additional substrates (nitrite and ammonia) for anammox
 351 bacteria. Comparative analysis of OMZ and Non-OMZ samples revealed that species such
 352 as *Zunongwangia profunda*, *Roseovarius nubinhibens*, *hydrocarbonoclasticus*,
 353 *Prochlorococcus marinus* and *Ruegeria pomeroyi* were common in oxygen depleted
 354 waters. These bacteria are known to be actively involved in dimethylsulfoniopropionate
 355 (DMSP) metabolism. This observation is important in the perspective of global climate
 356 change since DMS is thought to play a key role by decreasing the absorption of solar
 357 radiation and thereby influence temperature changes.

358 PICRUST analysis results obtained in the current study indicated presence of
 359 microorganisms harboring genes such as alkB, AlmA, CYP153A and AlkB are enriched in
 360 the OMZ in Arabian sea similar to the reports from Atlantic Ocean and Bay of Bengal⁵¹⁻⁵³.
 361 Presence of hydrocarbon degrading organisms points to presence of alkanes and
 362 hydrocarbon. In comparison to Calicut, Goa and Mangalore samples showed enrichment
 363 for polycyclic aromatic hydrocarbon degradation pathway suggesting anthropogenic
 364 activities in these areas. Deep sea water samples from depth 100-500 m depth were found
 365 to be enriched in (In total 22) Methylobacteriaceae, Halomonadaceae, Alcanivoracaceae

366 and Rhizobiaceae. Methylobacterium which derive energy from the oxidation of
 367 thiosulfate to sulfate. The current study provides baseline data related to the diversity and
 368 potential microbial communities in oxygen-depleted water which could provide a basis for
 369 better understanding of the microbiological function, dynamics, and distribution in the
 370 oceanic OMZ. Further the results obtained in this study indicate the location specific
 371 functional divergence in bacterial community and therefore it would be interesting to carry
 372 out detailed functional analysis for bacterial diversity from Arabian Sea at more locations
 373 and with multiple samples in different seasons. Although our understanding of the OMZ
 374 and interplay between geochemical processes and microbes has been improving in recent
 375 years, the potential impacts of OMZs on marine ecosystem structure and global
 376 geochemical cycling remains to be completely elucidated. In this context we need to
 377 accelerate the exploration and discovery of microbes and their interplay with geochemical
 378 processes in the OMZ.

379

380 **Acknowledgement**

381 We thank Prof. Dileep N. Deobagkar for valuable suggestions. The authors are grateful to
 382 Dr. M. Sudhakar, Dr. Saravannane, and Dr. R. A. Shivaji for their help during the sample
 383 collection.

384 **Author Contributions**

385 DDD, MSP designed the experiments; DDD, KA, MSP, SR, PM carried out the work;
 386 DDD, MSP, SR, PM interpreted the results; DDD, MSP wrote the manuscript.

387 **Funding**

388 The authors would like to acknowledge financial support provided by the Ministry of Earth
 389 Sciences (MoES), India under the Microbial Oceanography project. This project was
 390 coordinated through CMLRE.

391 **Conflict of interests**

392 The authors declare that they have no competing interests.

393

394 **References**

- 395 1. Naqvi S.W.A., Some aspects of the oxygen-deficient conditions and denitrification in
396 the Arabian Sea. *J. Mar. Res.* 1987, **45**, 1049–1072.
- 397 2. McCreary, J.P., Yu, Z., Hood, R.R., Vinaychandran, P.N., Furue, R., Ishida, A. and
398 Richards, K.J., Dynamics of the Indian-Ocean oxygen minimum zones. *Prog.*
399 *Oceanogr.*, 2013, **112**, 15-37.
- 400 3. Morrison, J.M., Codispoti, L.A., Gaurin, S., Jones, B., Manghnani, V. and Zheng, Z.
401 Seasonal variation of hydrographic and nutrient fields during the US JGOFS Arabian
402 Sea Process Study. *Deep Sea Res. II: Topical Stud. Oceanogr.*, 1998, **45**, 2053-2101.
- 403 4. Naqvi, W.A., Geographical extent of denitrification in the Arabian Sea in relation to
404 some physical processes. *Oceanol. Acta*, 1991, **143**, 281-290.
- 405 5. Stramma, L., Johnson, G.C., Sprintall, J. and Mohrholz, V., Expanding oxygen-
406 minimum zones in the tropical oceans. *Science*, 2008, **320**, 655-658.
- 407 6. Sarmiento, J.L., Hughes, T.M., Stouffer, R.J. and Manabe, S. Simulated response of
408 the ocean carbon cycle to anthropogenic climate warming. *Nature*, 1998, **393**, 245-
409 249.
- 410 7. Keeling, R.F., Körtzinger, A. and Gruber, N. Ocean deoxygenation in a warming
411 world. *Ann. Rev. Mar. Sci.* 2010, **2**,199-229.
- 412 8. DeSousa, S.N., Dileepkumar, M., Sardessai, S., Sarma, V.V.S.S. and Shirodkar, P.V.,
413 Seasonal variability in oxygen and nutrients in the central and eastern Arabian Sea.
414 *Curr. Sci.*, 1996, **71**, 847-851.

