



OPEN

# DNA metabarcoding reveals ecological patterns and driving mechanisms of archaeal, bacterial, and eukaryotic communities in sediments of the Sansha Yongle Blue Hole

Qingxia Li<sup>1,5</sup>, Yanli Lei<sup>1,3,4,6</sup>✉ & Tiegang Li<sup>2</sup>✉

The Sansha Yongle Blue Hole (SYBH) is the world's deepest marine blue hole with unique physicochemical characteristics. However, our knowledge of the biodiversity and community structure in SYBH sediments remains limited, as past studies have mostly focused on microbial communities in the water column. Here, we collected sediment samples from the aerobic zone (3.1 to 38.6 m) and the deep anaerobic zone (150 m, 300 m) of the SYBH and extracted DNA to characterize the archaeal, bacterial, and eukaryotic communities inhabiting these sediments. Our results showed that the archaeal and bacterial communities were dominated by Thaumarchaeota and Proteobacteria, respectively. The dominant taxa of eukaryotes in different sites varied greatly, mainly including Phaeophyceae, Annelida, Diatomea and Arthropoda. All three examined domains showed clear vertical distributions and significant differences in community composition between the aerobic and anaerobic zones. Sulfide played a prominent role in structuring the three domains, followed by salinity, nitrous oxide, pH, temperature and dissolved oxygen, all of which were positively correlated with the turnover component, the main contributor to beta diversity. Neutral community model revealed that stochastic processes contributed to more than half of the community variations across the three domains. Co-occurrence network showed an equal number of positive and negative interactions in the archaeal network, while positive interactions accounted for ~80% in the bacterial and eukaryotic networks. Our findings reveal the ecological features of prokaryotes and eukaryotes in SYBH sediments and shed new light on community dynamics and survival strategies in the special environment of marine blue holes.

Marine blue holes are special geological settings formed in the carbonate banks below the sea level and are dark blue in color<sup>1</sup>. They typically have steep-walled structures and stratified water columns characterized by oxygen deficit and high levels of hydrogen sulfide in deep layers<sup>2,3</sup>. The water exchange between marine blue holes and the open sea is severely restricted, resulting in a relatively stable environment with unique physical–chemical characteristics inside the holes, such as a stratified water column with a strong thermohaline<sup>2</sup>. The special geographical, physical, and chemical features of marine blue holes make them natural laboratories for studying marine biodiversity, biological adaptation, and evolution, as well as the interaction between biotic and abiotic factors<sup>4–6</sup>.

<sup>1</sup>Laboratory of Marine Organism Taxonomy and Phylogeny, Qingdao Key Laboratory of Marine Biodiversity and Conservation, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China. <sup>2</sup>Key Laboratory of Marine Sedimentology and Environmental Geology, First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China. <sup>3</sup>Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao 266237, China. <sup>4</sup>Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China. <sup>5</sup>Center for Marine Ranching Engineering Science Research of Liaoning, Dalian Ocean University, Dalian 116023, China. <sup>6</sup>University of Chinese Academy of Sciences, Beijing 100049, China. ✉email: leiyanli@qdio.ac.cn; tqgli@fio.org.cn

The Sansha Yongle Blue Hole (SYBH; 111°46′06″ E, 16°31′30″ N), located within a coral reef of the Yongle Atoll among the Xisha Islands in the South China Sea, is the world's deepest marine blue hole known so far, with the observed deepest portion at 301.19 m below the local 10-year mean sea level<sup>7</sup>. The three-dimensional (3D) morphology of the SYBH showed that the overall shape of the internal structure resembled a ballet dancer's shoe with two large transitions at 76–78 m and 158 m, and no water or material exchange with the outside open sea<sup>7</sup>. A comprehensive investigation of the hydrochemical and redox properties of the SYBH revealed the existence of two thermoclines, one at 13–20 m and the other at 70–150 m, which divided the water column into five stratified water layers, and the deep water below 90 m was anaerobic<sup>8</sup>. The existing topography and redox gradients in the water column of the SYBH have increased scientists' interest in the biodiversity and community structure.

Numerous studies on water column microeukaryotic<sup>9,10</sup> and prokaryotic plankton<sup>5,11–13</sup> including bacteria and archaea have expanded our knowledge of the composition, structure, and potential function of the biological communities in the SYBH. The study by Liu et al.<sup>9</sup> revealed obvious differences in the eukaryotic composition of the water column between the SYBH and the outer reef slope and demonstrated significant effects of turbidity and nitrite concentration on the eukaryotic community structure. Another metabarcoding study suggested that planktonic microeukaryotes in the SYBH were less sensitive to environments but significantly affected by cross-domain biointeraction<sup>10</sup>. Nevertheless, limited information is available concerning the diversity, community structure and spatial distribution of eukaryotic communities in SYBH sediments, and what shapes the eukaryotic assemblages in the sediments remains unclear. For prokaryotes, studies on bacterial community revealed the vertical stratification of community composition in SYBH water column and highlighted the dominance of the phylum Proteobacteria<sup>5,11–13</sup>. Phylogenetic analysis indicated that the SYBH microbial ecosystem was characterized by the presence of unidentified microorganisms in the deep water layer<sup>12</sup>, and a considerable number of new taxa were identified among culturable anaerobic bacteria in SYBH water<sup>13</sup>. However, whether there are also abundant unknown prokaryotic taxa in SYBH sediments remains to be explored.

While previous studies have demonstrated that the geographical isolation and in situ conditions of SYBH contribute to the high microbial and eukaryotic diversity in the water column<sup>5,9,14</sup>, SYBH sediment microorganisms have been poorly studied. This is because, the location of SYBH in the continental slope of the South China Sea, the presence of submerged reefs and the complex geological structures<sup>7,9</sup> limit access to the bottom of SYBH for sediment collection. The 18S ribosomal DNA study of SYBH sediments indicates high vertical diversity of foraminifera that is influenced by in situ O<sub>2</sub> conditions and reveals the preference of some foraminiferal taxa for anoxia which may be supported by their ability to denitrify<sup>14</sup>. These findings suggest that SYBH may serve as an excellent natural laboratory to explore additional eukaryotic adaptations associated with anoxia<sup>14</sup>. However, sediment communities have not been well documented in the SYBH, our understanding of the biodiversity and driving factors in SYBH sediments is still limited.

In our study, we collected a total of 12 sediment samples from the aerobic slope (3.1–38.6 m depth) and the anaerobic regime (150 m, and 300 m depth) of SYBH and we performed high-throughput sequencing of 16S and 18S rRNA gene amplicons to examine the prokaryotic (archaea and bacteria) and eukaryotic diversity and distribution in SYBH. We combined our sequencing data results with environmental data to: (1) describe the distribution, diversity, and co-occurrence of archaea, bacteria, and eukaryotes at different oxygen and geochemical regimes of SYBH; (2) compare the differences in community characteristics between the aerobic and anaerobic zones of SYBH; (3) analyze the main environmental factors that influence the composition of the prokaryotic and eukaryotic communities in SYBH sediments.

## Results

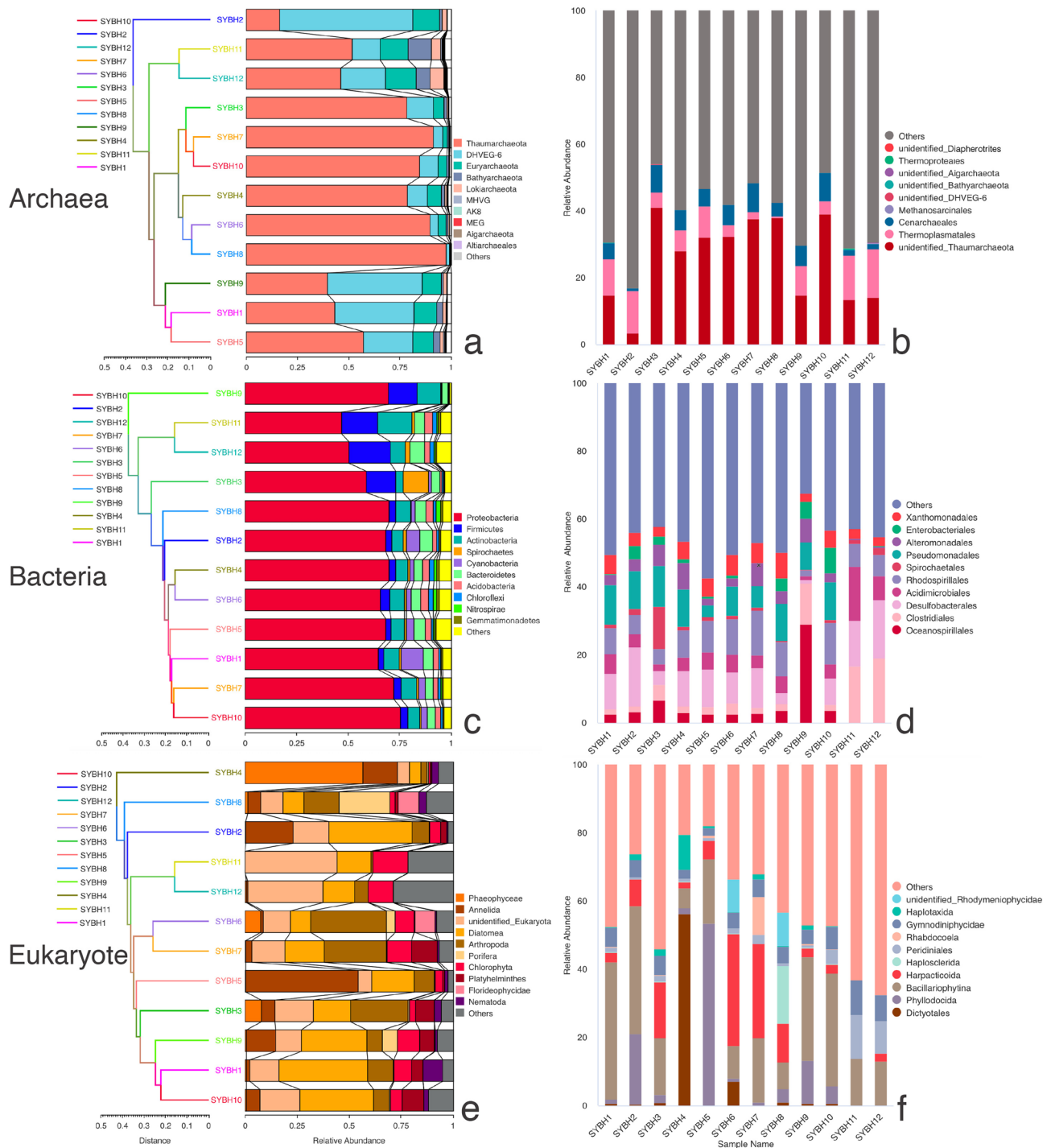
### Environmental metadata for SYBH sampling sites

The hydrochemical factors in the water column of SYBH at each sampling site are shown in Table S1. Temperature, pH and the concentrations of dissolved oxygen (DO) and nitrous oxide (N<sub>2</sub>O) generally decreased with depth. Temperature and pH decreased from 30 °C and 8.09 at the surface to 15 °C and 7.49 at the bottom, respectively. DO concentration in the upper water column was ~ 233 μmol L<sup>-1</sup> and decreased sharply, showing two concentration minima before becoming anoxic below 100 m (Table S1). Contrastingly, salinity, ammonia, dissolved inorganic carbon (DIC), methane, phosphate, silicate, and sulfide increased with depth (from surface to 150 m), maintaining nearly constant levels below 150 m. The concentration of nitrate reached maximum levels at 70 m (5.44 μmol L<sup>-1</sup>) and then sharply decreased to undetectable levels at a depth of 100 m. The concentrations of dissolved organic carbon (DOC), nitrite, particulate organic carbon (POC), and total suspended solid (TSS) fluctuated multiple times from the surface to the bottom, with a relatively stable trend below 200 m. The concentrations of DOC and POC reached a peak at the depth of 150 m.

### Taxonomic diversity and community composition

In the final dataset, we retained 2752 archaeal OTUs representing 223,545 reads, 6266 bacterial OTUs representing 384,644 reads, and 3608 eukaryotic OTUs representing 809,257 reads for downstream analysis (Table S2). The number of bacterial OTUs varied greatly among the 12 SYBH sites (SYBH1–SYBH10 aerobic; SYBH11–SYBH12 anaerobic), ranging from 1903 (SYBH9) to 3734 (SYBH11), with an average of 2864 OTUs per site. The difference in the number of eukaryotic OTUs among the 12 SYBH sites was small, ranging from 1056 (SYBH2) to 1615 (SYBH8), with an average of 1412 OTUs per site. The number of archaeal OTUs ranged from 535 (SYBH8) to 1606 (SYBH5), with an average of 965 OTUs per site. The number of bacterial OTUs and eukaryotic reads in the two anoxic sites (SYBH11 and SYBH12) was higher than that in the 10 aerobic sites (Fig. S1).

Our 16S and 18S rRNA gene analysis indicated 16 archaeal, 52 bacterial and 52 eukaryotic phyla. The top three archaeal taxa at the phylum level were Thaumarchaeota, DHVEG-6 and Euryarchaeota (Fig. 1). Thaumarchaeota was the most abundant archaeal phylum accounting for 73.49% of the total archaeal reads and occupied the



**Figure 1.** Community composition of archaea, bacteria, and eukaryotes at the phylum level (a, c, e) and order level (b, d, f). The UPGMA analysis was performed based on the Bray–Curtis dissimilarity index at the phylum level.

highest proportion of archaeal reads at 10 SYBH sites. DHVEG-6 had the highest proportion of archaeal reads at sites SYBH2 and SYBH9. Euryarchaeota, Bathyarchaeota and Lokiarchaeota presented higher relative abundance in the two anaerobic sites SYBH11 and SYBH12. The top three bacterial phyla were Proteobacteria, Firmicutes and Actinobacteria. Proteobacteria was the dominant phylum across all SYBH sites, accounting for 63.36% of the total bacterial reads. Firmicutes occupied a high proportion in the two anaerobic SYBH sites. The relative abundance of Actinobacteria in SYBH11 was the highest among the 12 sites. Cyanobacteria occupied higher relative abundance in the two shallow sites SYBH1 and SYBH2. The eukaryotic community varied greatly in composition across the 12 SYBH sites, showing different abundances at the examined sites. Annelida, Diatomea, and Arthropoda were relatively abundant phyla at all sites. The relative abundance of phylum Phaeophyceae at

the aerobic site SYBH4 exceeded 50% but was low in other sites. In the anaerobic sites SYBH11 and SYBH12, the eukaryotic community was dominated by the group of unidentified Eukaryota, and only Diatomea, Chlorophyta and Arthropoda had relative abundances greater than 1%. Additionally, the eukaryotic dataset contained 1005 fungal reads and 50 fungal OTUs representing three phyla (Chytridiomycota, Basidiomycota and Ascomycota). The relative abundance of Fungi was less than 0.3% in all SYBH sites (Table S2).

Detailed analysis at the site level revealed clear vertical variations in the alpha diversity of archaea, bacteria and eukaryotes, with different distributions across the 12 sites examined (Fig. S2). Notably, the alpha diversity of bacterial community in the anaerobic zone (SYBH11 and SYBH12) was higher than that in the aerobic zone. The unweighted pair group method with arithmetic mean (UPGMA) cluster analysis of archaeal, bacterial, and eukaryotic communities at the phylum level revealed that the two anaerobic sites, SYBH11 and SYBH12, were clustered on the same branch. The order-level analysis further confirmed that the community composition of SYBH11 and SYBH12 was similar. As the redox regime in the water column changed from aerobic to anaerobic/sulfidic, the relative abundance of some taxa such as Bathyarchaeota, Lokiarchaeota, Firmicutes, Thermoplasmatales, Acidimicrobiales, Clostridiales and Desulfobacterales became higher (Fig. 1). 2570 out of the 2752 archaeal OTUs were not assigned to any archaeal taxon at the order level, indicating that the vast majority of archaeal OTUs (over 93%) remain unidentified at the order level and below. The remaining 182 archaeal OTUs were attributed to 10 archaeal groups at the order level, of which 147 OTUs belong to Thermoplasmatales, five OTUs belong to Methanosarcinales, five OTUs belong to Cenarchaeales and only one OTU belong to Thermoproteales (Table S2). The group of unidentified Thaumarchaeota was represented by five OTUs but accounted for the highest proportion of archaeal reads at the order level (Fig. 1). In the bacterial and eukaryotic datasets, the number of OTUs identified at the order level were 4312 and 2553, accounting for 69% and 71% respectively.

### Correlations between community structure and environmental variables

To examine correlations between the dominant phyla and the environmental variables analyzed in this study, we performed nonparametric Spearman's correlation analysis (Table 1). The relative abundance of Euryarchaeota was negatively correlated with nitrate ( $r = -0.768$ ,  $p < 0.01$ ), nitrite ( $r = -0.839$ ,  $p < 0.01$ ), and  $N_2O$  ( $r = -0.837$ ,  $p < 0.01$ ), but positively correlated with sulfide ( $r = 0.640$ ,  $p < 0.05$ ), ammonia ( $r = 0.723$ ,  $p < 0.01$ ) and methane ( $r = 0.792$ ,  $p < 0.01$ ). The relative abundance of Proteobacteria exhibited positive correlations with nitrate ( $r = 0.824$ ,  $p < 0.01$ ), nitrite ( $r = 0.797$ ,  $p < 0.01$ ) and  $N_2O$  ( $r = 0.802$ ,  $p < 0.01$ ), but was negatively correlated with sulfide ( $r = -0.650$ ,  $p < 0.05$ ), ammonia ( $r = -0.754$ ,  $p < 0.01$ ) and methane ( $r = -0.844$ ,  $p < 0.01$ ). The relative abundance of Firmicutes was correlated with 11 environmental variables, of which temperature, DO, DOC, POC and pH had negative correlations with Firmicutes, whilst salinity, sulfide, phosphate, silicate, DIC and depth had positive correlations with Firmicutes. The correlations between Firmicutes and temperature and POC were significant ( $p < 0.01$ ), while the remaining nine correlations were not significant ( $p < 0.05$ ). The relative abundance of Arthropoda and Annelida exhibited a negative correlation with sulfide ( $p < 0.05$ ). Arthropoda showed a positive correlation with POC ( $p < 0.05$ ). No significant correlation was detected between the relative abundance of Thaumarchaeota, DHVEG-6, Actinobacteria and Diatomea and the examined environmental variables ( $p > 0.05$ ).

	Thaumarchaeota	DHVEG-6	Euryarchaeota	Proteobacteria	Actinobacteria	Firmicutes	Diatomea	Arthropoda	Annelida
T	-0.182	0.231	-0.056	0.021	-0.413	-0.713**	0.287	0.406	0.490
Salinity	0.091	-0.147	0.126	0.021	0.510	0.608*	-0.182	-0.434	-0.490
DO	-0.214	0.277	-0.018	0.032	-0.445	-0.669*	0.389	0.371	0.473
Sulfide	-0.253	0.124	0.640*	-0.650*	0.414	0.640*	-0.253	-0.591*	-0.650*
Nitrate	0.521	-0.408	-0.768**	0.824**	-0.113	-0.085	-0.007	0.289	0.380
Nitrite	0.566	-0.462	-0.839**	0.797**	0.014	-0.182	-0.070	0.420	0.343
Ammonia	-0.404	0.302	0.723**	-0.754**	-0.189	0.309	-0.189	-0.442	-0.211
Phosphate	0.141	-0.197	0.085	0.000	0.465	0.648*	-0.246	-0.451	-0.472
Silicate	0.179	-0.242	0.049	0.004	0.403	0.701*	-0.312	-0.452	-0.455
$N_2O$	0.564	-0.459	-0.837**	0.802**	-0.004	-0.179	-0.074	0.427	0.347
Methane	-0.515	0.410	0.792**	-0.844**	0.102	0.140	-0.025	-0.305	-0.385
DOC	-0.343	0.406	0.105	0.000	-0.476	-0.608*	0.497	0.168	0.517
POC	0.021	0.035	-0.268	0.197	0.070	-0.732**	0.296	0.599*	0.183
TSS	-0.403	0.414	0.342	-0.346	-0.150	-0.349	0.460	-0.018	0.089
DIC	0.168	-0.217	0.063	0.000	0.371	0.699*	-0.294	-0.420	-0.441
pH	-0.210	0.273	-0.021	0.027	-0.427	-0.671*	0.392	0.364	0.469
Depth	0.168	-0.217	0.063	0.000	0.371	0.699*	-0.294	-0.420	-0.441

**Table 1.** Results of Spearman's correlation analysis of dominant taxa at the phylum level in the archaeal, bacterial, and eukaryotic communities and the environmental variables examined, such as temperature (T), DO, ammonia, phosphate,  $N_2O$ , DOC, POC, TSS, and DIC. \*indicates insignificant correlation ( $p < 0.05$ ); \*\*indicates significant correlation ( $p < 0.01$ ).

Redundancy analysis (RDA) results showed that sulfide, salinity and  $N_2O$  had great impacts on the archaeal, bacterial and eukaryotic communities, among which sulfide was the most important factor affecting the three domains in SYBH sediments. The strongest influence of sulfide was at the two anaerobic sites, SYBH11 and SYBH12. Nitrate, nitrite, TSS and POC had relatively weak effects on the distribution of archaea, bacteria and eukaryotes (Fig. 2), but the effects were different for the three domains. As an example, BIOENV analysis revealed that nitrite and pH could best explain the archaeal community structure. In particular, the variation partitioning analysis (VPA) results indicated that nitrite and pH explained 13% of archaeal community variation. Environmental variables that correlated with the bacterial community assemblages included salinity, nitrate, and DO, explaining 22% of the bacterial community variation (Fig. 2). For eukaryotes, sulfide influenced the community structure and together with nitrite, explained 22% of community variation (Fig. 2).

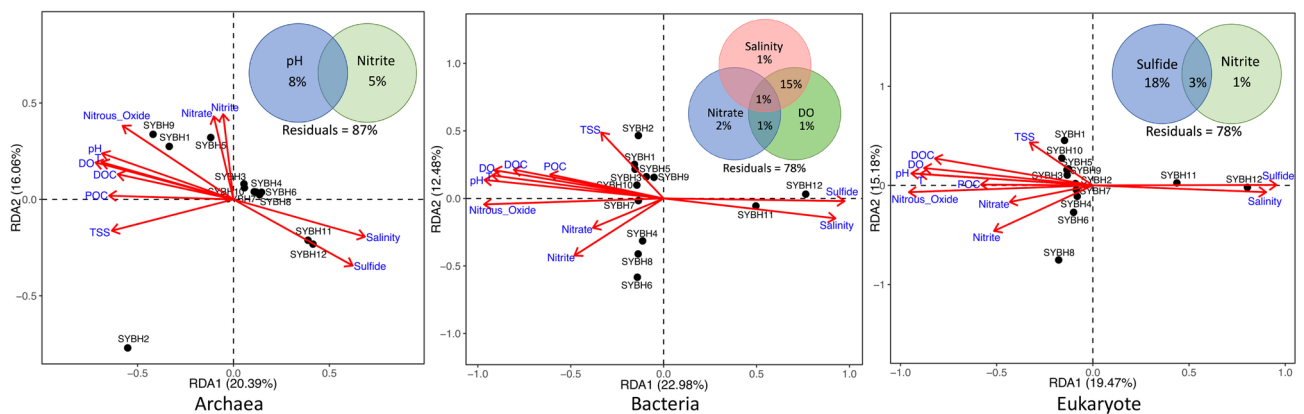
Partitioning beta-diversity using multiple-site dissimilarity measures showed that the turnover component ( $\beta_{sim}$ ) contributed at least eight times more to the Sørensen dissimilarity ( $\beta_{sor}$ ) of the archaeal, bacterial, and eukaryotic communities than the nestedness component ( $\beta_{sne}$ ). This indicates that the variation in species composition among sites was affected by species replacement between sites rather than species loss from site to site (Table 2).

Mantel test between the pairwise dissimilarity metrics and the Euclidean distance of environmental factors further revealed that the turnover component ( $\beta_{sim}$ ) of Sørensen dissimilarity of the archaeal, bacterial, and eukaryotic communities was influenced by a wide range of environmental variables, including water depth, temperature, salinity, DO, pH, ammonia, phosphate, silicate,  $N_2O$ , methane, sulfide and DIC (Fig. 3). There were nine significant correlations in the bacterial community, and five significant correlations in the eukaryotic community. Water depth, temperature, pH and methane had significant effects on the  $\beta_{sim}$  of both bacterial and eukaryotic communities ( $p < 0.01$ ). Furthermore, the  $\beta_{sim}$  of the bacterial and eukaryotic communities was correlated with DOC ( $p < 0.05$ ). Consistent with the RDA results, the  $\beta_{sim}$  of the three domains was not correlated with nitrate, nitrite, POC and TSS. The nestedness component ( $\beta_{sne}$ ) of the three domains was not significantly correlated with environmental variables.

### The community assembly and co-occurrence network analysis

The Sloan neutral community model showed that the  $R^2$  values in the archaeal, bacterial, and eukaryotic communities were 56.02%, 58.78% and 54.81%, respectively (Fig. 4). These results indicated that stochastic processes contributed to more than half of the community variations across the three domains. The  $m$  value in the bacterial community was the largest ( $m = 0.7163$ ), followed by that of the archaeal community ( $m = 0.4565$ ), and the  $m$  value in the eukaryotic community was the smallest ( $m = 0.1814$ ), indicating that the species dispersal of bacteria was relatively high, while the species dispersal in the eukaryotic community was limited.

The bacterial and eukaryotic genera with a relative abundance of more than 0.5% and all archaeal genera were selected to generate the co-occurrence network in the SYBH (Fig. 5). The bacterial network was more complex than the archaeal network and eukaryotic network at the genus level (Fig. 5a-c). The numbers of vertices in

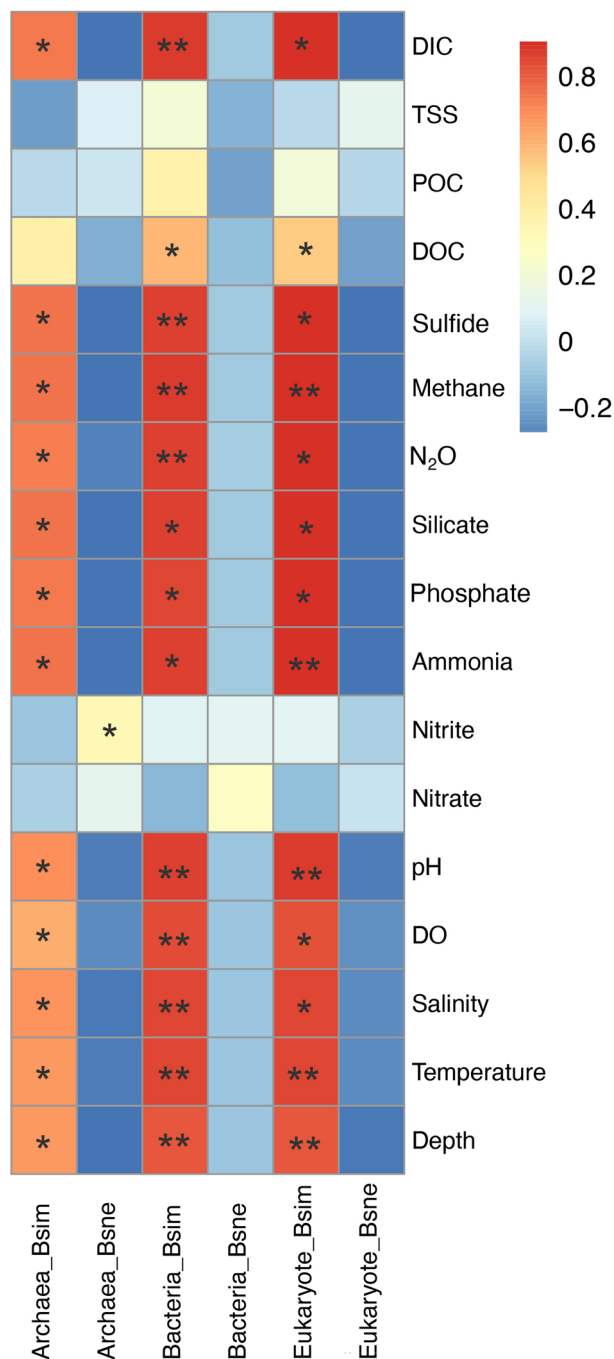


**Figure 2.** RDA for the archaeal, bacterial and eukaryotic communities to explore the relationships between community structure and environmental variables. Venn diagrams illustrate the relative contribution of the selected variables in shaping archaeal, bacterial, and eukaryotic communities as it was evaluated by VPA.

	$\beta_{sor}$	$\beta_{sim}$	$\beta_{sne}$
Archaea	0.8391	0.7504	0.0887
Bacteria	0.7411	0.6883	0.0528
Eukaryote	0.7630	0.7269	0.0361

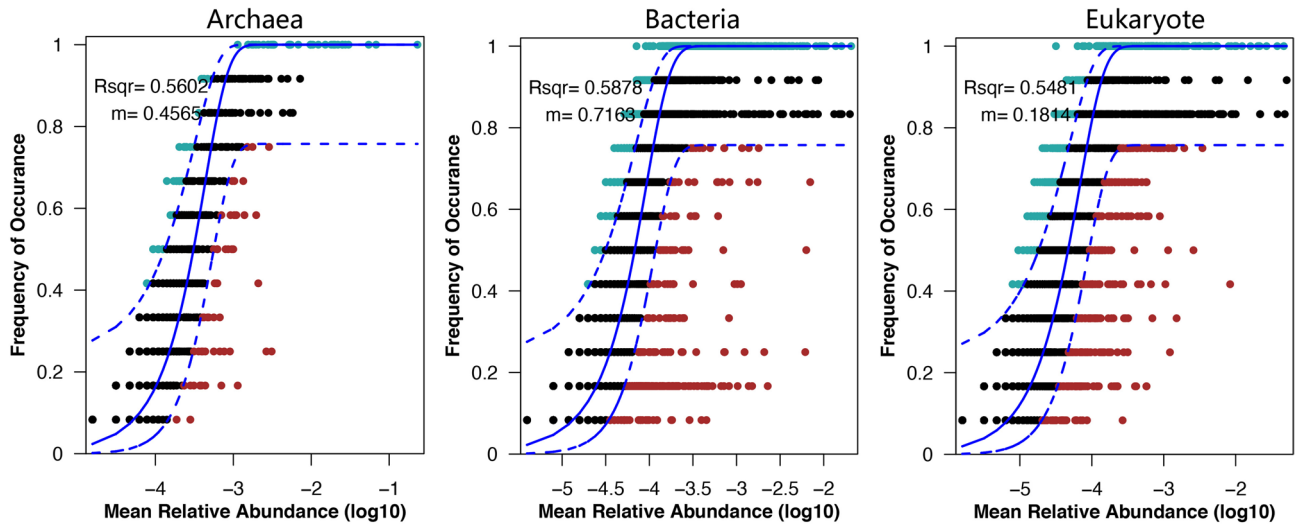
**Table 2.** Results of partitioning multiple-site beta diversity ( $\beta_{sor}$ ) into turnover component ( $\beta_{sim}$ ) and nestedness component ( $\beta_{sne}$ ) by using Baselga's method.



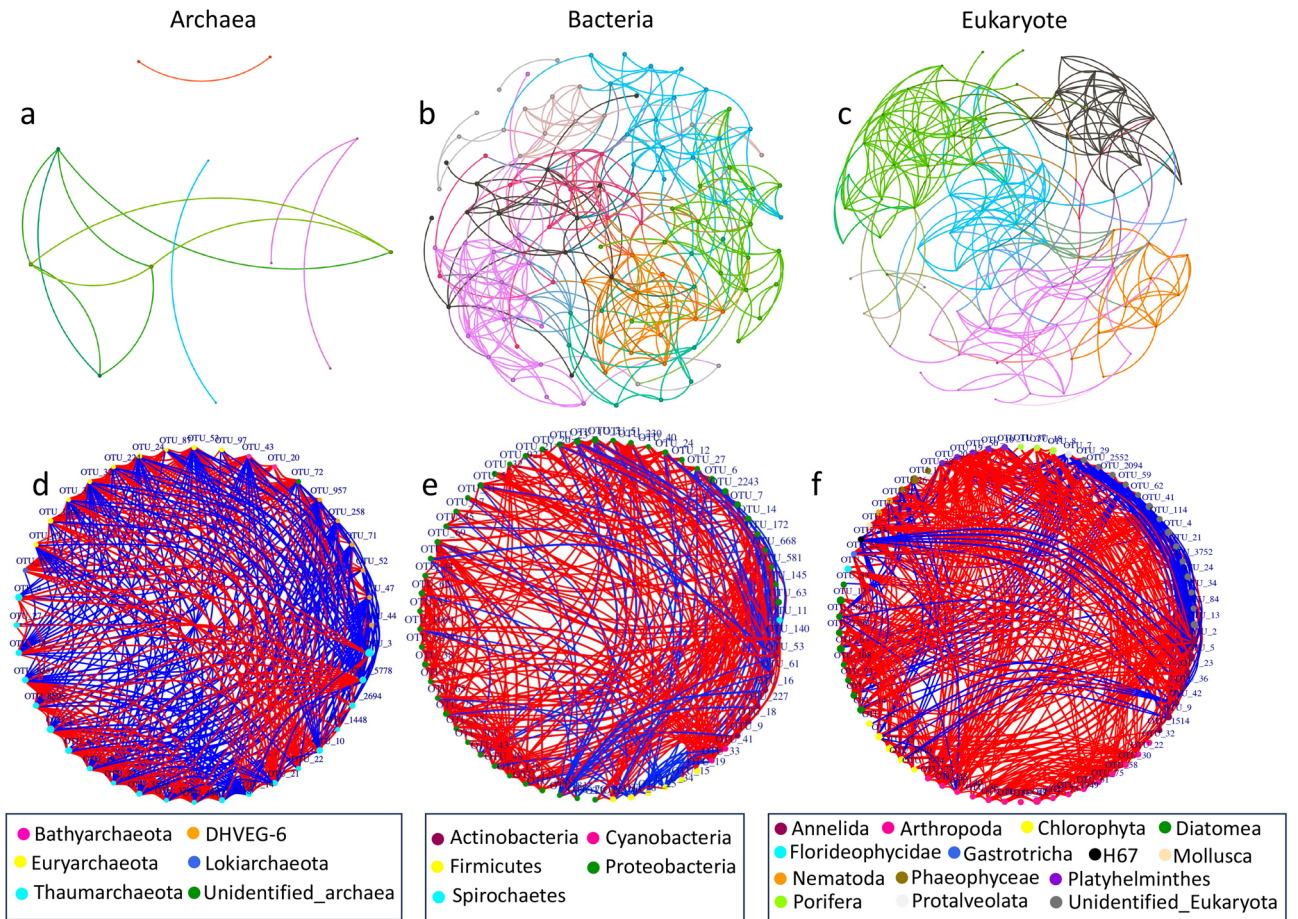


**Figure 3.** The relationships between two  $\beta$ -diversity components ( $\beta_{sim}$  and  $\beta_{sne}$ ) and the environmental variables which were determined by Mantel test based on Pearson's correlations. \* indicates insignificant correlation ( $p < 0.05$ ); \*\* indicates significant correlation ( $p < 0.01$ ).

the archaeal, bacterial, and eukaryotic networks at the genus level were 12, 121, and 87, respectively, and their edge numbers were 13, 385, and 291, respectively. The co-occurrence networks constructed based on OTU abundances reflected that the number of positive interactions was close to that of negative interactions (195 vs. 177) in the archaeal network. In the bacterial and eukaryotic networks, most of the interactions were positive, accounting for 82% and 80%, respectively (Fig. 5d-f). In the archaeal network, most of the keystone OTUs were derived from Thaumarchaeota and Euryarchaeota. Thaumarchaeota had negative interactions with Euryarchaeota and DHVEG-6, and Euryarchaeota showed positive interactions with Lokiarchaeota and Bathyarchaeota. Because only 7% of archaeal OTUs were assigned below the order level, more interactions may occur between archaeal taxa in SYBH, which warrants further investigation. In the bacterial network, most of the keystone OTUs belonged to the phylum Proteobacteria, which had strong positive interactions with Cyanobacteria. Strong negative interactions were observed between Firmicutes and Cyanobacteria. In the eukaryotic network, most of



**Figure 4.** Fitting of a neutral community model (NCM) to the community assemblage at the OTU level. The frequencies of occurrence of archaea, bacteria and eukaryotes were predicted. The parameters  $R^2$  and  $m$  represent the fit to the neutral model and the immigration rate, respectively.



**Figure 5.** Co-occurrence networks at the genus level (a-c) and at the OTU level (d-f). The networks at the genus level were colored according to the modularity class. In the networks based on OTUs, red lines represent positive interactions among OTUs, and the blue lines represent negative interactions among OTUs.

the highly connected OTUs were derived from the Unidentified Eukaryota, indicating putative importance of the unidentified fraction of eukaryotes in SYBH sediments.

## Discussion

As the deepest marine blue hole known in the world, Sansha Yongle Blue Hole has high scientific research values and has attracted widespread attention. Our study presented the first DNA metabarcoding dataset covering the three domains of cellular life (archaea, bacteria, and eukaryotes) from the SYBH sediment samples. Further it provides a comprehensive analysis of the molecular diversity, community structure and assembly mechanism, and indicates the role of environmental factors in shaping the prokaryotic and eukaryotic communities in the SYBH.

Marine blue holes are generally characterized by steep physicochemical gradients within geographically restricted areas<sup>4,13</sup>, creating conditions for new, unidentified microbial lineages to exist<sup>6</sup>. High-quality metagenome-assembled genomes (MAGs) retrieved recently from the Amberjack Hole (Gulf of Mexico blue hole) showed that microbial communities were dominated by understudied microbial taxa, indicating high levels of novel microbial lineages in the Amberjack Hole<sup>6</sup>. Similarly, some studies indicated that the microbial ecosystems in SYBH water column were also characterized by unidentified taxa<sup>11,12</sup>. A recent study on microbial community in SYBH water samples recovered high-quality MAGs belonging to various uncultivated lineages, reflecting the high novelty of the SYBH microbiome<sup>15</sup>. In this study, 29% of eukaryotic OTUs, 31% of bacterial OTUs, and more than 93% of archaeal OTUs remained unidentified at the order level, indicating the existence of a large number of undocumented prokaryotic and eukaryotic taxa in SYBH sediments. The high percentages of undocumented taxa in our study agree with the high number of unidentified lineages in the water column of SYBH<sup>11,12,15</sup>, which indicates that microorganisms inhabiting SYBH (both sediment and water column) are underrepresented in public databases. This hinders adequate taxonomic identifications of the in situ prokaryotic and eukaryotic communities in SYBH. The presence of high levels of novel lineages in the SYBH is not surprising, as this phenomenon has also been observed in other marine habitats. Recent studies have revealed novel lineages of Southern Ocean deep-sea foraminifera<sup>16</sup> and high genetic novelty of foraminifera on Western Pacific seamounts<sup>17,18</sup>. The unique environmental characteristics of habitats such as marine blue holes and seamounts provide conditions for the emergence of new species.

This study also revealed the community composition of the three domains at the phylum and order levels. To summarize, the bacterial community in SYBH sediments was dominated by Proteobacteria (Fig. 1), which is consistent with the results of bacterial communities in SYBH water<sup>5,11–13</sup>. The relative abundance of Proteobacteria in deep anaerobic sites was lower than that in upper aerobic sites (Fig. 1). This may be mainly related to nutrient availability, as the relative abundance of Proteobacteria was significantly positively correlated with nitrate and nitrite, which were undetectable below 100 m. As for the archaeal community, the study by He et al.<sup>5</sup> revealed the dominance of Euryarchaeota in SYBH water, but a recent study by Chen et al.<sup>19</sup> showed that Nanoarchaeota and Aagenigamarchaeota dominated the archaeal community in SYBH water. This divergence in results may be due to differences in the primers used, but the exact cause deserves further exploration. Different from the above research results<sup>5,19</sup>, a study on the archaeal communities in the deep-sea sediments of the South China Sea showed that archaeal communities in sediments were dominated by Thaumarchaeota (55%)<sup>20</sup>. Our results on the fraction of archaeal community that were able to assign taxonomy confirmed the dominance of Thaumarchaeota, which dominated the archaeal community at 10 SYBH sites with the relative abundance ranging from 43 to 98% (Fig. 1). DHVEG-6 was the second most abundant taxon in the archaeal community at the phylum level, with the relative abundance ranging from 2 to 65% across all SYBH sites. The dominant taxa of eukaryotic community varied greatly among the 12 SYBH sites. In aerobic sites, the dominant eukaryotic taxa included Annelida, Arthropoda and Diatomea, while the eukaryotic communities in anaerobic sites were dominated by unidentified Eukaryota. Notably, in the study of SYBH water column by Chen et al.<sup>19</sup>, a total of 7888 amplicon sequence variants (ASVs) were identified from the fungi dataset. However, only a small fraction of fungi (50 fungal OTUs representing 1005 reads) were detected in our eukaryotic dataset (Table S2). The fact that our dataset did not recruit many fungal OTUs may be caused by primer specificity, as we used general 18S rRNA primers rather than ITS primers.

Vertical variation is a typical distribution pattern for biological communities in the water column of the SYBH which has been observed in the eukaryotic<sup>14,19</sup> and prokaryotic communities<sup>5,13,21</sup>. The water column in SYBH is generally divided into an aerobic surface layer (at depths of 0–70 m), an intermediate layer (at depths of 70–100 m) and an anoxic bottom layer (below 100 m) along the depth gradient<sup>8</sup>. These different water layers have different microbial compositions<sup>5</sup> and redox processes<sup>8</sup>. Although we did not obtain sediment samples within chemocline (at depths of 70–100 m), our results still clearly reveal vertical changes in prokaryotic and eukaryotic communities in SYBH sediments with water depth. In the aerobic layer close to surface and within the euphotic zone (3.1–38.6 m), Cyanobacteria capable of oxygen-producing photosynthesis had the highest relative abundance in the shallowest site SYBH1 where the DO concentration was the highest. As the redox regime changed from aerobic to anaerobic/sulfidic, the relative abundance of some taxa became higher, such as Bathyarchaeota, Lokiarchaeota, Firmicutes, Thermoplasmatales, Acidimicrobiales, Clostridiales and Desulfobacterales (Fig. 1). These taxa may be associated with crucial biogeochemical processes and play key roles in maintaining the stability of ecosystems in anaerobic zones. Bathyarchaeota are a globally distributed archaeal phylum and important members of global biogeochemical cycles that are shown to be highly abundant in anoxic sediments and are considered to utilize a wide range of labile and recalcitrant carbon sources<sup>22</sup>. Firmicutes, a group of syntrophic bacterial taxa, can participate in biogeochemical cycles by degrading organic polymers and lignocellulosic plant material to H<sub>2</sub>, or converting small molecular compounds to H<sub>2</sub> and acetate<sup>12</sup>. Methanogenic archaea (phylum Euryarchaeota, order Methanosarcinales) were also identified from nine SYBH sites, and their relative abundance was higher in anaerobic zones with high levels of methane concentration (Table S2).



Our findings highlight differences in community composition between the aerobic and anaerobic zones of the SYBH and demonstrate that SYBH is a natural laboratory for understanding ecosystem function under oxygen deficiency across redox gradients.

As mentioned, the relative abundance of some prokaryotic and eukaryotic taxa varied greatly between SYBH sites, such as Thaumarchaeota, DHVEG-6, Actinobacteria, Annelida, Arthropoda and Diatomea. Inferring from the results of our analysis, the distinct variations in relative abundance of these taxa between sites may be determined not only by environmental factors but also by strong biotic interactions. A growing number of studies emphasize that biotic interactions play an important role in influencing community diversity<sup>23</sup>, driving key ecological processes<sup>24</sup>, and governing species distributions at macroecological scales<sup>25</sup>. Co-occurrence network analysis is a useful approach for exploring complex interactions within communities<sup>26</sup>. The co-occurrence networks in this study indicated that the above-mentioned taxa had complex interaction networks (Fig. 5). For example, Thaumarchaeota had negative interactions with Euryarchaeota and DHVEG-6, Annelida exhibited negative interactions with Chlorophyta, and Diatomea showed positive interactions with Chlorophyta. Members of Thaumarchaeota, Arthropoda and Annelida also showed a high proportion of intra-phyllum interactions. Chen et al.<sup>10</sup> suggested that abiotic factors played a minor role in shaping microeukaryotic plankton community in SYBH, and interspecies cooperation might be one of the ecological strategies. Here, co-occurrence networks suggest that biotic interactions (e.g., mutualism, commensalism, synergism) may play an important role in the distribution of these dominant taxa of prokaryotic and eukaryotic communities in SYBH sediments. Furthermore, co-occurrence network analysis can provide new insights into keystone species in communities<sup>26</sup>. In this study, the keystone OTUs in the archaeal, bacterial, and eukaryotic networks were mainly derived from Thaumarchaeota, Proteobacteria and unidentified Eukaryota, respectively, indicating that they potentially exerted a considerable impact on the community. At the OTU level, most correlations in bacterial and eukaryotic networks were positive (Fig. 5), consistent with previous research<sup>5</sup>, reflecting the widespread beneficial interspecific interactions such as mutualism, commensalism, and facilitation<sup>27</sup> in the bacterial and eukaryotic communities. In the archaeal network, the number of positive correlations was close to the number of negative correlations (Fig. 5), which indicated that negative interspecific interactions (e.g., competition) and the aforementioned positive interspecific interactions played roles in shaping community characteristics.

The mechanisms that shape microbial community structure have always been a key issue in understanding community dynamics in the SYBH. The results of He et al.<sup>5</sup> suggested that temperature and nitrate concentration had significant contribution to the heterogeneous distribution of major bacterial clades and salinity explained most variations of the archaeal communities in SYBH water column. Similarly, Zhang et al.<sup>12</sup> illustrated the significant correlations between the bacteria and environmental elements of DO, temperature, salinity, pH, sulfur and nutrient. However, the above research results are all focused on the microbial community in SYBH water. The RDA results in our study illustrated that sulfide, salinity, N<sub>2</sub>O, pH, DO and temperature had great impacts on the archaeal, bacterial and eukaryotic communities, among which sulfide was the most prominent factor affecting the community structure in SYBH sediments (Fig. 2). Finally, our beta diversity metrics may reflect spatial turnover and nestedness of assemblages, caused by species replacement and species loss, which are common traits that influence biodiversity<sup>28</sup>. By applying Baselga's approach<sup>29</sup>, our results show that the turnover component ( $\beta_{sim}$ ) is the main contributor to beta diversity ( $\beta_{sor}$ ) of the archaeal, bacterial and eukaryotic communities in SYBH sediments (Table 2). This indicates that the overall patterns of multiple-sites dissimilarity of the three domains in SYBH sediments are driven by species replacement rather than species loss, which may be the consequence of either environmental sorting or spatial and historical constraints<sup>30</sup>. The Mantel test further revealed that the turnover component ( $\beta_{sim}$ ) of the three domains was influenced by a wide range of environmental variables, including sulfide, salinity, N<sub>2</sub>O, pH, DO and temperature (Fig. 3). Compared with bacterial and eukaryotic communities, the  $\beta_{sim}$  of archaeal community seems to be less affected by environmental variables. Quantifying the relative importance of deterministic and stochastic processes that shape microbial community assembly is considered a central challenge in ecology<sup>31,32</sup>, and also for SYBH. The NCM proposed by Sloan et al.<sup>33</sup> is particularly useful in quantifying the importance of neutral processes and has been employed in bacterial community<sup>34–36</sup>, eukaryotic community<sup>37</sup> and microeukaryotic community<sup>38,39</sup>. Although stochastic processes seem to influence planktonic microeukaryotes and deterministic processes dominate in prokaryotic plankton in the SYBH<sup>10</sup>, our data from SYBH sediments indicate that stochastic processes appear to shape the sediment communities. Our results show that stochastic processes contributed to ~55% of the variations in archaeal, bacterial, and eukaryotic communities in SYBH sediments (Fig. 4), indicating that stochastic processes might play more important roles than deterministic processes in the sedimental communities. Overall, the SYBH supports an interesting ecosystem that warrants further investigation in the future to decipher the metabolic capabilities of the identified taxa. This will facilitate our understanding of the biogeochemistry that exists in marine blue holes and how it is shaped by the in situ microbial communities.

## Materials and methods

### Sample collection

Sediment samples were collected from 12 sites at different water depths in the SYBH on board R/V CHANGHE OCEAN on May 17–28, 2017. Among the 12 sites, 10 sites were located in the oxygenated mixed layer with water depth ranging from 3.1 to 38.6 m, and the sediment samples from these 10 sites were recovered by scuba diving. The other two sediment samples were collected from the anaerobic layer at 150 and 300 m respectively with a ROV. All sediment samples were stored in liquid nitrogen on board until further processing. At each site, the hydrochemical properties were measured in the water column at the depth where sediment samples were collected. The supplementary information of the sediment samples can be found in Table S1 and Figure S2 of Li et al.<sup>14</sup>.

### DNA extraction, PCR amplification and illumina sequencing

DNA was extracted from a 0.25 g sediment using the DNeasy PowerSoil kit (QIAGEN, Germany) following the manufacturer's instructions. One negative control without sediment was incorporated for each extraction session and three DNA replicates were extracted from each sample.

The V3-V4 hypervariable region of the bacterial 16S rRNA gene was amplified using the primers 341F (5'-CCTAYGGGRBGCASCAG-3') and 806R (5'-GGACTACNNGGTATCTAAT-3')<sup>40</sup>. The V4 region of 18S rRNA gene was amplified using the primers 528F (5'-GCGGTAATTCAGCTCCAA-3') and 706R (5'-AATCCRAGAATTTACCTCT-3')<sup>41</sup> for eukaryote. The V4 region of the archaeal 16S rRNA gene was amplified using the primers Arch519F (5'-CAGCCGCCGCGGTAA-3') and Arch915R (5'-GTGCTCCCCGCCAATTCCT-3')<sup>42</sup>. Forward and reverse primers were tagged with six nucleotide-long sequences appended at their 5'-end to multiplex the PCR products in a unique sequencing library. All PCR reactions were performed in a total volume of 30  $\mu$ L containing 15  $\mu$ L of Phusion High-Fidelity PCR Master Mix (New England Biolabs, Beverly, MA, USA), 0.2  $\mu$ M of forward and reverse primers, and about 10 ng template DNA. Thermal cycling consisted of a pre-denaturation at 98 °C for 1 min, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 50 °C for 30 s, and extension at 72 °C for 30 s, followed by a final extension step at 72 °C for 5 min. One PCR blank control without DNA was included during each amplification. Each PCR product was mixed with same volume of 1  $\times$  loading buffer, and electrophoresis was performed on 2% agarose gel for detection. The triplicated PCR products of one sample were mixed in equidensity ratios and the mixed PCR products were purified with Qiagen Gel Extraction Kit (Qiagen, Germany).

Sequencing libraries were generated using TruSeq DNA PCR-Free Sample Preparation Kit (Illumina, USA) following manufacturer's recommendations. The library quality was assessed by the Qubit 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. The library was sequenced on an Illumina HiSeq2500 platform and 250 bp paired-end reads were generated at the Novogene Bioinformatics Technology Co., Ltd (Beijing). The sequencing reads were submitted to the Sequence Read Archive under accession number PRJNA998750.

### Sequence processing

Raw paired-end reads were de-multiplexed to samples based on their unique barcode and truncated by cutting off the barcode and primer sequences. They were merged using FLASH<sup>43</sup>, and the spliced sequences were called raw reads. Quality filtering on the raw reads were performed under specific filtering conditions according to the QIIME<sup>44</sup> quality controlled process and low-quality sequences were filtered out. Chimera sequences were removed from reserved sequences using UCHIME algorithm. High-quality sequences were clustered into operational taxonomic units (OTUs) at a 97% similarity level using UPARSE pipeline<sup>45</sup>. The taxonomic assignment of bacterial OTUs and archaeal OTUs was performed using the Mothur method and the SILVA database (version 132)<sup>46</sup>, and the eukaryotic OTUs were taxonomically classified using the Ribosomal Database Project (RDP) classifier and SILVA database (version 132)<sup>46</sup>. After taxonomic assignment, we excluded the OTUs that could not be assigned to the target community.

### Diversity and statistical analysis

Before calculating alpha and beta diversity, we normalized the data of each sample using a standard of reads number corresponding to the sample with the least reads using the R package vegan. The standard of reads number for archaea, bacteria and eukaryotes was 5340, 21,029 and 52,561, respectively. Alpha diversity estimated by Chao1 and Shannon index of the archaeal, bacterial, and eukaryotic communities were calculated on the normalized data with R packages vegan and picante. Community composition of archaea, bacteria, and eukaryotes at the phylum and order levels were visualized in Fig. 1 with the R package ggplot2<sup>47</sup>. The UPGMA analysis was performed based on the Bray–Curtis dissimilarity index at the phylum level (Fig. 1).

Preliminary detrended correspondence analysis (DCA) was used to evaluate whether RDA or CCA was more suitable for exploring the relationships between community and environmental variables<sup>48</sup>. Preliminary DCA results revealed that the longest gradient length of the archaeal, bacterial, and eukaryotic communities was 2.69, 3.37, and 3.95 (> 3 standard deviations), respectively. Therefore, RDA was chosen for archaea, bacteria and eukaryotes to assess the relationships between community structure and environmental variables<sup>49,50</sup>. To improve normality and homoscedasticity, all environmental factors were  $\log(x + 1)$  transformed before analysis. BIOENV and VPA analysis in R package vegan were used to identify the major environmental variables and evaluate their relative contribution in driving distribution patterns across the three domains (Fig. 2).

### Partitioning of beta diversity

The beta-diversity as Sørensen index ( $\beta_{sor}$ ) of the archaeal, bacterial, and eukaryotic communities were partitioned into two components, turnover ( $\beta_{sim}$ ) and nestedness ( $\beta_{sne}$ ), by applying Baselga's approach (Table 2). The multiple-site measures of compositional dissimilarity across all sites and pairwise between-site dissimilarity matrices were computed from the presence-absence matrix using the function beta.multi and beta.pair of the R package "betapart", respectively<sup>29</sup>. The relationships between the environmental variables and the  $\beta_{sim}$  and  $\beta_{sne}$  were determined by Mantel test based on Pearson's correlations (Fig. 3).

### Community assembly process

The potential importance of stochastic processes to community assembly was evaluated using the Sloan's NCM, which predicted the relationship between the occurrence frequency of OTUs and their relative abundance in the metacommunity<sup>33</sup>. The model used here is an adaptation of the neutral theory<sup>51</sup> adjusted to fit large microbial populations, and the analysis was performed using non-linear least-squares fitting with the R package minpack.

$lm^{52}$ . The parameters  $R^2$  and  $m$  represent the fit to the neutral model and the immigration rate, respectively (Fig. 4). Calculation of 95% confidence intervals around all fitting statistics was done by bootstrapping with 1000 bootstrap replicates.

### Co-occurrence network construction

Co-occurrence network analysis was conducted at the genus level and OTU level using R packages psych, igraph and Hmisc and visualized in Fig. 5. At the genus level, the bacterial genus and eukaryotic genus with a relative abundance of more than 0.5% and all archaeal genera were selected to generate the co-occurrence patterns (Fig. 5a–c). A Spearman's coefficient of greater than 0.7 and a significance level of less than 0.05 indicated a significant correlation. The network diagram was generated by Gephi software<sup>53</sup>. At the OTU level, only the OTUs with a relative abundance of more than 3% were selected to generate the co-occurrence patterns to show the interactions among OTUs (Fig. 5d–f). A Spearman's coefficient of greater than 0.6 and a significance level of less than 0.05 indicated a significant correlation.

### Data availability

The sequencing data can be downloaded from the NCBI Sequence Read Archive (SRA) under accession number PRJNA998750 (<https://www.ncbi.nlm.nih.gov/bioproject?term=PRJNA998750&cmd=DetailsSearch>).

Received: 21 August 2023; Accepted: 15 March 2024

Published online: 21 March 2024

### References

1. Mylroie, J. E., Carew, J. L. & Moore, A. I. Blue holes: Definition and genesis. *Carb. Evapor.* **10**, 225–233. <https://doi.org/10.1007/BF03175407> (1995).
2. Canganella, F., Bianconi, G., Kato, C. & Gonzalez, J. Microbial ecology of submerged marine caves and holes characterized by high levels of hydrogen sulphide. In *Life in Extreme Environments* (eds Amils, R. et al.) 115–124 (Springer, 2006).
3. Seymour, J. R., Humphreys, W. F. & Mitchell, J. G. Stratification of the microbial community inhabiting an anchialine sinkhole. *Aquat. Microb. Ecol.* **50**, 11–24. <https://doi.org/10.3354/ame01153> (2007).
4. Iliffe, T. M. & Kornicker, L. S. Worldwide diving discoveries of living fossil animals from the depths of anchialine and marine caves. *Smithson Contrib. Mar. Sci.* **38**, 269–280 (2009).
5. He, H. et al. Community structure, abundance and potential functions of bacteria and archaea in the Sansha Yongle Blue Hole, Xisha, South China Sea. *Front. Microbiol.* **10**, 2404 (2019).
6. Patin, N. V. et al. Gulf of Mexico blue hole harbors high levels of novel microbial lineages. *ISME J.* **15**, 2206–2232 (2021).
7. Li, T. G. et al. Three-dimensional (3D) morphology of Sansha Yongle Blue Hole in the South China Sea revealed by underwater remotely operated vehicle. *Sci. Rep.* **8**, 17122. <https://doi.org/10.1038/s41598-018-35220-x> (2018).
8. Xie, L. P. et al. Hydrochemical properties and chemocline of the Sansha Yongle Blue Hole in the South China Sea. *Sci. Total Environ.* **649**, 1281–1292. <https://doi.org/10.1016/j.scitotenv.2018.08.333> (2019).
9. Liu, Y. T. et al. Environmental DNA sequencing reveals a highly complex eukaryote community in Sansha Yongle Blue Hole, Xisha, South China Sea. *Microorganisms* **7**, 624. <https://doi.org/10.3390/microorganisms7120624> (2019).
10. Chen, T. Y. et al. Metabarcoding reveals the differential sensitivity of planktonic microbiome to environmental filtering and biointeraction in Sansha Yongle blue hole. *Front. Mar. Sci.* **9**, 1046808. <https://doi.org/10.3389/fmars.2022.1046808> (2022).
11. He, P. Q. et al. Microbial diversity and metabolic potential in the stratified Sansha Yongle Blue Hole in the South China Sea. *Sci. Rep.* **10**, 5949. <https://doi.org/10.1038/s41598-020-62411-2> (2020).
12. Zhang, Y. X. et al. Unique bacterial communities and potential function along the vertical gradient in the deepest marine blue hole. *Environ. Microbiol. Rep.* **13**(6), 911–927. <https://doi.org/10.1111/1758-2229.13001> (2021).
13. Zhou, S. et al. Unique bacterial communities and lifestyles in deep ocean blue holes: Insights from the Yongle Blue Hole (South China Sea). *Front. Mar. Sci.* **10**, 1086117. <https://doi.org/10.3389/fmars.2023.1086117> (2023).
14. Li, Q. X., Lei, Y. L., Morard, R., Li, T. G. & Wang, B. D. Diversity hotspot and unique community structure of foraminifera in the world's deepest marine blue hole—Sansha Yongle Blue Hole. *Sci. Rep.* **10**, 10257. <https://doi.org/10.1038/s41598-020-67221-0> (2020).
15. Chen, X. et al. Phylogenetically and metabolically diverse autotrophs in the world's deepest blue hole. *ISME Commun.* **3**(1), 117. <https://doi.org/10.1038/s43705-023-00327-4> (2023).
16. Pawlowski, J., Fontaine, D., da Silva, A. A. & Guiard, J. Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. *Deep Sea Res. Part II Topical Stud. Oceanography* **58**(19–20), 1996–2003 (2011).
17. Shi, J. F., Lei, Y. L., Li, Q. X., Lyu, M. & Li, T. G. Molecular diversity and spatial distribution of benthic foraminifera of the seamounts and adjacent abyssal plains in the tropical Western Pacific Ocean. *Marine Micropaleontol.* **156**, 101850. <https://doi.org/10.1016/j.marmicro.2020> (2020).
18. Shi, J. F., Lei, Y. L., Li, H. T. & Li, T. G. NGS-metabarcoding revealing novel foraminiferal diversity in the Western Pacific Magellan Seamount sediments. *J. Oceanol. Limnol.* **39**, 1718–1729. <https://doi.org/10.1007/s00343-021-0427-0> (2021).
19. Chen, B. et al. The diversity, community dynamics, and interactions of the microbiome in the world's deepest blue hole: Insights into extreme environmental response patterns and tolerance of marine microorganisms. *Environ. Microbiol.* **11**(6), e0053123. <https://doi.org/10.1128/spectrum.00531-23> (2023).
20. Li, Y. T. et al. Archaeal communities in the deep-sea sediments of the South China Sea revealed by Illumina high-throughput sequencing. *Ann. Microbiol.* **69**, 839–848. <https://doi.org/10.1007/s13213-019-01477-4> (2019).
21. Li, B. et al. Vertical variation in *Vibrio* community composition in Sansha Yongle Blue Hole and its ability to degrade macromolecules. *Mar. Life Sci. Technol.* **2**, 60–72. <https://doi.org/10.1007/s42995-019-00003-4> (2020).
22. Feng, X., Wang, Y., Zubin, R. & Wang, F. Core metabolic features and hot origin of Bathyarchaeota. *Engineering* **5**(3), 498–504. <https://doi.org/10.1016/j.eng.2019.01.011> (2019).
23. Wardle, D. A. The influence of biotic interactions on soil biodiversity. *Ecol. Lett.* **9**, 870–886. <https://doi.org/10.1111/j.1461-0248.2006.00931.x> (2006).
24. Blois, J. L., Zarnetske, P. L., Fitzpatrick, M. C. & Finnegan, S. Climate change and the past, present, and future of biotic interactions. *Science* **341**(6145), 499–504. <https://doi.org/10.1126/science.1237184> (2013).
25. Araújo, M. B. & Luoto, M. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecol. Biogeography* **16**(6), 743–753. <https://doi.org/10.1111/j.1466-8238.2007.00359.x> (2007).
26. Chen, J. et al. The structure and species co-occurrence networks of soil denitrifying bacterial communities differ between a coniferous and a broadleaved forests. *Microorganism* **7**(9), 361. <https://doi.org/10.3390/microorganisms7090361> (2019).

27. Silknetter, S. *et al.* Positive biotic interactions in freshwaters: A review and research directive. *Freshwater Biol.* **65**, 811–832. <https://doi.org/10.1111/fwb.13476> (2020).
28. Baselga, A. Partitioning the turnover and nestedness components of beta diversity. *Global Ecol. Biogeography* **19**(1), 134–143. <https://doi.org/10.1111/j.1466-8238.2009.00490.x> (2010).
29. Baselga, A. & Orme, C. D. L. betapart: An R package for the study of beta diversity. *Methods Ecol. Evol.* **3**(5), 808–812. <https://doi.org/10.1111/j.2041-210X.2012.00224.x> (2012).
30. Qian, H., Ricklefs, R. E. & White, P. S. Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecol. Lett.* **8**, 15–22. <https://doi.org/10.1111/j.1461-0248.2004.00682.x> (2005).
31. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **10**, 497–506. <https://doi.org/10.1038/nrmicro2795> (2012).
32. Zhou, J. & Ning, D. Stochastic community assembly: Does it matter in microbial ecology?. *Microbiol. Mol. Biol. Rev.* **81**, e00002-17. <https://doi.org/10.1128/mmb.00002-17> (2017).
33. Sloan, W. T. *et al.* Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* **8**, 732–740. <https://doi.org/10.1111/j.1462-2920.2005.00956.x> (2006).
34. Liao, J. *et al.* The importance of neutral and niche processes for bacterial community assembly differs between habitat generalists and specialists. *FEMS Microbiol. Ecol.* **92**(11), fiw174. <https://doi.org/10.1093/femsec/fiw174> (2016).
35. Ofiteru, I. D. *et al.* Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci.* **107**, 15345–15350. <https://doi.org/10.1073/pnas.100060410> (2010).
36. Roguet, A. *et al.* Neutral community model explains the bacterial community assembly in freshwater lakes. *FEMS Microbiol. Ecol.* **91**(11), fiv125. <https://doi.org/10.1093/femsec/fiv125> (2015).
37. She, Z. X. *et al.* Contrasting prokaryotic and eukaryotic community assembly and species coexistence in acid mine drainage-polluted waters. *Sci. Total Environ.* **856**, 158954. <https://doi.org/10.1016/j.scitotenv.2022.158954> (2023).
38. Chen, W. *et al.* Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. *Microbiome* **7**, 138. <https://doi.org/10.1186/s40168-019-0749-8> (2019).
39. Pan, Y. B., Yang, J., McManus, G. B., Lin, S. J. & Zhang, W. J. Insights into protist diversity and biogeography in intertidal sediments sampled across a range of spatial scales. *Limnol. Oceanography* **65**(5), 1103–1115. <https://doi.org/10.1002/lno.11375> (2020).
40. Stibal, M. *et al.* Different bulk and active bacterial communities in cryoconite from the margin and interior of the Greenland ice sheet. *Environ. Microbiol. Rep.* **7**(2), 293–300. <https://doi.org/10.1111/1758-2229.12246> (2015).
41. Cheung, M. K., Au, C. H., Chu, K. H., Kwan, H. S. & Wong, C. K. Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. *ISME J.* **4**, 1053–1059. <https://doi.org/10.1038/ismej.2010.26> (2010).
42. Long, Y. *et al.* Influences of plant type on bacterial and archaeal communities in constructed wetland treating polluted river water. *Environ. Sci. Pollut. Res.* **23**, 19570–19579. <https://doi.org/10.1007/s11356-016-7166-3> (2016).
43. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**(21), 2957–2963 (2011).
44. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336. <https://doi.org/10.1038/nmeth.f.303> (2010).
45. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. Preprint at <https://doi.org/10.1101/081257> (2016).
46. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. <https://doi.org/10.1093/nar/gks1219> (2013).
47. Wickham, H. Ggplot2: Elegant graphics for data analysis. *Springer Verlag* **2016**, 1–260 (2016).
48. Wang, W. P. *et al.* Seven-year dynamics of testate amoeba communities driven more by stochastic than deterministic processes in two subtropical reservoirs. *Water Res.* **185**, 116232. <https://doi.org/10.1016/j.watres.2020.116232> (2020).
49. Wang, L. J. *et al.* Polycyclic aromatic hydrocarbons in urban soil in the semi-arid city of Xi'an, northwest China: Composition, distribution, sources, and relationships with soil properties. *Arch. Environ. Contam. Toxicol.* **75**, 351–366. <https://doi.org/10.1007/s00244-018-0522-1> (2018).
50. