Check for updates

Diversity, structure, and distribution of bacterioplankton and diazotroph communities in the Bay of Bengal during the winter monsoon

Chao Wu[1,2], Dhiraj Dhondiram Narale[3], Zhengguo Cui[1,2], Xingzhou Wang[3,4], Haijiao Liu[3,4], Wenzhe Xu[3], Guicheng Zhang[3] and Jun Sun[3,4,5]*

[1]Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, [2]Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology, Qingdao, China, [3]Research Centre for Indian Ocean Ecosystem, Tianjin University of Science and Technology, Tianjin, China, [4]Institute for Advanced Marine Research, China University of Geosciences, Guangzhou, China, [5]State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, China

The Bay of Bengal (BoB) is conventionally believed to be a low productive, oligotrophic marine ecosystem, where the diazotroph communities presumed to play a vital role in adding "new" nitrogen through the nitrogen fixation process. However, the diazotroph communities in the oceanic region of the BoB are still poorly understood though it represents most of the seawater volume. The present study investigated a detailed account of the bacterioplankton community structure and distribution in the oceanic BoB during the winter monsoon using high throughput sequencing targeting the 16S rRNA and *nif*H genes. Our study observed diverse groups of bacterioplankton communities in the BoB including both cyanobacterial and non-cyanobacterial phylotypes. Cyanobacteria (*Prochlorococcus* spp. and *Synechococcus* spp.) and Proteobacteria (mainly α-, γ-, and δ-Proteobacteria) were the most abundant groups within the bacterial communities, possessing differential vertical distribution patterns. Cyanobacteria were more abundant in the surface waters, whereas Proteobacteria dominated the deeper layers (75 m). However, within the diazotroph communities, Proteobacteria (mainly γ-Proteobacteria) were the most dominant groups than Cyanobacteria. Function prediction based on PICRUSt revealed that nitrogen fixation might more active to add fixed nitrogen in the surface waters, while nitrogen removal pathways (denitrification and anammox) might stronger in deeper layers. Canonical correspondence analysis (CCA) indicated that temperature, salinity, and silicate were major environmental factors driving the distribution

of bacterial communities. Additionally, phosphate was also an important factor in regulating the diazotroph communities in the surface water. Overall, this study provided detailed information on bacterial communities and their vital role in the nitrogen cycles in oligotrophic ecosystems.

Introduction

Seasonal reversal monsoon systems and the resulting meso- or basin-scale sea surface circulation patterns in the Bay of Bengal (BoB) are the most fascinating climatological features of the seas around the world (Schott and Mccreary, 2001; Vecchi and Harrison, 2002; Chen et al., 2018). The hydrography and circulation are influenced by the seasonal reversal monsoons in the BoB, with poleward flowing East Indian Coastal Current (EICC), eastward Southwest Monsoon Current during June-September (summer monsoon) and equatorward EICC, and westward Northeast Monsson Currents during December-February (winter monsoon) (Shankar et al., 2002; Miyama et al., 2003; Schott et al., 2009). This reversal pattern of monsoon currents plays a significant role in the water mass exchange between the BoB and the Arabian Sea (AS) (Shankar et al., 2002; Prasanna et al., 2004). Except for the two monsoons, winds over the northern Indian Ocean are generally weak and consistent westerly during the transition (inter) monsoon periods, i.e., spring (March-May) and fall (September-October) (Schott et al., 2009). The surface water during these inter-monsoon periods is weakly churned by the persistently steady zephyr, and hence most regions in the bay are calm, stratified, and oligotrophic (Narvekar and Prasanna, 2006, 2014; Jinadasa et al., 2020). In general, winds in the northern Indian Ocean are strongest in the summer monsoon, intense in the winter monsoon, and weakest in the transition periods (Vecchi and Harrison, 2002; Singh and Ramesh, 2015).

Though the BoB is located in the same latitudinal belt as the AS, the former is influenced by a more intense monsoon rainfall and river runoff (Prasanna et al., 2002; Wijesekera et al., 2015; Sarma et al., 2016). The immense annual freshwater runoff from the Himalayan Rivers (Ganges and Brahmaputra) and other peninsular rivers (Mahanadi, Krishna-Godavari, etc.) into the BoB decreases average surface salinity on a large scale (Prasanna et al., 2002; Narvekar and Prasanna, 2014; Jinadasa et al., 2020). The freshwater capping creates a strong and highly stable "barrier layer" in the upper layers of the northern BoB, thereby inhibiting the vertical turbulent transport of nutrients, heat, momentum, and tracers (Prasanna et al., 2002; Kumari et al., 2018). Even under strong summer monsoon wind forcing, surface water in the BoB is weakly churned, and

hence restricts the vertical transport of nutrients. More than that, major physical processes such as summer coastal upwelling and winter convective mixing are very weak in the BoB, but alternatively, eddies, gyres, and other episodic events dominate in this region and make the bay locally productive (Shetye et al., 1991; Prasanna et al., 2002, 2004; Vidya and Kumar, 2013; Singh et al., 2015; Chen et al., 2018). In brief, the BoB is conventionally believed to be an oligotrophic system relative to the AS, partly due to the depletion of surface nutrients, intense cloud cover, narrow shelf region, and high turbidity waterbody jointly restricting the growth of phytoplankton (Prasanna et al., 2002; Madhupratap et al., 2003; Gauns et al., 2005; Narvekar and Prasanna, 2014; Bhushan et al., 2018).

The other characteristic of the BoB is that subsurface waters contain extremely low but persistent concentrations of dissolved oxygen (DO) (Löscher et al., 2020). Low DO concentrations have profound influences on the marine microbial community which could eventually change marine biogeochemical cycling such as nitrogen and sulfur cycling (Cheung et al., 2015; Fernandes et al., 2020). Bristow et al. (2016) found that OMZs in the BoB support denitrifies and anammox microbial communities, mediating low but significant nitrogen loss. One more recent research reported a significant variation in nitrogen and sulfur metabolism in the OMZs of the BoB (Fernandes et al., 2019, 2020). Further analysis demonstrated that δ-Proteobacteria and γ-Proteobacteria affiliated with Proteobacteria showed high abundance in the deoxygenated waters in OMZs of BoB (Fernandes et al., 2020). Bacterial communities have also been explored in the coastal and offshore waters of the BoB, with *Synechococcus*, *Erythrobacter*, and *Psychrobacter* dominating in the coastal region while *Prochlorococcus* and *Vibrio* in the offshore waters (Vijayan et al., 2020). However, these studies focused on specific regions (e.g., coastal region, offshore region, and OMZs), no study was conducted in the open regions though they represent most of the seawater volume of the BoB.

Most parts of the BoB, especially in the open region, are lacking bioavailable nitrogen in the surface waters (Löscher et al., 2020). Thus, the nitrogen-transforming microorganisms which control the balance of bioavailable nitrogen are particularly important in this oligotrophic ecosystem (Wu et al., 2018, 2019). Diazotroph communities play a vital role

in sustaining primary productivity by adding new nitrogen to oligotrophic marine ecosystems (Levitan et al., 2007; Gradoville et al., 2017). Earlier studies had reported that the globally significant marine nitrogen fixer *Trichodesmium* spp. inhabits the northern Indian Ocean throughout the year (Hegde et al., 2008; Achary et al., 2014; Sahu et al., 2017). Particularly, this genus forms periodic blooms in the BoB during the spring inter-monsoon periods (Li et al., 2012; Jyothibabu et al., 2017). Compared with the high frequency of blooms during the inter-monsoon season, *Trichodesmium* spp. blooms are a small probability event in the monsoon seasons. Many studies showed that *Trichodesmium* spp. usually thrives in a relatively stable environment (Jyothibabu et al., 2017; Wu et al., 2018). The unusually strong winds and high turbulent surface water during the active monsoon seasons are not favorable for the flourishing of *Trichodesmium* spp. In addition to *Trichodesmium* spp., symbiotic diatom-diazotrophic cyanobacteria associations were also reported widely distributed in the northern Indian Ocean (Madhu et al., 2013; Manjumol et al., 2018). Other diazotrophs, for instance, unicellular diazotrophic cyanobacteria (UCYN) and heterotrophic diazotrophs (various Proteobacteria populations), are less studied in the BoB.

This study presents a detailed account of bacterioplankton communities (including diazotrophs) in the oligotrophic BoB waters during the winter monsoon. Furthermore, to better understand the variation within the diazotrophs in different months, the result was compared with the earlier study carried out during the spring inter-monsoon (Wu et al., 2019). The high throughput sequencing was used to assess the detailed information of dominant, rare species and many uncultured bacterioplankton populations. In addition, quantitative methods including real-time fluorescent quantitative polymerase chain reaction (qPCR) assay and microscope counting were also used to determine the factual abundance of diazotrophs. Thus, overall, this study will enhance our understanding of spatial-temporal heterogeneity and distribution of bacterioplankton communities and their vital roles in nitrogen cycling in oligotrophic marine ecosystems.

Materials and methods

Sampling strategies

To better understand the bacterioplankton communities in the BoB, sampling was carried out during R/V "*Dongfanghong II*" (8 October to 20 December 2016) at 24 stations located between 2 and 18°N along 88°E (**Figure 1**). All 24 stations were sampled for the assessment of the environmental parameters, whereas molecular samples were collected only at seven selected stations (details in **Figure 1**). Water samples were collected at 6 depths (0, 30, 75, 100, and 150–300 m) using 12-L Go-Flo bottles
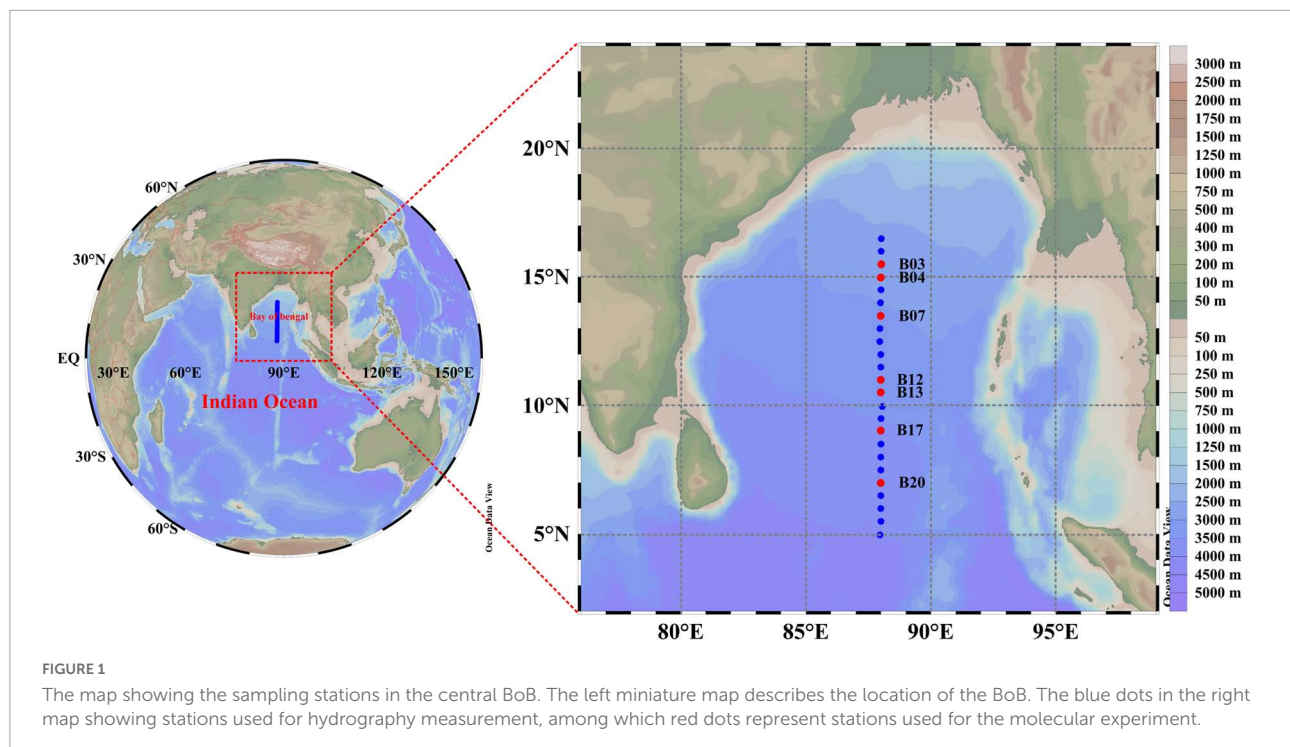
attached to a rosette multi-sampler installed with CTD probes (Seabird SBE 911Plus, Sea-Bird Electronics, Inc., Bellevue, WA, USA). Environmental parameters such as temperature and salinity were recorded vertically with the CTD sensor. The vertical profiles of the environmental parameters were visualized by Ocean Data View (ODV V5.0.0; Schlitzer, 2008). Collected water samples were transferred into 15-L HCl-rinsed buckets. For nutrient measurement, subsamples were transferred into 100-ml HCl-rinsed bottles and kept in a refrigerator until further processing. To quantify the filamentous cyanobacteria, a 1-L water sample was fixed with 1% formaldehyde in a plastic bottle and stored in dark. For chlorophyll a (Chl$a$) analysis, 1-L subsamples were vacuum-filtered (<100 mm Hg) through a 25-mm Whatman^{TM} GF/F filter (Whatman, Florham Park, NJ, USA). The filters were packed in aluminum foil to prevent light and stored at 4°C until further analysis. For molecular analysis, 2–4 L subsamples were filtered through 0.22-μm polycarbonate filters (Millipore, Eschbonn, Germany) under a low-pressure vacuum (<100 mm Hg). The filters were placed into 20-ml microtubes, flash-frozen immediately in liquid nitrogen, and stored at –80°C in the lab until analysis.

Environmental parameter measurement

The nutrients, including nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4), and phosphate (PO_4^{3-}), were analyzed with Technicon AA3 Auto-Analyzer (Bran+ Luebbe) using the standard methods (Grasshoff et al., 1982). NO_3^-, NO_2^-, and NH_4 were measured using the copper-cadmium column reduction methods, while PO_4^{3-} was measured by typical spectrophotometric methods. For Chl$a$ estimation, filters were extracted using the acetone (90%) extraction method. The Chl$a$ concentrations were measured using a Trilogy (CHL NA, Model # 046) fluorometer.

Deoxyribonucleic acid extraction, polymerase chain reaction amplification, and sequencing

Genomic DNA was extracted using DNeasy PowerWater® Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The quality and quantity of the extracted DNA were measured using an ND-2000 Nanodrop spectrometer (Thermal Scientific, Wilmington, DE, USA). A total of 14 samples from two depths (0 and 75 m) at each station were used for high throughput sequencing analysis. Two genes (16S rRNA and *nif*H) were amplified from the 14 samples in the present study. To characterize the total bacterioplankton composition, the V3–V4 region of the 16S rRNA gene was amplified with the pair-wise common primer 338F (5′-ACTCCTACGGGAGGCAGCA-3′)

FIGURE 1
The map showing the sampling stations in the central BoB. The left miniature map describes the location of the BoB. The blue dots in the right map showing stations used for hydrography measurement, among which red dots represent stations used for the molecular experiment.

and 806R (5′-GGACTACHVGGGTWTCTAAT-3′) (Mori et al., 2014). Reactions contained 5X Q5 Reaction Buffer (10 µl), 5X Q5 High GC Enhancer (10 µl), 10 µM forward and reverse primers (1.5 µl each), 10 µM dNTPs (1 µl), Q5 High-Fidelity DNA Polymerase (0.2 µl) (New England Biolabs, Whitby, ON, Canada), 40–60 ng of DNA templates, and nuclease-free water to adjust the final volume of 50 µl. The thermal profile used for the 16S rRNA gene amplification was denatured (95°C, 5 min), followed by 25 cycles of denaturation (95°C, 30 S), annealing (50°C, 30 S), and extension (72°C, 40 S) with a final extension (72°C, 7 min). The diazotroph community composition was defined by the *nif*H gene. The *nif*H gene fragment was amplified by nested PCR using the protocol given by Zehr et al. (1998). The details of the PCRs reaction system and thermal profile are illustrated by Wu et al. (2019). Note that negative controls were set up by replacing template DNA with nuclease-free water. All the PCR amplifications were conducted in a Veriti 9902 thermocycler (Applied Biosystems, Foster City, CA, USA). Remarkably, primers used in the 16S rRNA gene amplification and the second round of the *nif*H gene amplification were composed of dual-indexed barcodes to distinguish different samples. After amplification, all the PCR products were checked in 1.8% agarose gel. Samples with bright bands of approximately 450 bp for the 16S rRNA gene and 360 bp for the *nif*H gene were thought to be successfully amplificated. In addition, negative controls have no obvious bands that were considered contamination free. The PCR products were purified by MinElute® PCR Purification Kit (Qiagen, Hilden, Germany), and the purified PCR products

were validated using Nanodrop 2000. Finally, the libraries were sequenced using paired-end chemistry (PE250) on an Illumina Hiseq2500 platform (Illumina, San Diego, California, USA) at Biomarker Technologies, Beijing, China. The 16S rRNA gene and *nif*H gene raw sequences obtained from this study were deposited in NCBI Sequence Read Archive with BioProject no. PRJNA637963 and PRJNA637983, respectively.

Quality control and bioinformatics analysis

The 16S rRNA gene data were analyzed in the BMK Cloud.[1] The detailed procedure of the bioinformatics analysis for the 16S rRNA gene is given in Wu et al. (2022a). For *nif*H gene analysis, the raw sequence data were first quality filtered to remove low-quality tags. Then these sequences were separated by samples, according to their barcode sequences, permitting up to one mismatch (Zhang et al., 2017). The paired-end reads were subsequently merged into full-length sequences by FLASH v1.2.7 software to get raw tags (Magoč and Salzberg, 2010). The minimal overlapping length and maximum mismatch ratio in this step were 10 bp and 0.2, respectively. Notably, the paired-end reads without overlaps were removed from the pool. To remove low-quality raw tags, Trimmomatic v0.33 software was used to filter sequences less than 300 bp in length, or that

---

1   http://www.biocloud.net/

contained homopolymers longer than 8 bp and ambiguous base-pair (Huse et al., 2007; Kunin et al., 2010; Bolger et al., 2014). The chimera sequences were also removed from the raw tags by comparing tags with the reference database in the UCHIME v4.2 software (Edgar et al., 2011), and the remaining effective tags were grouped into operational taxonomic units (OTUs) at 97% similarly by USEARCH v10.0 (Edgar, 2010). In the present study, the most common sequences in each OTU were selected as representative sequences.

For taxonomic classification of the diazotroph communities, representative sequences were first translated into amino acid sequences and searched in the protein sequences database of the National Center for Biotechnology Information (NCBI) using BLASTX v2.8.1+ (Altschul et al., 1997). The most closely related sequences (>96% similarly) were chosen as the alignment sequences. Finally, the representative and alignment sequences were aligned with ClustalW in MEGA v7.0, and a phylogenetic neighbor-joining tree was subsequently constructed using the maximum likelihood method (Kumar et al., 2016). Bootstrap values were determined by resampling 1,000 times, and bootstrap values greater than 50% were shown near nodes. The constructed tree was further edited by the Interactive Tree of Life (iTOL), an online tool for managing the phylogenetic tree (Letunic and Bork, 2011).

Statistical analysis

Rarefaction curves were calculated by Past V3.0 software and further visualized in the online software based on the standard operating procedure shown on the website.[2] Alpha and beta diversity were calculated based on the OTU tables for comparing the relative complexity of bacterioplankton communities. Alpha diversity including the Chao1 richness estimator, Shannon-Weiner diversity index, and Simpson similarity index were calculated in the R V3.3.2 software. Non-metric multidimensional scaling (NMDS) analysis was conducted in PRIMER V6.0 software to characterize the vertical and horizontal distribution patterns of bacterioplankton communities (Clarke and Gorley, 2006). To evaluate the relationship between environmental factors and bacterial-diazotroph communities, canonical correspondence analysis (CCA) was carried out using CANCO 4.5 software. The detailed procedure of CCA is given by Kong et al. (2011) and Wu et al. (2019). The network was used to explore the co-occurrence patterns of bacterioplankton communities by Gephi v0.9.2. For convenience, only the OTUs with relative abundance greater than 0.1% were selected for network analysis (Jiao et al., 2016). The possible pairwise Spearman's rank correlation (r) between OTUs was calculated within the package of *psych* in R V3.3.2 software. Only statistically robust ($|r| > 0.9$ for bacterial

---

2 https://chao.shinyapps.io/iNEXTOnline/

communities, and $|r| > 0.7$ for diazotroph communities) and significant ($P < 0.01$) correlations were included in the further analysis. The topology of the network correlations was visualized in Gephi. Furthermore, node-level topological properties were also obtained in the Gephi. Nodes with high degree and low betweenness centrality values in the networks were identified as keystone species (Wu et al., 2022a). The functions of bacterial communities were predicted by PICRUSt genome prediction software v0.9.2 (Natale et al., 2000). The predicted proteins related to nitrogen cycles were discovered and visualized by heatmap using R v3.6.1 software.

Quantification of *nif*H phylotypes

The filamentous cyanobacteria population was microscopically quantified and enumerated from the preserved seawater samples. Prior to the analysis, the 1-L seawater sample was concentrated to a 200 ml solution by siphoning method (Liu et al., 2020). Then, 100-ml concentrated solution was transferred into an Utermöhl chamber for 24 h sedimentation (Uthermöhl, 1958). Finally, the filamentous cyanobacteria and their corresponding cells were identified and counted under an inverted microscope (Motic, AE2000) using 200 or 400 magnifications.

The quantity of unicellular diazotrophs, qPCR targeting UCYN-A, UCYN-B, and two uncultured proteobacteria were conducted using an ABI Step One Plus Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). The corresponding *nif*H standards were obtained from the clone library of environmental samples, except for the alpha-proteobacteria, which directly use the specific primers amplified from the environmental samples. The specific primers, probes, and standard clones used in the present study are described in **Table 1**. The qPCR reactions were performed in duplicate in a final volume of 10 μl which contained 5 μl 2 × Premix Ex Taq^{TM} (Takara Bio, Tokyo, Japan), 50 × ROX Reference Dye (0.2 μl), 10 μM forward and reverse primers (0.4 μl), TaqMan probe (0.4 μl), template DNA (1 μl), and finally nuclease-free water to adjust the final volume. qPCR condition was denatured (95°C, 30 S), followed by 45 cycles of denaturation (95°C, 5 S), and annealing (60°C, 30 S) (Wu et al., 2019). Standard curves were determined by 10-fold dilution series from 10 to 10^7 gene copies per reaction. The linear regression (R^2) values and amplification efficiencies (E) of each standard curve greater than 0.99 and 90% were thought effective. The amplification efficiency was calculated by the equation $E = 10^{-1/m} - 1$, where m is the slope of the standard curve. Non-target templates were also tested in the same condition as the standards and samples. Where amplification of non-target templates occurred (Ct values ranged from 35 to 38), the non-target template gene copies were subtracted from the sample values to adjust for slight

TABLE 1 Primers, Taqman probes, and standard clones for quantitative polymerase chain reaction (qPCR) analysis targeting the *nif*H gene of different cyanobacterial diazotrophic groups.

Targets	Forward primer (5′–3′)	Probe	Reverse primer (5′–3′)	References
UCYN-A	GGTTACAACAACGTTTTATGTGTTGA	TCTGGTGGTCCTGAGCCCGGA	GCAGTAATAATACCACGACCAGCAC	Goebel et al. (2010)
UCYN-B	TGGTCCTGAGCCTGGAGTTG	TGTGCTGGTCGTGGTAT	CTTCTTCTAGGAAGTTGATGGAGGTG	Goebel et al. (2010)
Sagittula castanea	ATCACCGCCATCAACTTCCT	CGCCTACGATGACGTGGATTACGTGTCC	AGACCACGTCGCCCAGAAC	Zhang et al. (2011)
γ-HM210363	CCTGGACTTCGTATTCTA	CGATGTTGTATGCGGTGGCTT	CCACAATGTAGATTTCCTG	Halm et al. (2012)

The 5′ and 3′ of TaqMan probes were labeled with the fluorescent reporter FAM (6-carboxyfluorescein) and the quenching dye TAMRA (6-carboxytetramethylrhodamine), respectively.

contamination. The depth-integrated gene abundances were computed by trapezoidal integration over the sampling depth.

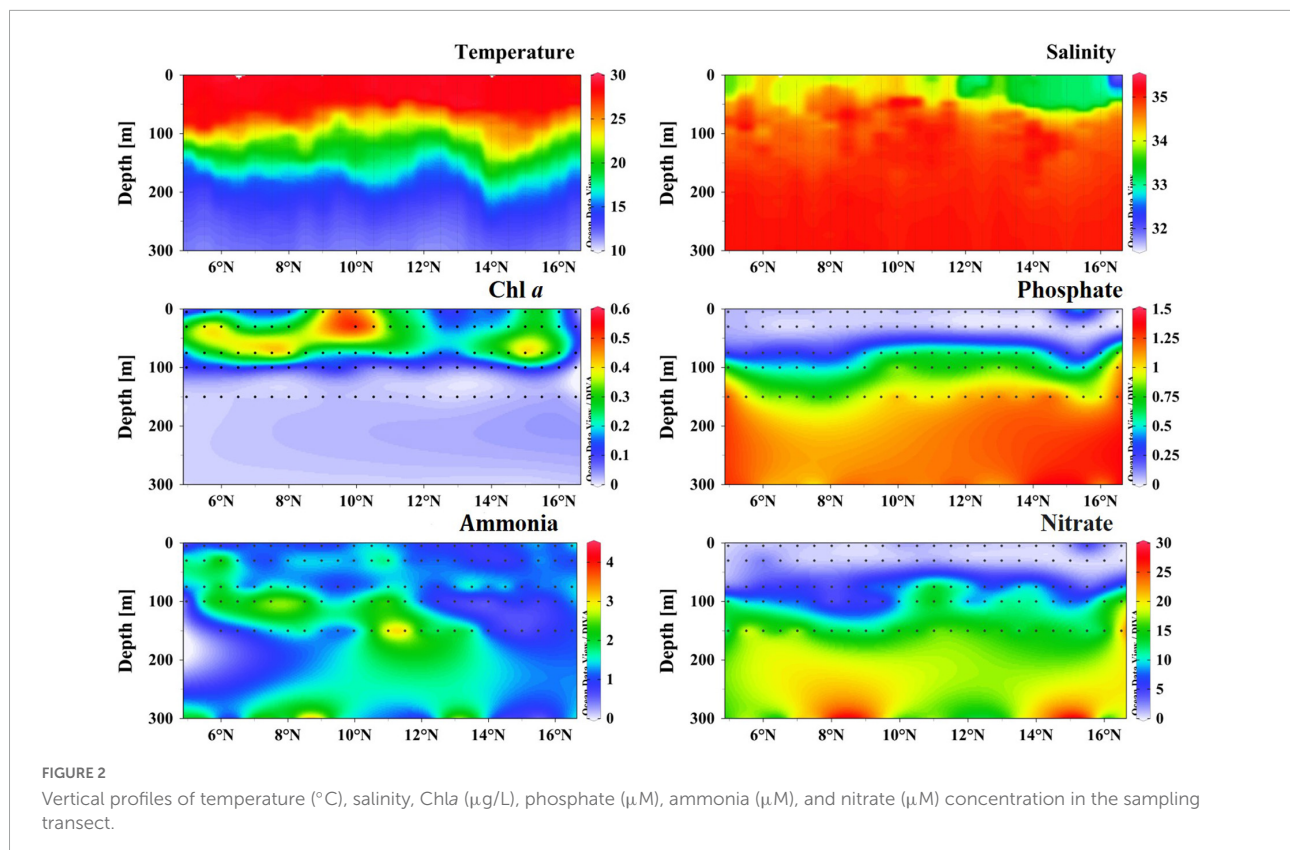# Results

Hydrography and environmental parameters

Hydrography and environmental parameters were supplied in **Figure 2** and **Supplementary Table 1**. During the study period, the sea surface temperature (SST) ranged from 27.7 to 29.3°C with an average of 28.7°C and showed an equatorward increasing trend (**Figure 2**). The thermocline was estimated between 75 and 125 m across this transect, whereas temperature decreased drastically under the thermocline. Overall, sea surface salinity (SSS) ranged from 31.754 to 34.312 (average 33.642) and revealed significant regional variation (**Figure 2**). Two distinct low SSS areas were observed at both ends of the transect (5 and 16°N), partly due to the freshwater runoff. The halocline was limited between 75 and 100 m, and the halocline depth increased toward the equator (**Figure 2**). The Chl*a* concentration in the surface water varied between 0.03 μg/L (at Sta. B01) and 0.53 μg/L (Sta. B15). Chl*a* maximum layer was estimated at approximately 50 m except for 10°N, where a typical high value was observed in the upper 50 m (**Figure 2**). Furthermore, Chl*a* concentration was almost undetectable less than 130 m. The surface nutrient concentrations, i.e., phosphate (0.01–0.574 μM), ammonia (0.629–1.571 μM), and nitrate (0.294–2.070 μM) were comparatively low, which indicated the BoB as a typical oligotrophic ecosystem (**Figure 2**).

Bacterioplankton community structure and diversity using 16S rRNA gene

In total, 429,576 high-quality sequences were obtained from the 16S rRNA gene sequencing, which clustered into 573 OTUs at 97% sequence similarity. According to the rarefaction curves, we observed that most samples were plateaued (**Supplementary Figure 1**). The alpha diversity of the 16S rRNA gene was shown in **Supplementary Table 2**. Good's coverage of all 16S

rRNA gene sequencing was greater than 99.8%, revealing full rarefaction saturation of all the samples and recovery of most bacterioplankton taxa in samples. The OTU numbers ranged from 299 to 353 in 0 m and 417–470 in 75 m samples. In general, the surface samples presented to low OTUs than the 75 m samples (one-way ANOVA, $P < 0.01$). Shannon-Weiner and Simpson diversity indices also revealed that the 75 m samples owned higher species abundance than the surface samples (one-way ANOVA, $P < 0.01$).

Non-metric multidimensional scaling (NMDS) analysis revealed an obvious separation of the 16S rRNA gene in the surface and deeper layer (**Figure 3A**). The predominated phylum in the study area was Proteobacteria (48.8% of the total 16S rRNA gene sequence), followed by Cyanobacteria (29.8%), Actinobacteria (10.1%), Bacteriodetes (3.9%), and Marinimicrobia (3.2%) (**Figure 3B**). These top five phyla accounted for more than 95.8% of the total bacterial communities across all the samples in the BoB. Other bacterial communities, such as Chloroflexi and Planctomycetes, were also detected but in very low relative abundance. Proteobacteria was the most diverse and abundant bacterial phylum across all the samples (53.6% of OTUs). Among which, the α-Proteobacteria (20.8%), γ-Proteobacteria (17.6%), and δ-Proteobacteria (13.6%) were the most diverse Proteobacteria in the BoB, accounting for 52% of the total number of OTUs. The relative abundance of Proteobacteria contributed about 48.8% to the total sequences and showed little difference between the two depths (22.2% on the surface and 26.6% in the 75 m layers). The α-proteobacteria was the most abundant Proteobacteria at each station (27% of the total sequences) and revealed a little difference in relative abundance among the two depths. However, γ-Proteobacteria and δ-proteobacteria were distinct from the α-proteobacteria, with more abundant in the deep layer. Unlike Proteobacteria, Cyanobacteria were more abundant in the surface water (22.7 and 7.1% of the total sequence in 0 and 75 m, respectively). Among the Cyanobacteria, *Prochlorococcus* always dominated the surface layers, whereas *Synechococcus* showed minor differences within the two depths. The dominant subgroup of the Actinobacteria, *Candidatus,* and *Actinomarina* was also detected in all samples and accounted for 5.4% of the total sequences. The rest subgroups of Actinobacteria have occurred in extremely low abundance. Flavobacteria represented the
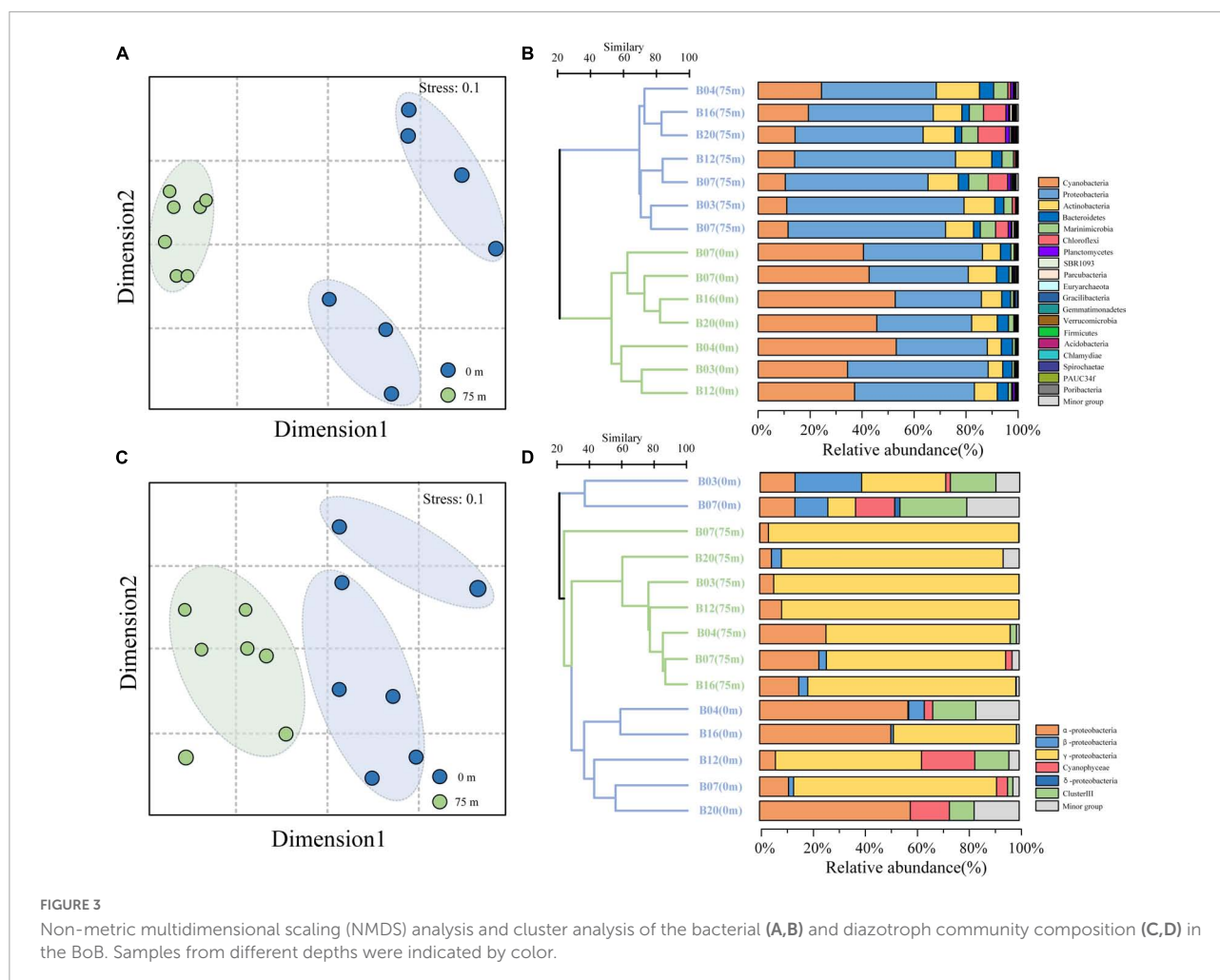
FIGURE 2
Vertical profiles of temperature (°C), salinity, Chl*a* (µg/L), phosphate (µM), ammonia (µM), and nitrate (µM) concentration in the sampling transect.

most fractions of Bacteriodetes, which accounted for 3.8% of the total sequence. Marinimicrobia was another dominant group in the bacterial communities, represented solely by the clade SAR406.

Diazotroph community structure and diversity using *nif*H gene

The rarefaction curves of the *nif*H gene were supplied in **Supplementary Figure 1**. We found that many OTUs in two samples (B07, 75 m and B20, 75 m) only contained 2 sequences from the OTU table. Thus, here it can be hypothesized that some errors might occur in the process of sequencing. To decrease the influence of these errors, OTUs that contained less than 10 sequences across all samples were removed from the OTU table. The alpha diversity of the *nif*H gene is given in **Supplementary Table 2**. In total, 901,603 high-quality sequences and 495 OTUs were included in our analysis. The sequencing coverages (C) were all greater than 99%, which suggested that the sequencing was deep enough for *nif*H gene diversity analysis in the present study. Unlike with 16S rRNA sequencing, the *nif*H gene OTUs in the surface samples were greater than the 75 m samples (one-way ANOVA, $P < 0.05$). Shannon-Weiner and Simpson diversity indices also showed the same vertical variation pattern (one-way ANOVA, $P < 0.05$).

As the 16S rRNA gene, the *nif*H gene also presented obvious separation in the surface and deeper layer (**Figure 3C**). For the taxonomic classification of the OTUs, a neighbor-joining phylogenetic tree was constructed (**Figure 4**). The OTUs containing sequences greater than 1,000 were defined as top OTUs. The top 59 OTUs, which accounted for more than 94.4% of the total sequences, were included in the present analysis. In addition, 37 representative sequences were also selected from GenBank for comparison. Here, **Figure 3D** together with **Figure 4** revealed the predominance of γ-*Proteobacteria* (63.4% of the total *nif*H gene sequence) in the *nif*H phylotypes in the BoB, followed by α-Proteobacteria (16.7%), δ-Proteobacteria (5.9%), Cyanophyceae (4.4%), and β-Proteobacteria (3.7%). γ-Proteobacteria was more abundant in the 75 m layer than the surface (one-way ANOVA, $P < 0.01$). On the contrary, *Cyanophyceae* and ClusterIII species were more abundant on the surface than the 75 m layer (one-way ANOVA, $P < 0.05$). However, no significant difference was observed in the occurrence of α-Proteobacteria within the two layers. From the phylogenetic tree (**Figure 4**), we observed that γ-Proteobacteria was the most diverse diazotroph communities (18 / 59 OTUs). OTU1 was the most abundant γ-Proteobacteria in the 14 samples. OTU1 showed 100% similarity to HM210363, a diazotroph reported from the south pacific gyre (Halm et al., 2012). OTU50 showed 100% identity with γ-24774A11 which was once reported as dominant in the AS during winter

FIGURE 3

Non-metric multidimensional scaling (NMDS) analysis and cluster analysis of the bacterial **(A,B)** and diazotroph community composition **(C,D)** in the BoB. Samples from different depths were indicated by color.

monsoon (BAN66892). However, in the present study, OTU50 was only detected in the surface waters at stations B12 and B20. In addition, many unidentified γ-Proteobacteria were also presented in high abundance. Here, α-Proteobacteria such as OTU4 and OTU16 were the most prevalent in the BoB. OTU4 had 100% identity with the isolation of *Methylocystis* strain m261 (ACH61854) from freshwater. OTU16 had 100% similarity with an uncultured α-Proteobacteria clone CE1-100m-2 (HQ586738) detected in the South China Sea, and 99% similarity with new isolation belonged to the genus of *Sagittula* from an OMZs in the eastern tropical South Pacific (CP021913). The dominance of *Sagittula* sp. was reported in the Eastern Indian Ocean (EIO) earlier during the spring inter-monsoon. In the BoB, however, OTU16 was not the dominant species. The neighbor-joining phylogenetic tree illustrated that 16 OTUs belonged to ClusterIII among the top 59 OTUs, including diverse δ-Proteobacteria. The δ-*Proteobacteria* was mainly observed in the surface waters and detected very rarely in the 75 m layers (Figure 4). Together with diverse *Proteobacteria*, we also identified 7 OTUs belonged

to *Cyanophyceae*, including OTU6 (*Trichodesmium* sp.), OTU7 (UCYN-A3), OTU29 (*Calothrix* sp.), OTU34 (*Crocosphaera* sp.), OTU32 (*Richelia* sp.), OTU256, and OTU27. As expected, *Cyanophyceae* was not prevalent in the BoB (4.4% of the total sequences) during the winter monsoon. Among the 7 OTUs, *Trichodesmium* sp. and UCYN-A3 were the most abundant *Cyanophyceae* in the study region, but the two *Cyanophyceae* were only detected in the surface layer.

Co-occurrence pattern of bacterial and diazotroph communities

To understand the organization of the bacterial and diazotroph communities, network analysis was conducted based on the high-throughput sequencing data (Figure 5). Detailed information on node-level topological properties is presented in Table 2. The bacterial community network was composed of 71 nodes and 289 edges with a mean of 8.141 edges per node (Table 2). The average path length (APL) was 2.816 edges with a diameter of 8 edges. The modularity index
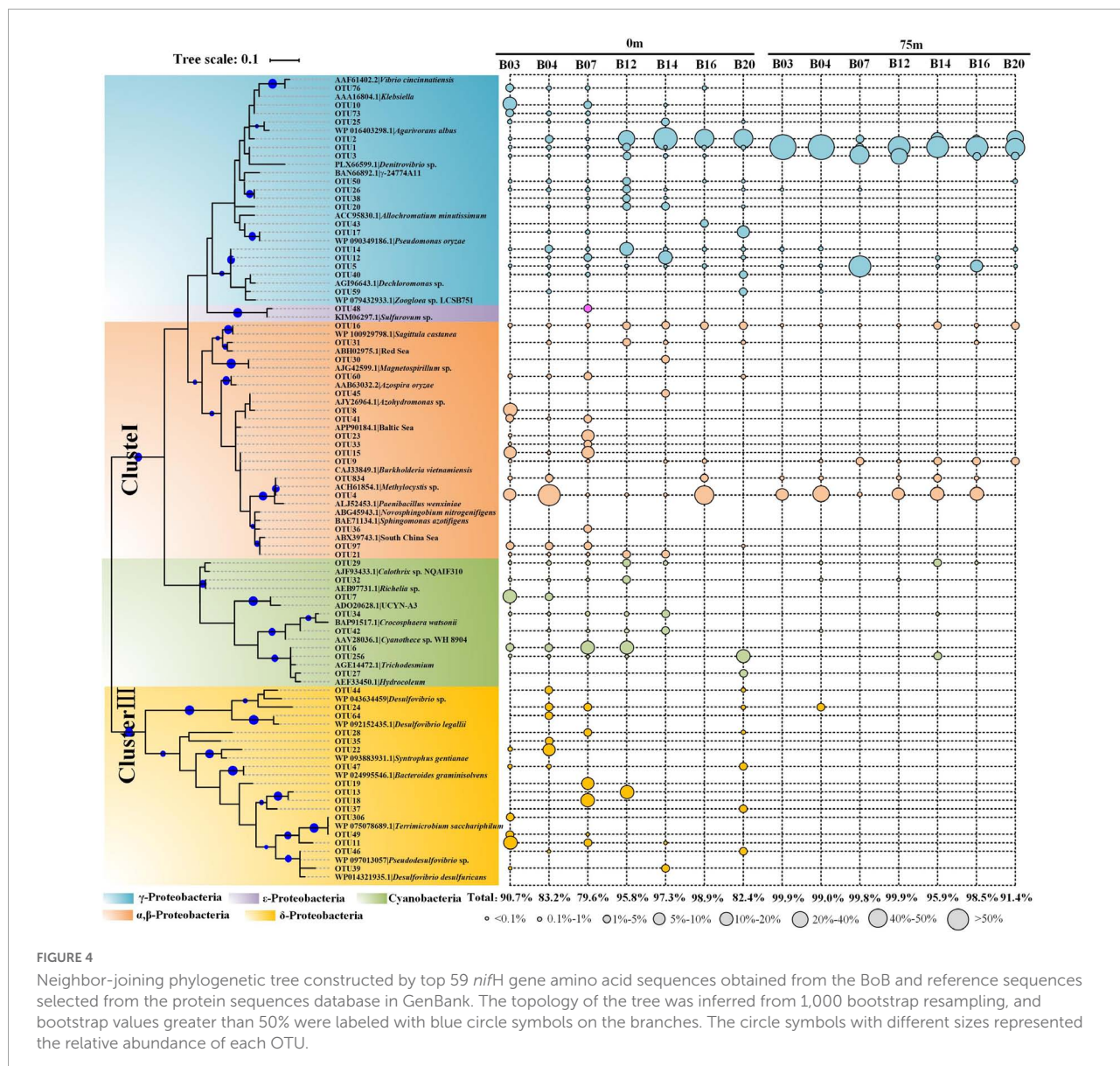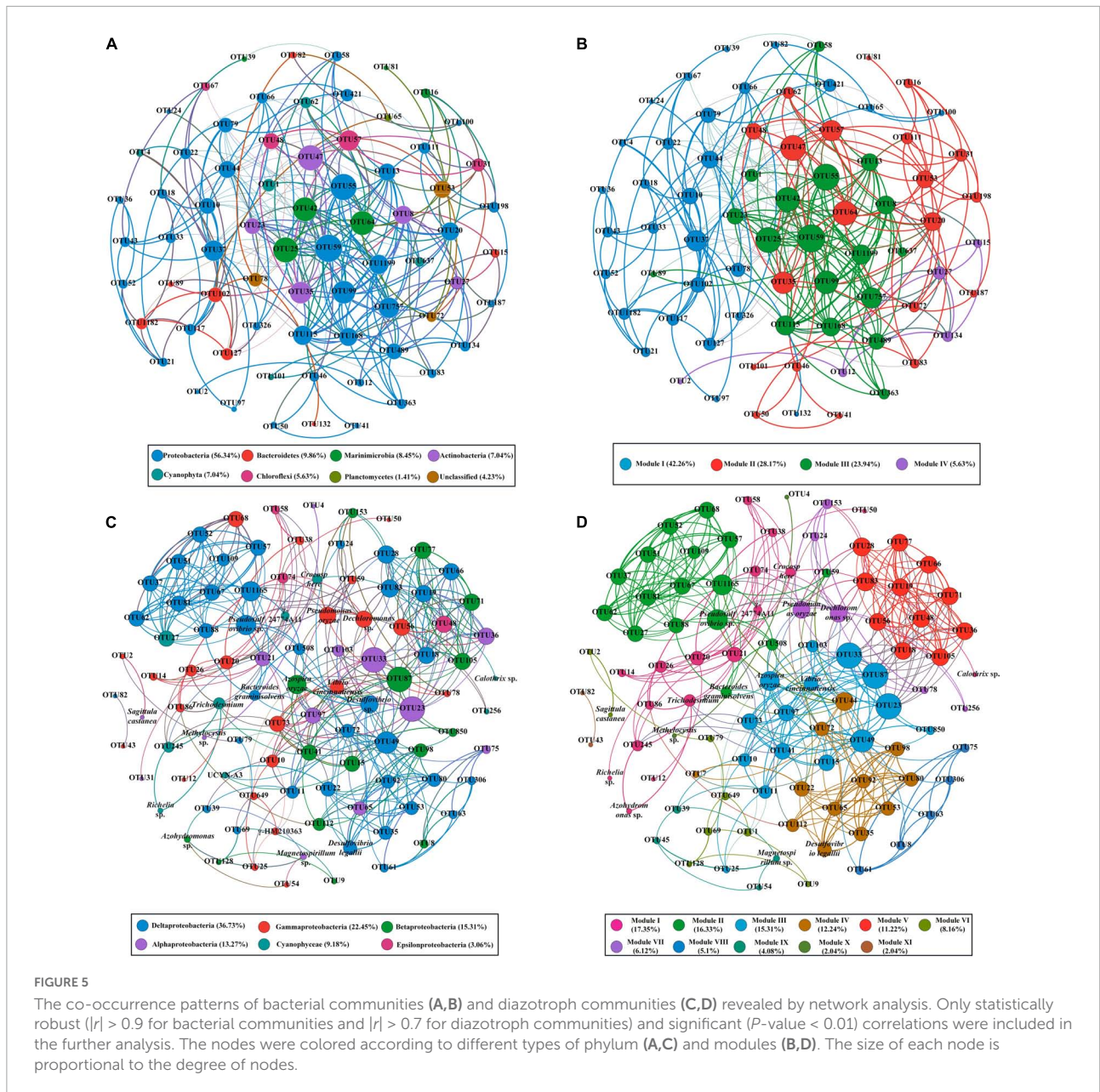
FIGURE 4

Neighbor-joining phylogenetic tree constructed by top 59 *nif*H gene amino acid sequences obtained from the BoB and reference sequences selected from the protein sequences database in GenBank. The topology of the tree was inferred from 1,000 bootstrap resampling, and bootstrap values greater than 50% were labeled with blue circle symbols on the branches. The circle symbols with different sizes represented the relative abundance of each OTU.

was greater than 0.4 which implied the network for bacterial has a modular structure. In brief, our results indicated that the bacterial communities were composed of highly connected genera and formed distinguishable modules. The network was composed of six modules, in which *Proteobacteria* was the most abundant Phylum (**Figure 5A**). The further modularized analysis revealed that the network was mainly composed of four modules (**Figure 5B**). The nodes from module I was mostly composed of Proteobacteria and Bacterodetes; nodes from module III mostly belonged to Proteobacteria and Marinimicrobia (**Figures 5A,B**). Based on the centrality scores, the top five OTUs were identified as keystone taxa including OTU47 (Actinobacteria), OTU25 (Marinimicrobia), OTU64 (Marinimicrobia), OTU78 (Unclassified), and OTU102 (Bacteroidetes). For diazotroph communities, the network

was composed of 98 nodes and 411 edges (**Table 2**). The average edges for each node were 10.299. Similar to the bacterial network, the modularity index was also greater than 0.4 which implies a good modular structure for diazotroph communities. The APL of the diazotroph network was 3.447 edges with a diameter of 9 edges. From the network, nodes were assigned to six classes (**Figure 5C**). Among these, δ-Proteobacteria accounted for about 36.73% of all nodes and presented high-level connectedness, indicating the vital ecological niche of these low abundance species. According to the value of betweenness centrality, six nodes were recognized as keystone species, including OTU49 (δ-Proteobacteria), OTU15 (β-Proteobacteria), OTU21 (α-Proteobacteria), OTU20 (γ-Proteobacteria), OTU74 (ε-Proteobacteria), and OTU60 (δ-Proteobacteria). All keystone species belonged to

FIGURE 5

The co-occurrence patterns of bacterial communities **(A,B)** and diazotroph communities **(C,D)** revealed by network analysis. Only statistically robust ($|r| > 0.9$ for bacterial communities and $|r| > 0.7$ for diazotroph communities) and significant (*P*-value < 0.01) correlations were included in the further analysis. The nodes were colored according to different types of phylum **(A,C)** and modules **(B,D)**. The size of each node is proportional to the degree of nodes.

Proteobacteria in the present study. After modularizing, these nodes were grouped into 11 major modules (**Figure 5D**). We found that most of the cyanobacteria were grouped into Module I, and most of the δ-*Proteobacteria* were grouped into Module II (**Figures 5C,D**).

Nitrogen cycles simulation in the Bay of Bengal using PICRUSt

The proteins related to nitrogen cycles were predicted and visualized in **Figure 6**. According to the COG orthologs, the denitrification and ammonium oxidation (anammox) in the

75 m layers were decreased from north to south (Sta. B03 to Sta. B20), which implies the BoB losing fixed nitrogen in deeper layers. However, the two nitrogen-removed pathways

TABLE 2 Topological properties of the co-occurring networks of bacterial and diazotroph communities.

	Nodes	Edges	MD	CC	APL	ND	AD	GD
Bacterial network	71	289	0.808	0.528	2.816	8	8.141	0.116
Diazotroph network	98	411	0.700	0.670	3.447	9	8.388	0.086

MD, modularity; CC, clustering coefficient; APL, average path length; ND, network diameter; AD, average degree; GD, graph density.
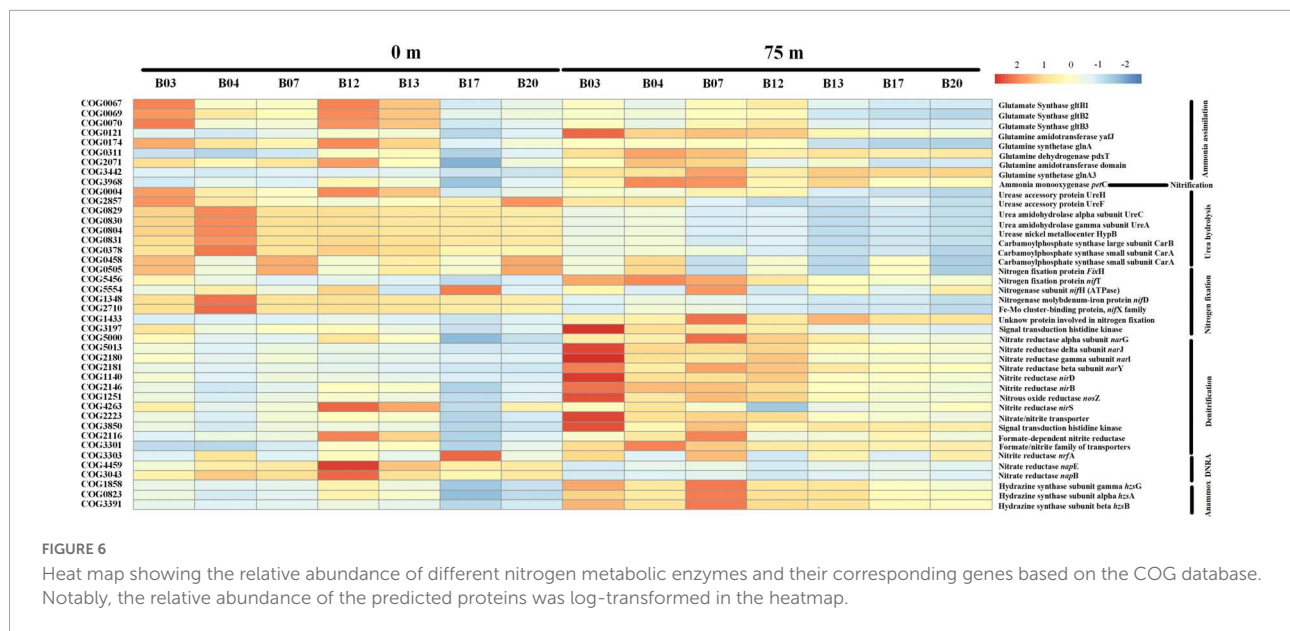
FIGURE 6
Heat map showing the relative abundance of different nitrogen metabolic enzymes and their corresponding genes based on the COG database. Notably, the relative abundance of the predicted proteins was log-transformed in the heatmap.

were not active in the surface water. Ammonia oxidation, the first step of nitrification, was also detected by PICRUSt. It was also presented as more active in the 75 m layers than that in the surface water. Dissimilatory nitrate reduction to ammonium (DNRA) and urea hydrolysis were also detected in this study and were only active in the surface waters. Unlike denitrification and anammox, nitrogen fixation was more active in the surface waters. However, nitrogen fixation was also detected in the deeper layer and replenished some of the nitrogen loss.

Quantification of cyanobacterial *nif*H phylotypes

Depth profiles of six main diazotrophic phylotypes are shown in Figure 7. *Trichodesmium* and *Richelia intracellularis* were only detected in the upper layer from 0 to 30 m, in very low abundance (Figures 7A,B). Three main types of *Trichodesmium* spp. were observed during the survey time: *Trichodesmium erythraeum*, *Trichodesmium thiebauti*, and *Trichodesmium hildebrandtii*. The average number of cells per trichome in all samples ranged from 70 to 150 cells/trichome. *Richelia intracellularis is* only detected at station B16 (<30 m) and only formed symbiont with *Rhizomatous rhizophyta* (Figure 7B). The other four types of diazotrophic phylotypes were measured by qPCR. UCYN-A and UCYN-B were generally concentrated in the upper water layers (<75 m). These two types of unicellular cyanobacteria were typically high at Sta. B16 and B20 (Figures 7C,D). γ-HM210363 was the most dominant diazotroph as observed in high-throughput sequencing data. Similarly, qPCR also demonstrated a high abundance of γ-HM210363 in the BoB during the winter

monsoon. The gene abundance of γ-HM210363 (copies/L) was two orders of magnitude higher than *Sagittula castanea*, which was recognized as the most abundant species during the pre-southwest monsoon.

Influence of environmental factors on 16S rRNA gene and *nif*H gene communities

To assess the influence of environmental factors on the bacterioplankton communities in the BoB, a multivariate analysis (CCA) was performed. In the bacterial community CCA biplot, the first two axes of CCA explained 66.07 and 19.61% of the total variation in the bacterial communities, respectively (Figure 8A). Significant tests of bacterial communities and the environmental factors indicted that temperature, salinity, and silicate contributed significantly ($P < 0.01$) to the total variance, and were closely associated with the first and second axes (999 times Monte–Carlo permutations). Due to the high collinearity with silicate, nitrate was removed from the CCA analysis. The bacterial communities in the CCA biplot also showed a distinct vertical distribution pattern (Figure 8A) similar to NMDS analysis (Figure 3A). The surface samples at Sta. B03, B04, and B12 were positive related to silicate, while in other stations they were more closely related to phosphorus. However, in 75 m samples, the bacterial community composition was mainly influenced by temperature. In addition, the CCA biplot showed that the abundance of most Cyanobacteria, α-Proteobacteria, and γ-Proteobacteria species were positively correlated with temperature, while negatively influenced by nutrients and salinity.
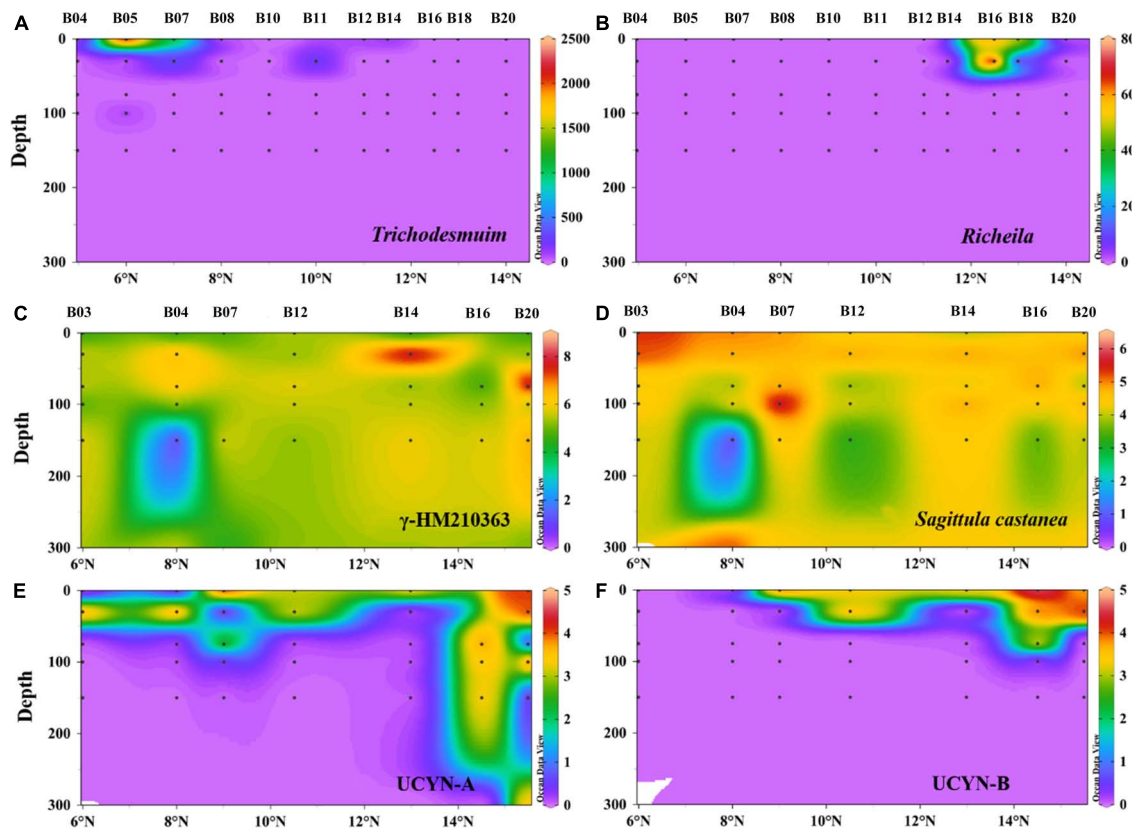
FIGURE 7
Depth profiles of filamentous cyanobacteria abundance (cell L^{-1}) for *Trichodesmium* **(A)** and *Richelia intracellularis* **(B)** and *nif*H gene abundance ($\log_{10}$ copies L^{-1}) for γ-HM210363 **(C)**, *Sagittula castanea* **(D)**, UCYN-A **(E)**, UCYN-B **(F)** in the BoB revealed by qPCR. For convenience, one gene copy used to represent where *nif*H genes were under detection.
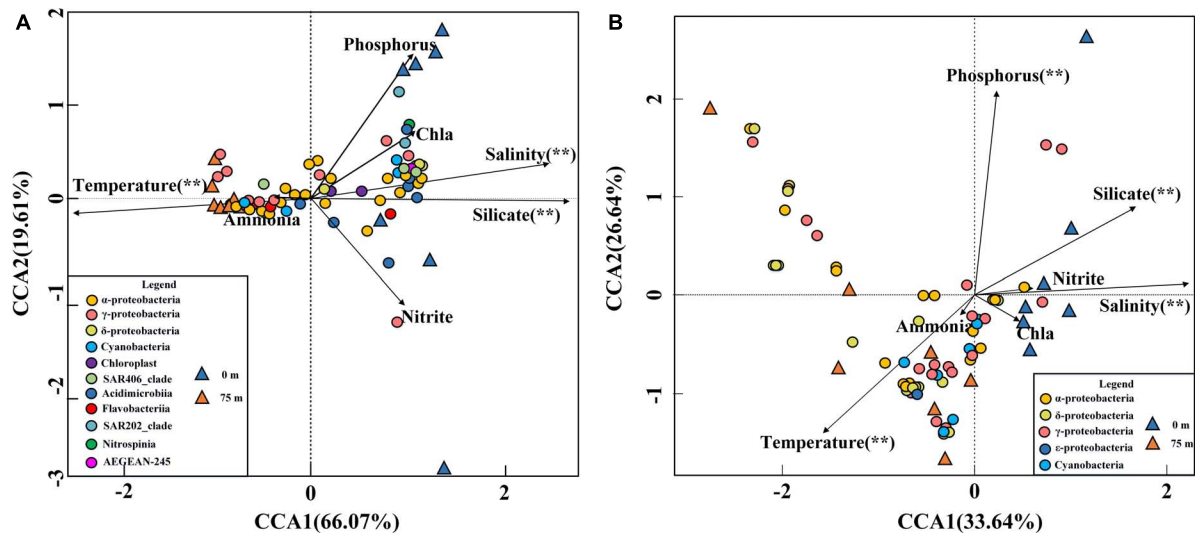

FIGURE 8
CCA for the relationship between the 16S rRNA gene **(A)** and *nif*H gene **(B)** distributions with the environmental parameters in the BoB. Black arrows represented environmental parameters and different colored triangles represented different layer samples. Significant (**) was determined by 999 Monte Carlo permutation tests in R V.3.0 software.

For diazotroph communities, the environmental factors in the first two axes explained 60.28% of the total variance. Significant tests revealed that temperature, phosphorus, salinity, and silicate contributed significantly ($P < 0.01$) to the total variance and were closely associated with the first two axes (999 times Monte–Carlo permutations). In the CCA biplot, surface samples were more closely related to nutrients and salinity, while the 75 m samples were always influenced by temperature (**Figure 8B**). The relationship between the environmental factors and samples in the *nif*H Gene biplot closely resembled the 16S rRNA gene CCA biplot. The CCA biplot showed that all the Cyanobacteria were positively correlated with temperature while negatively correlated with nutrients and salinity. In addition, we observed that the most abundant three γ-Proteobacteria (OTU1, OTU3, and OTU5) were positively influenced by nutrients and salinity and negatively correlated with temperature. However, the other less abundance γ-Proteobacteria species exhibited the opposite relationship with environmental factors. α-Proteobacteria and β-Proteobacteria also showed the same relationship with environmental factors and most of these species were positively correlated with temperature except for OTU9, OTU30, OTU39, and OTU45.

Discussion

Distribution of bacterial communities in the Bay of Bengal during the winter monsoon

Here, we assessed the bacterial communities along a transect across the central BoB with Illumina Hiseq sequencing targeting the V3–V4 region of the 16S rRNA gene. Overall, highly diversified groups of bacterial communities were detected in the study area. Five dominant phyla were reported through the 16S rRNA gene sequencing, among which Proteobacteria and Cyanobacteria were the most abundant phyla than Actinobacteria, Bacteroidetes, and Marinimicrobia. A similar bacterioplankton community structure was also reported earlier in many of the global oligotrophic oceans such as the South China Sea, the Northwestern Pacific Ocean, and the South Atlantic Ocean (Milici et al., 2016; Li et al., 2018; Fernandes et al., 2020). Our results were nevertheless different from recent studies conducted in the coastal areas of BoB (Rajpathak et al., 2018) and the equatorial EIO (Wang et al., 2016) where Cyanobacteria were rarely discovered. This could suggest the distinct spatio-temporal variations within the bacterioplankton communities in the EIO.

The 16S rRNA gene sequencing revealed the distinct distribution pattern within the dominant bacterioplankton groups, i.e., Cyanobacteria, Proteobacteria, Actinobacteria, Chloroflexi, and Marinimicrobia (**Supplementary**

Figure 2). The dominance of the pico-cyanobacteria genus, *Prochlorococcus*, was more abundant in the surface water than in the deeper layers (**Figure 3B**) and could be attributed to the light-limited conditions in the deeper layers (Rusch et al., 2010). It agreed well with previous reports that *Prochlorococcus* was more abundant at or near the surface, and decreased dramatically along the water column (Beatriz et al., 2016; Kent et al., 2016). The previous study conducted in the neighboring equatorial EIO reported that the mean cell diameter of *Prochlorococcus* was 0.60 ± 0.22 μm, and it further increased with depth (Wei et al., 2018). The smaller *Prochlorococcus* cell on the surface has a competitive advantage over larger phytoplankton for resource acquisition, utilization, and reproduction (Raven, 1998). However, in the deeper layers, their cell diameters are gradually increased to acquire more pigments for growth to compensate for the decreased light levels (Goericke and Welschmeyer, 1998). The *Prochlorococcus* population was less diverse (4 OTUs) in the present study, compared to previous reports where specific primers were used (Kent et al., 2016). *Prochlorococcus* group members shared >97% similarity of their 16S rRNA genes, and this group was always designated as a single species (Babić et al., 2017). Therefore, further investigations targeting the functional gene of *Prochlorococcus* are greatly needed to better understand the real composition of this pico-cyanobacteria in the BoB. *Synechococcus* was another well-recovered Cyanobacteria in this study, though presented in lower relative abundance than *Prochlorococcus*. This group has a wider geographical distribution even in colder, eutrophic waters and high-latitude oceans where *Prochlorococcus* occurs in low abundance or absent (Xia et al., 2019). A recent metagenomic study conducted in the Indian Ocean confirmed the presence of nitrate utilization genes (*nar*B) in *Synechococcus*, which might explain their tolerance to nutrient-rich waters (Beatriz et al., 2016). Like *Prochlorococcus*, 16S rRNA gene sequencing also underestimated the diversity of *Synechococcus* groups (2 OTUs). Two distinct types of *Synechococcus* were observed in the BoB with OTU 4 abundant in the surface waters and OTU 1038 dominant in the 75 m layer. Here possibly OTU 4 group contain higher phycourobilin (bright type) which is always dominant in oligotrophic waters, whereas OTU 1038 was a lower phycourobilin (dim type) containing group which is always found in eutrophic waters (Mazard et al., 2011; Xia et al., 2016). The representative sequence of OTU 338 shared 100% similar with *Richelia* sp. Consistent with *nif*H gene sequencing, the 16S rRNA sequences also revealed a low relative abundance of *Richelia* sp. at all stations. Other nitrogen-fixing cyanobacteria were not detected in the 16S rRNA gene libraries though they might be sequenced.

Compared to Cyanobacteria, Proteobacteria represented the highest diversity and largest fractions in the BoB during the winter monsoon. Among which, α- and γ-Proteobacteria were the most abundant in both two layers, and δ-Proteobacteria

was also commonly detected in low abundance. Despite the difference in other dominant species, the dominance of α- and γ-Proteobacteria within the Proteobacteria group report in the BoB agreed with an earlier study from the equatorial EIO (Wang et al., 2016). In α-Proteobacteria, members of the *SAR11* clade and *Rhodobacterales* spp. were most abundant in this study. Here, we have identified four clades of *SAR11* in the BoB according to the previous definition (**Supplementary Figure 3**). Earlier studies had also reported the α-Proteobacteria clade *SAR11* as the most abundant and diverse group in the global ocean ecosystem (Morris et al., 2002; Rappe et al., 2002). The dominance of this clade in oxygen-rich surface waters as well as in OMZs suggests its adaptive capabilities to confront oxygen decrease (Hawley et al., 2014). The growth of *SAR11* clade also requires exogenous sources of reduced sulfur (e.g., 3-dimethylsulphoniopropionate), suggesting that this group of organisms may be active in sulfur cycles (Morris et al., 2002; Grote et al., 2012). In addition, the efficient genome streamlining of the *SAR11* clade can minimize the cellular nutrient requirements, thus supporting fast growth and replication under extreme oligotrophic conditions (West et al., 2016). The dominant other α-Proteobacteria, *Rhodobacterales* spp., can be correlated with fundamental metabolic processes such as sulfur oxidation, nitrogen fixation, and autotrophic carbon fixation (Lee et al., 2013). Strikingly, the nitrogen fixer *Sagittula castanea* quantified by qPCR in this study belonged to *Rhodobacterales* spp., whereas the phylogenetic analysis revealed that *Sagittula*-related groups were more correlated with the AEGEAN-169 marine group. The relative abundance of another dominant Proteobacteria, γ-Proteobacteria, decreased with depth, which was consistent with the previous study conducted in the North Atlantic Ocean (HéLèNE et al., 2011). Sequence alignment revealed that *Oceanospirillales* and *SAR86* clade were the dominant groups at the order level within γ-Proteobacteria. Some *Oceanospirillales* members were reported to contain genes for carbon fixation (RuBisCO genes) and sulfur oxidation (Swan et al., 2011). In addition, many studies revealed that *Oceanospirillales* members always formed symbiotic with various marine invertebrates (Cao et al., 2014). *SAR86* clade was more abundant in the surface waters. The metagenomic analysis demonstrated that the *SAR86* clade lacks several pathways, such as amino acid and vitamin synthesis as well as sulfate reduction which was commonly detected in other abundant marine microorganisms (Dupont et al., 2012). The genome streamlining strategy supported this group to have a selective advantage over other oligotrophic groups. In δ-Proteobacteria, members of the order *SAR324* clade were abundant in the study area (**Supplementary Figure 3**). Earlier this clade was reported as a deeper water clade (HéLèNE et al., 2011); however, it has been also detected in the surface waters in our study. This suggests their survival capability in diverse environmental conditions. In brief, adaptive mechanisms such

as small cell size and genome streamlining could support the dominance of Proteobacteria in oligotrophic waters of the BoB.

Co-occurrence network analysis revealed that all keystone species belonged to rare species (<0.1% of the total sequences) as also reported in other aquatic ecosystems (Milici et al., 2016; Jia et al., 2018). These "rare biospheres" played a vital role in the process of community assemblage and contributed significantly to the biogeochemical process albeit they only account for a minor fraction of the whole communities (Jiao et al., 2016). In the present study, temperature and salinity were major factors to determine the distribution of bacterial communities. This agreed well with previous studies in many other oceans such as the South China Sea and the Bohai Sea (Zhang et al., 2011; Wu et al., 2022a). Notably, silicate was also the major factor to influence the distribution of bacterial communities in our study. This is because pico-cyanobacteria were the major component of bacterial communities in the BoB. In recent years, pico-cyanobacteria were found to have significant silica accumulation (Wei et al., 2018). Furthermore, nitrate was deleted from the CCA analysis due to the high collinearity with other environments. However, it does not mean that nitrate was not important in structuring the spatial heterogeneity of bacterial communities. It is generally believed that these inorganic nutrient concentrations have important roles in regulating the abundance of various bacterial communities (Wu et al., 2019). According to the PICRUSt simulations, the BoB was losing fixed nitrogen in the deeper layer *via* denitrification and anammox (**Figure 6**). A decreasing trend of denitrification and anammox was observed from north to south in 75 m layers of the BoB. This suggested that the deeper layer of the BoB was losing a great deal of fixed nitrogen. Thus, we can speculate that the part of the present sampling stations (Sta. B03-B13) was occupied in the OMZs of the BoB. Our results coincided with previous reports by the isotope labeling method which also showed nitrogen loss was significant in OMZs of the BoB (Bristow et al., 2016). However, denitrification and anammox were all suppressed in the surface waters, and the net results were that nitrogen fixation supplied fixed nitrogen in the upper layers. Our previous study using an isotope ^{15}N tracer technique also showed that nitrogen fixation rates were high in surface waters in the BoB and the southeastern Indian Ocean (Wu et al., 2022b). Overall, our study demonstrated that nitrogen fixation is significantly important in adding new nitrogen to the upper layers of this oligotrophic marine ecosystem.

Distribution of diazotroph communities in the Bay of Bengal during the winter monsoon

In the present study, lower diversity of the diazotrophic community was observed in the BoB during the winter monsoon. The diazotrophic communities were present with

higher diversity in the surface than deeper layers in contrast to the 16S rRNA gene sequencing. The high diversity of diazotrophic communities on the surface revealed that most of the diazotrophs were lucipetal and more active in the upper ocean. In the phylogenetic analysis, the top 59 OTUs were grouped into two defined clusters of *nif*H genes (Cluster I and Cluster III). Further, these communities were strikingly similar to a recent study conducted in the OMZs of the BoB, so these diazotrophs could be the characteristic of OMZs (Löscher et al., 2020).

Here, the results generated from high-throughput sequencing and microscopic enumeration indicated that nitrogen-fixing Cyanobacteria were not abundant in the study area during the winter monsoon. This was quite different from our earlier study conducted during the spring inter-monsoon, where *Trichodesmium* spp. were dominant in the surface waters of the BoB (Wu et al., 2019, 2022b). Wind-drift current was thought to be the main factor that influences the seasonal rhythm of *Trichodesmium* spp. in the BoB (Hegde et al., 2008; Jyothibabu et al., 2017). *Trichodesmium* colonies can be easily destroyed by strong winds (Capone et al., 1997) and, therefore, the high turbulence during the winter monsoon is not in favor of the growth of *Trichodesmium*. The main types of *Trichodesmium* spp. in the present study were *Trichodesmium thiebauti* and *Trichodesmium erythraeum*. These two species were also reported to appear in both Indian coastal waters and open oceans (Krishnan et al., 2007). Unfortunately, *Trichodesmium* spp. (OTU347) was only retrieved from the *nif*H gene sequencing, while they were not recognized in the 16S rRNA gene sequencing. Unlike *Trichodesmium* spp., *Richelia* spp. was recognized in both gene sequencing though all presented in low abundance. Our results showed that *Richelia* spp. only formed symbionts with *Rhizomatous rhizophyta*, which was different from previous studies conducted in Indian oceans (Kulkarni et al., 2010; Lyimo, 2011).

Except for filamentous Cyanobacteria, unicellular diazotrophs were also detected in both high-throughput sequencing and qPCR in this study. However, the unicellular diazotrophs, including UCYN-A, *Crocosphaera watsonii*, and *Cyanothece* sp. WH 8904, were presented to very low abundance, consistent with our previous study (Wu et al., 2019). The low abundance of unicellular diazotrophs could be attributed to the shallower nitracline depths and continuous high temperature as described earlier (Shiozaki et al., 2014; Wu et al., 2019). Overall, the Cyanobacterial population was not abundant in the central BoB during the winter monsoon. From the neighbor-joining phylogenetic tree, OTU7 shared 100% similarity with UCYN-A3 (ADO20628.1) and was the only species that belonged to UCYN-A among the top 59 OTUs. UCYN-A3 is a newly characterized open ocean sublineage that formed symbiotic with uncultivated unicellular alga (Cornejo-Castillo et al., 2019). Turk-Kubo et al. (2016) reported that UCYN-A3 is always commonly detected in oligotrophic waters and co-occurs with UCYN-A1. However,

in this study, UCYN-A1 was not identified in the top 59 OTUs and could be a minor group to detect in this region. Here, UCYN-A was also quantified by qPCR with Taqman assays, and the gene abundance was the pool of UCYN-A2, UCYN-A3, and UCYN-A4 as stated elsewhere (Henke et al., 2018). Similar to high-throughput sequencing, qPCR also showed UCYN-A sublineages presented in low abundance toward the surface layers.

Unlike the Cyanobacteria, Proteobacteria represented the highest fractions of diazotrophs in the BoB during the winter monsoon. Here, the diazotroph communities mainly consisted of heterotrophic γ-Proteobacteria, which was significantly different from our previous pre-southwest monsoon study, where α-Proteobacteria dominated at the class level (Wu et al., 2019). The sequencing and qPCR analysis demonstrated that OTU1 (γ-HM210363) was the most abundant species and omnipresent in the BoB as compared with other proteobacterial subgroups. γ-HM210363 was once reported as relatively ubiquitous in the South Pacific Gyre, whereas not necessarily in the gene numbers (Halm et al., 2012). The gene abundance of *Sagittula castanea* was one or two orders of magnitude lower than that during the pre-southwest monsoon (Wu et al., 2019), which indicated high turbulence suppressing the growth of *Sagittula castanea*. Proteobacteria also exhibited significant temporal differences, where Cluster III was commonly detected in this study. Cluster III groups consisted mostly of sulfate-reducing bacteria and were always dominant in oxygen-depleted waters (Cheung et al., 2015). The subsurface waters in the BoB contain extremely low, but persistent concentrations of oxygen in the nanomolar range (Bristow et al., 2016). Thus, the low oxygen concentrations favor the growth of Cluster III species. However, here strangely Cluster III species were most detected in the surface waters, while it was sporadically detected in the 75 m layers. This discrepancy could be caused due to PCR bias because γ-Proteobacteria accounted for the majority of the diazotroph communities in the 75 m layers.

To date, none of the studies are focused on the assembly process of marine diazotroph communities in the oceans. Our study demonstrated that diazotroph communities were highly modular, and Proteobacteria played a vital role in the process of community assembling (**Figures 5C,D**). Strikingly, all recognized keystone species were rare species (<0.1% of the total sequences) and identical to the bacterial network. In this respect, the "rare biosphere" mediated multiple metabolic capabilities and ecosystem functions including nitrogen cycles (Jia et al., 2018). The CCA biplot of diazotroph communities explained less total variance (60.28%) in the present study than that during the pre-southwest monsoon (83.6%) (Wu et al., 2019). Wind stress was presumably to be a neglected parameter that could explain a considerable variance in the present study. The rest significant environmental parameters (temperature, salinity, phosphorus, and silicate) in the CCA biplot were generally the same as our study conducted during pre-southwest monsoon after excluding highly collinearity

parameters (Wu et al., 2019). This might be caused by different diazotroph communities that are suitable for living in different gradients of temperature and salinity. For example, UCYN-A tended to live in high salinity but low-temperature seawater, whereas *Trichodesmium* spp. tended to thrive when the temperature is 25°C or warmer (Wu et al., 2018; Li et al., 2021; Christian et al., 2022). Phosphorus is also an important factor to determine the distribution of diazotroph communities in our study. This might be the cause that phosphorus is the major component of cells and is thought to be a key factor in limiting diazotroph communities and their nitrogen fixation (Wu et al., 2019). Notably, we observed that the influence of silicate on diazotroph communities is also significant in our study. This phenomenon was also detected by a recent study conducted in the Eastern Indian Ocean and might be attributed to some heterotrophic diazotrophs entering a mutualistic symbiosis with diatoms (Wu et al., 2022b).

Conclusion

Here, our research focused on how the diazotroph communities play a significant role in the bacterioplankton communities, and their genetic potential to carry out nitrogen fixation. Generally, this study provided a comprehensive and intensive investigation of bacterial and diazotroph communities in the open oceans of the BoB during the winter monsoon. Although a recent study has analyzed the diazotroph communities based on clone library in the same region, our study provided a deeper sequencing of diazotroph communities as well as the bacterial communities based on high-throughput sequencing. The sequencing analysis presented that 75 m layers had higher diversity than surface samples of the bacterial communities, while diazotroph communities were presented with a reverse trend. Proteobacteria were the most abundant groups at the phylum level in both bacterial and diazotroph communities. Combined with our previous study, we found that both bacterial and diazotroph communities exhibited distinct spatio-temporal variations in the BoB. Co-occurrence networks demonstrated that rare biosphere might play a significant role in the assemble process of both bacterial and diazotroph communities. Limited by sequencing depth, our study lost some extremely rare OTUs and that is why we did not conduct a more in-depth analysis of abundant and rare communities. Further studies based on deeper sequencing are greatly in need to better understand the coupling and response between abundant and rare populations.

Data availability statement

The 16S rRNA gene and *nif*H gene raw sequences obtained from this study were deposited in NCBI Sequence Read Archive with BioProject nos. PRJNA637963 and PRJNA637983, respectively.

Author contributions

CW: investigation, methodology, software, validation, formal analysis, data curation, visualization, writing – original draft, and writing – review and editing. DN: software, writing – review and editing, and supervision. ZC: writing – review and editing and supervision. XW, WX, and GZ: investigation, writing – review and editing, and supervision. JS: investigation, writing – review and editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.987462/full#supplementary-material

References

Achary, M. S., Panigrahi, S., Satpathy, K. K., Sahu, G., Mohanty, A. K., Selvanayagam, M., et al. (2014). Nutrient dynamics and seasonal variation of phytoplankton assemblages in the coastal waters of southwest Bay of Bengal. *Environ. Monit. Assess.* 186, 5681–5695. doi: 10.1007/s10661-014-3812-8

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Babić, I., Petrić, I., Bosak, S., Mihanović, H., Radić, I. D., and Ljubešić, Z. (2017). Distribution and diversity of marine picocyanobacteria community: Targeting of *Prochlorococcus* ecotypes in winter conditions (southern Adriatic Sea). *Mar. Genomics* 36, 3–11. doi: 10.1016/j.margen.2017.05.014

Beatriz, D., Nylander, J. A. A., Ininbergs, K., Dupont, C. L., Allen, A. E., Yooseph, S., et al. (2016). Metagenomic analysis of the Indian Ocean picocyanobacterial community: Structure, potential function and evolution. *PLoS One* 11:e0155757. doi: 10.1371/journal.pone.0155757

Bhushan, R., Bikkina, S., Chatterjee, J., Singh, S. P., Goswami, V., Thomas, L. C., et al. (2018). Evidence for enhanced chlorophyll-a levels in the Bay of Bengal during early north-east monsoon. *J. Oceanogr.* 9, 15–23. doi: 10.5897/JOMS2017.0144

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bristow, L. A., Callbeck, C. M., Larsen, M., Altabet, M. A., Dekaezemacker, J., Forth, M., et al. (2016). N2 production rates limited by nitrite availability in the Bay of Bengal oxygen minimum zone. *Nat. Geosci.* 10, 24–29. doi: 10.1038/ngeo2847

Cao, Y., Chastain, R. A., Eloe, E. A., Nogi, Y., Kato, C., Bartlett, D. H. et al. (2014). Novel psychropiezophilic *Oceanospirillales* species *Profundimonas piezophila* gen. nov., sp nov., isolated from the deep-dea environment of the Puerto Rico Trench. *Appl. Environ. Microbiol.* 80, 54–60. doi: 10.1128/AEM.02288-13

Capone, D. G., Zehr, J. P., Paerl, H. W., Bergman, B., and Carpenter, E. J. (1997). *Trichodesmium*, a globally significant marine *Cyanobacterium*. *Science* 276, 1221–1229. doi: 10.1126/science.276.5316.1221

Chen, G. X., Li, Y. L., Xie, Q., and Wang, D. X. (2018). Origins of eddy kinetic energy in the Bay of Bengal. *J. Geophys. Res. Oceans* 123, 2097–2115. doi: 10.1002/2017JC013455

Cheung, S. Y., Xia, X. M., Guo, C., and Liu, H. B. (2015). Diazotroph community structure in the deep oxygen minimum zone of the Costa Rica Dome. *J. Plankton Res.* 38, 380–391. doi: 10.1093/plankt/fbw003

Christian, F. R., Ina, S., Jamileh, J., and Carolin, R. L. (2022). Salinity as a key control on the diazotrophic community composition in the southern Baltic Sea. *Ocean Sci.* 18, 401–417. doi: 10.5194/os-18-401-2022

Clarke, K. R., and Gorley, R. N. (2006). *PRIMER v6: User manual/tutorial.* Plymouth: Marine Laboratory.

Cornejo-Castillo, F. M., Muñoz-Marín, M. D. C., Turk-Kubo, K. A., Royo-Lionch, M., Farnelid, H., Acinas, S. G., et al. (2019). UCYN-A3, a newly characterized open ocean sublineage of the symbiotic N2-fixing *Cyanobacterium Candidatus Atelocyanobacterium thalassa*. *Environ. Microbiol.* 21, 111–124. doi: 10.1111/1462-2920.14429

Dupont, C. L., Rusch, D. B., Yooseph, S., Lombaerdo, M. J., Richter, R. A., Valas, R., et al. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* 6, 1186–1199. doi: 10.1038/ismej.2011.189

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Fernandes, G. L., Shenoy, B. D., and Damare, S. R. (2020). Diversity of bacterial community in the oxygen minimum zones of Arabian Sea and Bay of Bengal as deduced by illumina sequencing. *Front. Microbiol.* 10:3153. doi: 10.3389/fmicb.2019.03153

Fernandes, G. L., Shenoy, B. D., Menezes, L. D., Meena, R. M., and Damare, S. R. (2019). Prokaryotic diversity in oxygen depleted waters of the Bay of Bengal inferred using culture-dependent and independent methods. *Indian J. Microbiol.* 59, 193–199. doi: 10.1007/s12088-019-00786-1

Gauns, M., Madhupratap, M., Ramaiah, N., Jyothibabu, R., Fernandes, V., Paul, J. T., et al. (2005). Comparative accounts of biological productivity characteristics and estimates of carbon fluxes in the Arabian Sea and the Bay of Bengal. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 52, 2003–2017. doi: 10.1016/j.dsr2.2005.05.009

Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Herson, L., Morrison, A. E., et al. (2010). Abundance and distribution of major groups of diazotrophic *Cyanobacteria* and their potential contribution to N2 fixation in the tropical Atlantic Ocean. *Environ. Microbiol.* 12, 3272–3289. doi: 10.1111/j.1462-2920.2010.02303.x

Goericke, R., and Welschmeyer, N. A. (1998). Response of Sargasso Sea phytoplankton biomass, growth rates and primary production to seasonally varying physical forcing. *J. Plankton Res.* 20, 2223–2249. doi: 10.1093/plankt/20.12.2223

Gradoville, M. R., Bombar, D., Crump, B. C., Letelier, R. M., Zehr, J. P., and White, A. E. (2017). Diversity and activity of nitrogen-fixing communities across ocean basins. *Limnol. Oceanogr.* 62, 1895–1909. doi: 10.1002/lno.10542

Grasshoff, K., Ehrhardt, M., and Kremling, K. (1982). *Methods of seawater analysis*, 2nd Edn. New York, NY: Verlag Chemie Weinhein.

Grote, J., Thrash, J. C., Huggett, M. J., Landry, Z. C., Carini, P., Giovannoni, S. J., et al. (2012). Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* 3, e00252–12. doi: 10.1128/mBio.00252-12

Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., et al. (2012). Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *ISME J.* 6, 1238–1249. doi: 10.1038/ismej.2011.182

Hawley, A. K., Brewer, H. M., Norbeck, A. D., Paša-Tolic, L., and Hallam, S. J. (2014). Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11395–11400. doi: 10.1073/pnas.1322132111

Hegde, S., Anil, A. C., Patil, J. S., Mitbavkar, S., Krishnamurthy, V., and Gopalakrishna, V. V. (2008). Influence of environmental settings on the prevalence of *Trichodesmium* spp. in the Bay of Bengal. *Mar. Ecol. Prog. Ser.* 356, 93–101. doi: 10.3354/meps07259

HéLèNE, A., Lamy, D., Neal, P. R., Sogin, M. L., and Herndl, G. L. (2011). Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol. Ecol.* 20, 258–274. doi: 10.1111/j.1365-294X.2010.04932.x

Henke, B. A., Turk-Kubo, K. A., Sophie, B., and Zehr, J. P. (2018). Distributions and abundances of sublineages of the N2-fixing *Cyanobacterium Candidatus Atelocyanobacterium thalassa* (UCYN-A) in the New Caledonian Coral Lagoon. *Front. Microbiol.* 9:554. doi: 10.3389/fmicb.2018.00554

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., and Welch, D. M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 8:R143. doi: 10.1186/gb-2007-8-7-r143

Jia, X., Dini-Andreote, F., and Salles, J. F. (2018). Community assembly processes of the microbial rare biosphere. *Trends Microbiol.* 26, 738–747. doi: 10.1016/j.tim.2018.02.011

Jiao, S., Liu, Z. S., Lin, Y. B., Yang, J., Chen, W. M., and Wei, G. H. (2016). Bacterial communities in oil contaminated soils: Biogeography and co-occurrence patterns. *Soil Biol. Biochem.* 98, 64–73. doi: 10.1016/j.soilbio.2016.04.005

Jinadasa, S. U. P., Pathirana, G., Ranasinghe, P. N., Centurioni, L., and Hormann, V. (2020). Monsoonal impact on circulation pathways in the Indian Ocean. *Acta Oceanol. Sin.* 39, 103-112. doi: 10.1007/s13131-020-1557-5

Jyothibabu, R., Karnan, C., Jagadeesan, L., Arunpandi, N., Pandiarajan, R. S., Muraleedharan, K. R., et al. (2017). *Trichodesmium* blooms and warm-core ocean surface features in the Arabian Sea and the Bay of Bengal. *Mar. Pollut. Bull.* 121, 201–215. doi: 10.1016/j.marpolbul.2017.06.002

Kent, A. G., Dupont, C. L., Yooseph, S., and Martiny, A. C. (2016). Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J.* 10, 1856–1865. doi: 10.1038/ismej.2015.265

Kong, L. L., Jing, H. M., Kataoka, T., Sun, J., and Liu, H. B. (2011). Phylogenetic diversity, and spatio-temporal distribution of nitrogenase genes (*nif*H) in the northern South China Sea. *Aquat. Microb. Ecol.* 65, 15–27. doi: 10.3354/ame01531

Krishnan, A. A., Krishnakumar, P. K., and Rajagopalan, M. (2007). *Trichodesmium erythraeum* (Ehrenberg) bloom along the southwest coast of India (Arabian Sea) and its impact on trace metal concentrations in seawater. *Estuar. Coast. Shelf Sci.* 71, 641–646. doi: 10.1016/j.ecss.2006.09.012

Kulkarni, V. V., Chitari, R. R., Narale, D. D., Patil, J. S., and Anil, A. C. (2010). Occurrence of cyanobacteria-diatom symbiosis in the Bay of Bengal: Implications in biogeochemistry. *Curr. Sci.* 99, 736–737.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kumari, A., Prasanna, K. S., and Chakraborty, A. (2018). Seasonal and interannual variability in the barrier layer of the Bay of Bengal. *J. Geophys. Res. Oceans* 123, 1001–1015. doi: 10.1002/2017JC013213

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12, 118–123. doi: 10.1111/j.1462-2920. 2009.02051.x

Lee, D. H., Cho, S. J., Kim, S. M., and Lee, S. B. (2013). *Sagittula marina* sp. nov. isolated from seawater and emended description of the genus *Sagittula*. *Int. J. Syst. Evol.* 63(Pt 6), 2101–2107. doi: 10.1099/ijs.0.040766-0

Letunic, I., and Bork, P. (2011). Interactive tree of life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39, 475–478. doi: 10.1093/nar/gkr201

Levitan, O., Rosenberg, G., Setlik, I., Setlikova, E., Grigel, J., Klepetar, J., et al. (2007). Elevated CO2 enhances nitrogen fixation and growth in the marine *Cyanobacterium Trichodesmium*. *Glob. Change Biol.* 13, 531–538. doi: 10.1111/j. 1365-2486.2006.01314.x

Li, G., Ke, Z. X., Lin, Q., Ni, G. Y., Shen, P. P., Liu, H. X., et al. (2012). Longitudinal patterns of spring-intermonsoon phytoplankton biomass, species compositions and size structure in the Bay of Bengal. *Acta Oceanologica Sin.* 031, 121–128. doi: 10.1007/s13131-012-0198-8

Li, L. Y., Wu, C., Huang, D. Y., Ding, C. L., Wei, Y. Q., and Sun, J. (2021). Integrating stochastic and deterministic process in the biogeography of N_2-fixing *Cyanobacterium Candidatus Atelocyanobacterium Thalassa*. *Front. Microbiol.* 12:654646. doi: 10.3389/fmicb.2021.654646

Li, Y. Y., Chen, X. H., Xie, Z. X., Li, D. X., Wu, P. F., Kong, L. F., et al. (2018). Bacterial diversity and nitrogen utilization strategies in the upper layer of the Northwestern Pacific Ocean. *Front. Microbiol.* 9:797. doi: 10.3389/fmicb.2018. 00797

Liu, H. J., Wu, C., Wang, X. Z., Thangaraj, S., Zhang, G. C., Zhang, X. D., et al. (2020). Surface phytoplankton assemblages and controlling factors in the strait of Malacca and Sunda Shelf. *Front. Mar. Sci.* 7:33. doi: 10.3389/fmars.2020.00033

Löscher, C. R., Mohr, W., Bange, H. W., and Canfield, D. (2020). No nitrogen fixation in the Bay of Bengal? *Biogeosciences* 17, 851–864. doi: 10.5194/bg-17-851-2020

Lyimo, T. J. (2011). Distribution and abundance of the *Cyanobacterium Richelia intracellularis* in the coastal waters of Tanzania. *J. Ecol. Nat. Environ.* 3, 85–99.

Madhu, N. V., Paul, M., Ullas, N., Ashwini, R., and Rehitha, T. V. (2013). Occurrence of cyanobacteria (*Richelia intracellularis*)-diatom (*Rhizosolenia hebetata*) consortium in the Palk Bay, southeast coast of India. *Indian J. Geo Mar. Sci.* 42, 453–457.

Madhupratap, M., Gauns, M., Ramaiah, N., Kumar, S. P., Muraleedharan, P. M., de Sousa, S. N., et al. (2003). Biogeochemistry of the Bay of Bengal: Physical, chemical and primary productivity characteristics of the central and western Bay of Bengal during summer monsoon 2001. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 50, 881–896. doi: 10.1016/S0967-0645(02)00611-2

Magoč, T., and Salzberg, S. L. (2010). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Manjumol, C. C., Linoy, L. C., Albert, I. K. A., and Kripa, V. (2018). Occurrence of diatom – diazotrophic association in the coastal surface waters of south Andaman. *India. Symbiosis* 76, 293–302. doi: 10.1007/s13199-018-0559-y

Mazard, S., Ostrowski, M., Partensky, F., and Scanlan, D. J. (2011). Multi-locus sequence analysis, taxonomic resolution, and biogeography of marine *Synechococcus*. *Environ. Microbiol.* 14, 372–386. doi: 10.1111/j.1462-2920.2011. 02514.x

Milici, M., Deng, Z. L., Tomasch, J., Decelle, J., Wos-Oxley, M. L., Wang, H., et al. (2016). Co-occurrence analysis of microbial taxa in the Atlantic Ocean reveals high connectivity in the free-living bacterioplankton. *Front. Microbiol.* 7:649. doi: 10.3389/fmicb.2016.00649

Miyama, T., McCreary, J. P., Jensen, T. G., Loschnigg, J., Godfrey, S., and Ishida, A. (2003). Structure and dynamics of the Indian-Ocean cross-equatorial cell. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 50, 2023–2047. doi: 10.1016/S0967-0645(03) 00044-4

Mori, H., Maruyama, F., Kato, H., Toyoda, A., Dozono, A., Ohtsubo, Y., et al. (2014). Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. *DNA Res.* 21, 217–227. doi: 10.1093/dnares/dst052

Morris, R. M., Michael, S. R., Connon, S. A., Vergin, K. L., Siebold, W. A., Carison, C. A., et al. (2002). SAR11 clade dominate ocean surface bacterioplankton communities. *Nature* 420, 806–810. doi: 10.1038/nature 01240

Narvekar, J., and Prasanna, K. S. (2006). Seasonal variability of the mixed layer in the central Bay of Bengal and associated changes in nutrient and chlorophyll. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 53, 820–835. doi: 10.1016/j.dsr.2006.01.012

Narvekar, J., and Prasanna, K. S. (2014). Mixed layer variability and chlorophyll a biomass in the Bay of Bengal. *Biogeosciences* 11, 3819–3843. doi: 10.5194/bg-11-3819-2014

Natale, D. A., Shankavaram, U. T., Galperin, M. Y., Wolf, Y. I., Aravind, L., and Koonin, E. V. (2000). Genome annotation using clusters of orthologous groups of proteins (COGs) – towards understanding the first genome of a crenarchaeon. *Genome Biol.* 1:research0009.1. doi: 10.1186/gb-2000-1-5-research0009

Prasanna, K. S., Muraleedharan, P. M., Prasad, T. G., Gauns, M., Ramaiah, N., de Souza, S. N., et al. (2002). Why is the Bay of Bengal less productive during summer monsoon compared to the Arabian Sea? *Geophys. Res. Lett.* 29:2235. doi: 10.1029/2002GL016013

Prasanna, K. S., Narvekar, J., Kumar, A., Shaji, C., Sabu, P., Rijomon, G., et al. (2004). Intrusion of the Bay of Bengal water into the Arabian Sea during winter monsoon and associated chemical and biological response. *Geophys. Res. Lett.* 31:L15304. doi: 10.1029/2004GL020247

Rajpathak, S. N., Roumik, B., Mishra, P. G., Khedkar, A. M., Patil, Y. M., Joshi, S. R., et al. (2018). An exploration of microbial and associated functional diversity in the OMZ and non-OMZ areas in the Bay of Bengal. *J. Biosci.* 43, 635–648. doi: 10.1007/s12038-018-9781-2

Rappe, M. S., Connon, S. A., Vergin, K. L., and Glovannonl, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418, 630–633. doi: 10.1038/nature00917

Raven, J. A. (1998). Small is beautiful: The picophytoplankton. *Funct. Ecol.* 12, 503–513. doi: 10.1046/j.1365-2435.1998.00233.x

Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L., and Venter, J. C. (2010). Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16184–16189. doi: 10.1073/pnas. 1009513107

Sahu, B. K., Baliarsingh, S. K., Lotliker, A. A., Parida, C., Srichandan, S., and Sahu, K. C. (2017). Winter thermal inversion and *Trichodesmium* dominance in north-western Bay of Bengal. *Ocean Sci. J.* 52, 301–306. doi: 10.1007/s12601-017-0028-1

Sarma, V. V. S. S., Rao, G. D., Viswanadham, R., Sherin, C. K., Salisbury, J., Omand, M. M., et al. (2016). Effects of freshwater stratification on nutrients, dissolved oxygen, and phytoplankton in the Bay of Bengal. *Oceanography* 29, 222–231. doi: 10.5670/oceanog.20 16.54

Schlitzer, R. (2008). *Ocean data view*. Available online at: http://odv.awi.de (accessed May 19, 2018).

Schott, F. A., and Mccreary, J. P. (2001). The monsoon circulation of the Indian Ocean. *Prog. Oceanogr.* 51, 1–123. doi: 10.1016/S0079-6611(01)00083-0

Schott, F. A., Xie, S. P., and McCreary, J. P. (2009). Indian Ocean circulation and climate variability. *Rev. Geophys.* 47:RG1002. doi: 10.1029/2007RG000245

Shankar, D., Vinayachandran, P., and Unnikrishnan, A. (2002). The monsoon currents in the north Indian Ocean. *Prog. Oceanogr.* 52, 63–120. doi: 10.1016/S0079-6611(02)00024-1

Shetye, S., Shenoi, S., Gouveia, A., Michael, G. S., Sundar, D., and Nampoothiri, G. (1991). Wind-driven coastal upwelling along the western boundary of the Bay of Bengal during the southwest monsoon. *Cont. Shelf Res.* 11, 1397–1408. doi: 10.1016/0278-4343(91)90042-5

Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S., and Furuya, K. (2014). Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean. *Global Biogeochem. Cycles* 28, 1096–1110. doi: 10.1002/2014GB004886

Singh, A., and Ramesh, R. (2015). Environmental controls on new and primary production in the northern Indian Ocean. *Prog. Oceanogr.* 131, 138–145. doi: 10.1016/j.pocean.2014.12.006

Singh, A., Gandhi, N., Ramesh, R., and Prakash, S. (2015). Role of cyclonic eddy in enhancing primary and new production in the Bay of Bengal. *J. Sea Res.* 97, 5–13. doi: 10.1016/j.seares.2014.12.002

Swan, B. K., Martinez-Garcia, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamu, D., et al. (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333, 1296–1300. doi: 10.1126/science.1203690

Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B., and Zehr, J. P. (2016). Distinct ecological niches of marine symbiotic N2-fixing *Cyanobacterium Candidatus atelocyanobacterium thalassa* sublineages. *J. Phycol.* 53, 451–461. doi: 10.1111/jpy.12505

Uthermöhl, H. (1958). On the perfecting of quantitative phytoplankton method. *Int. Assoc. Theor. Appl. Limnol. Commun.* 9, 1–38.

Vecchi, G. A., and Harrison, D. E. (2002). Monsoon breaks and Subseasonal Sea surface temperature variability in the Bay of Bengal. *J. Clim.* 15, 1485–1493. doi: 10.1175/1520-0442(2002)015<1485:MBASSS>2.0.CO;2

Vidya, P. J., and Kumar, S. P. (2013). Role of mesoscale eddies on the variability of biogenic flux in the northern and central Bay of Bengal. *J. Geophys. Res. Oceans* 118, 5760–5771. doi: 10.1002/jgrc.20423

Vijayan, J., Kumar, V., and Ammini, P. (2020). Comparison of bacterial community structure in coastal and offshore waters of the Bay of Bengal, India. *Reg. Stud. Mar. Sci.* 39:101414. doi: 10.1016/j.rsma.2020.101414

Wang, J., Kan, J. J., Borecki, L., Zhang, X. D., Wang, D. X., and Sun, J. (2016). A snapshot on spatial and vertical distribution of bacterial communities in the East Indian Ocean. *Acta Oceanol. Sin.* 35, 85–93. doi: 10.1007/s13131-016-0871-4

Wei, Y. Q., Sun, J., Zhang, X. D., Wang, J., and Huang, K. (2018). Picophytoplankton size and biomass around equatorial eastern Indian Ocean. *Microbiologyopen* 8:e00629. doi: 10.1002/mbo3.629

West, N. J., Lepère, C., Manes, C. L., Catala, P., Scanlan, D. J., and Lebaron, P. (2016). Distinct spatial patterns of SAR11, SAR86 and *Actinobacteria* diversity along a transect in the ultra-oligotrophic South Pacific Ocean. *Front. Microbiol.* 7:234. doi: 10.3389/fmicb.2016.00234

Wijesekera, H. W., Jensen, T. G., Jarosz, E., Teagur, W. J., Metzger, E. J., Wang, D. W., et al. (2015). Southern Bay of Bengal currents and salinity intrusions during the northeast monsoon. *J. Geophys. Res. Oceans* 120, 6897–6913. doi: 10.1002/2015JC010744

Wu, C., Fu, F. X., Sun, J., Thangaraj, S., and Pujari, L. (2018). Nitrogen fixation by *Trichodesmium* and unicellular diazotrophs in the northern South China Sea and the Kuroshio in summer. *Sci. Rep.* 8:2415. doi: 10.1038/s41598-018-20743-0

Wu, C., Kan, J. J., Liu, H. J., Pujari, L., Guo, C. C., Wang, X. Z., et al. (2019). Heterotrophic bacteria dominate the diazotrophic community in the Eastern Indian Ocean (EIO) during pre-Southwest monsoon. *Microb. Ecol.* 78, 804–819. doi: 10.1007/s00248-019-01355-1

Wu, C., Kan, J. J., Narale, D. D., Liu, K., and Sun, J. (2022a). Dynamics of bacterial communities during a seasonal hypoxia at the Bohai Sea: Coupling and response between abundant and rare populations. *J. Environ. Sci.* 111, 324–339. doi: 10.1016/j.jes.2021.04.013

Wu, C., Sun, J., Liu, H. J., Xu, W. Z., Zhang, G. C., Lu, H. F., et al. (2022b). Evidence of the significant contribution of heterotrophic diazotrophs to nitrogen fixation in the Eastern Indian Ocean during pre-Southwest monsoon period. *Ecosystems* 25, 1066–1083. doi: 10.1007/s10021-021-00702-z

Xia, X. M., Cheung, S. Y., Endo, H., Suzuki, K., and Liu, H. B. (2019). Latitudinal and vertical variation of *Synechococcus* assemblage composition along 170° W transect from the South Pacific to the Arctic Ocean. *Microb.l Ecol.* 77, 333–342. doi: 10.1007/s00248-018-1308-8

Xia, X. M., Partensky, F., Garczarek, L., Suzuki, K., Guo, C., Cheung, S. Y., et al. (2016). Phylogeography and pigment type diversity of *Synechococcus* cyanobacteria in surface waters of the northwestern Pacific Ocean. *Environ. Microbiol.* 19, 142–158. doi: 10.1111/1462-2920.13541

Zehr, J. P., Mellon, M. T., and Zani, S. (1998). New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (nifH) genes. *Appl. Environ. Microbiol.* 64, 3444–3450. doi: 10.1128/AEM.64.9.3444-3450.1998

Zhang, Y. Y., Yang, Q. S., Ling, J., Nostrand, J. D. V., Shi, Z. J., Zhou, J. Z., et al. (2017). Diversity and structure of diazotrophic communities in mangrove rhizosphere, revealed by high-throughput sequencing. *Front. Microbiol.* 8:2032. doi: 10.3389/fmicb.2017.02032

Zhang, Y., Zhao, Z. H., Sun, J., and Jiao, N. Z. (2011). Diversity and distribution of diazotrophic communities in the South China Sea deep basin with mesoscale cyclonic eddy perturbations. *FEMS Microbiol. Ecol.* 78, 417–427. doi: 10.1111/j.1574-6941.2011.01174.x