- 415 9. Morrison, J.M., Codispoti, L.A., Smith, S.L, Wishner, K, Flagg C, Gardner WD, *et*
416 *al.*, The oxygen minimum zone in the Arabian Sea during 1995. *Deep Sea Res. II:*
417 *Topical Stud. Oceanogr.*, 1999, **468**, 1903-1931.
- 418 10. Resplandy, L., Lévy, M., Bopp, L., Echevin, V., Pous, S., Sarma, V.V.S.S. and
419 Kumar D., Controlling factors of the oxygen balance in the Arabian Sea's OMZ.
420 *Biogeosci.*, 2012, **9**, 5095-5109.
- 421 11. Gilly, W.F., Beman, J.M., Litvin, S.Y. and Robison, B.H. Oceanographic and
422 biological effects of shoaling of the oxygen minimum zone. *Ann. Rev. Mar. Sci.*,
423 2013, **5**, 393-420.
- 424 12. Thamdrup, B., Dalsgaard, T., Jensen, M.M., Ulloa, O., Farias, L. and Escibano, R.
425 Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile.
426 *Limnol. Oceanogr.*, 2006, **51**, 2145-2156.
- 427 13. Cohen, Y. and Gordon, L.I., Nitrous oxide in the oxygen minimum of the eastern
428 tropical North Pacific: Evidence for its consumption during denitrification and
429 possible mechanisms for its production. *Deep-Sea Res.*, 1978, **25**, 509–524.
- 430 14. Diaz, R.J. and Rosenberg, R., Spreading dead zones and consequences for marine
431 ecosystems. *Science* 2008, **321**, 926-929.
- 432 15. Lam, P. and Kuypers, M.M. Microbial nitrogen cycling processes in oxygen
433 minimum zones. *Ann. Rev. Mar. Sci.*, 2011, **3**, 317–345.
- 434 16. Ward, B, and Zafiriou, O., Nitrification and nitric oxide in the oxygen minimum of
435 the eastern tropical North Pacific. *Deep Sea Res.*, 1988, **35**, 1127-1142.
- 436 17. Ward, B., Glover, H. and Lipschultz, F., Chemoautotrophic activity and nitrification
437 in the oxygen minimum zone off Peru. *Deep Sea Res.* 1989, **36**, 1031–1051.

- 438 18. Füssel, J., Lam, P., Lavik, G., Jensen, M.M., Holtappels, M., Günter, M., *et al.*,
439 Nitriteoxidation in the Namibian oxygen minimum zone. *ISME J.*, 2011, **6**, 1200–
440 1209.
- 441 19. Hamersley, M.R., Lavik, G., Woebken, D., Rattray, J.E., Lam, P., Hopmans, E.C., *et*
442 *al.*, Anaerobic ammonium oxidation in the Peruvian oxygen minimum zone. *Limnol.*
443 *Oceanogr.*, 2007, **52**, 923-933.
- 444 20. Lam, P., Lavik, G., Jensen, M.M., van de Vossenberg, J., Schmid, M., Woebken, D.,
445 *et al.*, Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc. Nat.*
446 *Acad. Sci. USA.*, 2009, **106**, 4752-4757.
- 447 21. Stewart, F.J., Ulloa, O., and DeLong, E.F., Microbial metatranscriptomics in a
448 permanent marine oxygen minimum zone. *Environ. Microbiol.*, 2012, **14**, 23-40.
- 449 22. Ganesh, S., Parris, D.J., DeLong, E.F. and Stewart, F.J., Metagenomic analysis of
450 size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.*, 2014, **8**,
451 187–211.
- 452 23. Hawley, A.K., Brewer, H.M., Norbeck, A.D., Paša-Toliae, L. and Hallam, S.J.,
453 Metaproteomics reveals differential modes of metabolic coupling among ubiquitous
454 oxygen minimum zone microbes. *Proc. Nat. Acad. Sci. USA*, 2014, **111**, 11395–
455 11400.
- 456 24. Ganesh, S., Bristow, L.A., Larsen, M., Sarode, N., Thamdrup, B. and Stewart, F.J.,
457 Size-fraction partitioning of community gene transcription and nitrogen metabolism
458 in a marine oxygen minimum zone. *ISME J.*, 2015, **9**, 2682–2696.
- 459 25. Jayakumar, D.A., Francis, C.A., Naqvi, S.W.A. and Ward, B.B., Diversity of nitrite
460 reductase genes *nirS* in the denitrifying water column of the coastal Arabian Sea.
461 *Aqua. Microb. Ecol.*, 2004, **34**, 69-78.

- 462 26. Jayakumar, A., Al-Rshaidat, M.M.D., Ward, B.B. and Mulholland, M.R., Diversity,
463 distribution, and expression of diazotroph nifH genes in oxygen-deficient waters of
464 the Arabian Sea. *FEMS Microb. Ecol.*, 2012, **82**, 597-606.
- 465 27. Kalvelage, T., Lavik, G., Jensen, M.M., Revsbech, N.P., Löscher, C., Schunck, H, *et*
466 *al.*, Aerobic microbial respiration in oceanic oxygen minimum zones. *PLoS One.*,
467 2015, **107**, e0133526.
- 468 28. Stevens, H. and Ulloa, O., Bacterial diversity in the oxygen minimum zone of the
469 eastern tropical South Pacific. *Environ. Microbiol.*, 2008, **10**, 1244–1259.
- 470 29. Glass, J.B., Kretz, C.B., Ganesh, S., Ranjan, P., Seston, S.L., Buck, K.N., *et al.*,
471 Meta-omic signatures of microbial metal and nitrogen cycling in marine oxygen
472 minimum zones. *Front. Microbiol.*, 2015, **6**, 1-13.
- 473 30. Carlson, C.A., Morris, R., Parsons, R., Treusch, A.H., Giovannoni, S.J. and Vergin,
474 K., Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones
475 of the northwestern Sargasso Sea. *ISME J.* 2009, **3**, 283-295.
- 476 31. Gonsalves, M.J., Paropkari, A. L., Fernandes, C.E.G., LokaBharathi, P.A.,
477 Krishnakumari, L., Fernando, V., *et al.*, Predominance of anaerobic bacterial
478 community over aerobic community contribute to intensify ‘oxygen minimum zone’
479 in the eastern Arabian Sea. *Continental Shelf Res.* 2011, **31**, 1224-1235.
- 480 32. Pitcher, A., Villanueva, L., Hopmans, E.C., Schouten, S., Reichart, G.J., Sinninghe,
481 J. and Damste, S., Niche segregation of ammonia-oxidizing archaea and anammox
482 bacteria in the Arabian Sea oxygen minimum zone. *ISME J.*, 2011, **5**, 1896–1904.
- 483 33. Wyman, M., Hodgson, S. and Bird C., Denitrifying Alphaproteobacteria from the
484 Arabian Sea that express nosZ, the gene encoding nitrous oxide reductase, in oxic
485 and suboxic waters. *App. Environ. Microbiol.*, 2013, **798**, 2670-2681.

- 486 34. Lüke, C., Speth, D.R., Kox, M.A.R., Villanueva, L. and Jetten, M.S.M.,
487 Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen
488 minimum zone. *PeerJ.*, 2016, **4**, e1924. doi 10.7717/peerj.1924
- 489 35. Kim, O.S., Cho, Y.J., Lee, K., Yoon, S.H., Kim, M., Na, H., *et al.*, Introducing
490 EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that
491 represent uncultured species *Internat. J. Syst. Evol. Microbiol.* 2012, **62**, 716-721,
492 doi: 10.1099/ijs.0.038075-0
- 493 36. Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes,
494 J.A., *et al.*, Predictive functional profiling of microbial communities using 16S rRNA
495 marker gene sequences. *Nature Biotechnol.* 2013, **31**,814–821.
- 496 37. Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M.
497 Data, information, knowledge and principle: back to metabolism in KEGG. *Nucl.*
498 *Acids Res.*, 2014, **42(D1)**, D199–D205.
- 499 38. Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S. and
500 Hutten, C., Metagenomic biomarker discovery and explanation. *Genome Biol.*, 2011,
501 **12**:R60.
- 502 39. Labrenz, M., Sintes, E., Toetzke, F., Zumsteg, A., Herndl, G.J., Seidler, M. and
503 Jürgens, K., Relevance of a crenarchaeotal subcluster related to *Candidatus*
504 *Nitrosopumilus maritimus* 830 to ammonia oxidation in the suboxic zone of the
505 central Baltic Sea. *ISME J.*, 2010, **412**, 1496-1508.
- 506 40. Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., *et al.*,
507 *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and
508 autotrophy in globally distributed marine crenarchaea. *Proc. Nat. Acad. Sci. USA.*,
509 2010, **107**, 8818-8823.

- 510 41. Konneke, M.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B. and Stahl, D.A.
511 Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, 2005, **437**,
512 543-46.
- 513 42. Naqvi, S.W.A., Kumar, M.D., Narvekar, P.V., de Souza, S.N., George, M.D., and
514 D'Silva, C., An Intermediate nepheloid layer associated with high microbial
515 metabolic rates and denitrification in the Northwest Indian Ocean. *J. Geophys. Res.*,
516 1993, **98**, 16469-16479.
- 517 43. Koppelman, R., Zimmermann-Timm, H. and Weikert, H. Bacterial and zooplankton
518 distribution in deep waters of the Arabian Sea. *Deep-Sea Res. I*, 2005, **52**, 2184–
519 2192.
- 520 44. LokaBharathi, P.A. and Chandramohan, D., Sulfate-reducing bacteria from the
521 Arabian Sea-their distribution in relation to thiosulfate-oxidising and hetero-trophic
522 bacteria. *Bull. Mar. Sci.*, 1990, **47**, 622–630.
- 523 45. LokaBharathi, P.A., Nair, S. and Chandramohan, D., Anaerobic sulfide-oxidation in
524 marine colorless sulfur-oxidizing bacteria. *J. Mar. Biotechnol.*, 1997, **52/3**, 172-177.
- 525 46. Jayakumar, A., O'Mullan, G.D., Naqvi, S.W.A. and Ward, B.B., Denitrifying
526 bacterial community composition changes associated with stages of denitrification in
527 oxygen minimum zones. *Microb. Ecol.*, 2009, **58**, 350–362.
- 528 47. Ward, B.B., Devol, A.H., Rich, J.J., Chang, B.X., Bulow, S.E., Naik, H., Pratihary,
529 A. and Jayakumar, A., Denitrification as the dominant nitrogen loss process in the
530 Arabian Sea. *Nature*, 2009, **461**, 78–81.
- 531 48. Lavik, G., Stuhmann, T., Bruchert, V., Van der Plas, A., Mohrholz, V., Lam, P. *et*
532 *al.*, Detoxification of sulphidic African shelf waters by blooming chemolithotrophs.
533 *Nature*, 2009, **457**, 581-586.

- 534 49. Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C., Wright, J.J., Tringe, S.G., *et al.*,
535 Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead
536 Zones. *Science*, 2009, **326**, 578-582.
- 537 50. Canfield, D.E., Stewart, F.J., Thamdrup, B., De Brabandere, L., Dalsgaard, T.,
538 DeLong, E.F., *et al.*, A cryptic sulfur cycle in oxygen-minimum-zone waters off the
539 Chilean coast. *Science*, 2010, **330**, 1375-1378.
- 540 51. Wang, L.P., Wang, W.P., Lai, Q.L. and Shao, Z.Z., Gene diversity of CYP153A and
541 AlkB alkane hydroxylases in oil-degrading bacteria isolated from the Atlantic Ocean.
542 *Environ. Microbiol.*, 2010, **12**, 1230-1242.
- 543 52. Wang, W.P. and Shao, Z.Z., Diversity of flavin-binding monooxygenase genes
544 (almA) in marine bacteria capable of degradation long-chain alkanes. *FEMS Microb.*
545 *Ecol.*, 2012, **80**, 523–33.
- 546 53. Rajpathak, S.N., Banerjee, R., Mishra, P.G., Khedkar, A. M., Patil, Y.M., Joshi, S.R.
547 and Deobagkar, D.D. An exploration of microbial and associated functional diversity
548 in the OMZ and non-OMZ areas in the Bay of Bengal. *J. Biosci.*, 2018, 1-14
549 <https://doi.org/10.1007/s12038-018-9781-2>.
- 550
- 551
- 552

Figure Legends

Fig. 1. Location of sampling sites in Arabian Sea. (Distance from coast is depicted in Supplementary Information Table SII)

557 **Fig. 2.** Venn diagrams showing the number of unique and shared species between the (A)
 558 three sampling sites (B) different depth of sea. Distribution of predominant bacterial class
 559 in sample based on 16S rRNA gene sequencing. (C) Phyla distribution. Observations are
 560 displayed as stacked bar charts for individual mangrove sample (x-axis) against the taxa
 561 abundance (y-axis). (D) Abundance of species. Observations are displayed as scatter plot
 562 for individual sample (x-axis) against the species abundance (y-axis).

563

564 **Fig. 3.** Principal Coordinates Analysis (PCoA) representation of the similarity matrix
 565 generated by cluster analysis. Samples from (A) depth and (B) collection site are
 566 represented by a different shape, and the distance between dots represents relatedness
 567 obtained from the similarity matrix.

569 **Fig. 4.** Association of environmental parameters and diversity indices (A) Canonical
570 correlation analysis (CCorA) (B) Mantel Test

573 **Fig. 5.** Comparison of differentially present microbial diversity in OMZ and non-OMZ
574 areas

575

576

577 **Fig. 6.** Probable nitrogen cycle in oxygen depleted zones in Arabian Sea.

578

579

580 **TABLE 1. Metagenomics reads information and taxonomic affiliations of bacteria present in samples collected from Arabian Sea**

Place	Site	Valid reads	OTUs	Average Read Length	Goods Library Coverage	Bacteria			
						Order	Family	Genus	Species
Goa									
Goa	GAS1	48317	39457	423.53	0.23528	278	594	973	1390
Goa	GAS2	47364	38991	422.3	0.224305	334	673	1003	1338
Goa	GAS3	47953	38157	419.8	0.259608	332	712	1081	1460
Mangalore	GAS4	44960	34992	416.53	0.283029	285	563	897	1256
Mangalore	MGS5	43168	36154	421.56	0.208024	319	668	1045	1421
Mangalore	MGS6	47834	38732	421.25	0.242735	314	653	1052	1459
Mangalore	MGS7	48669	40730	423.61	0.213277	291	569	923	1288
Calicut	MGS8	48702	39181	421.18	0.249805	265	565	931	1333
Calicut	CLS9	30013	27721	438.47	0.087662	235	451	637	786
Calicut	CLS10	45356	38356	423.02	0.198585	318	633	961	1318
Calicut	CLS11	45726	37224	421.58	0.236824	294	608	965	1380

581

582 **TABLE 2. Effect of depth and sampling location on bacterial diversity of**
583 **samples collected from Arabian Sea**

Test	F	P
Permanova		
Sampling Site	1.041	0.416
Depth	3.503	0.0002
Pair wise comparison (t test)		
Goa vs Magalore	0.9932	0.5742
Goa vs Calicut	1.294	0.2914
Calicut vs Mangalore	0.8273	0.6308
100m vs 200m	4.952	0.0278
100m vs 500m	5.166	0.2496
100m vs 1000m	3.771	0.1022
200m vs 500m	2.472	0.1949
200m vs 1000m	3.668	0.0273
500m vs 1000m	1.75	0.2585

584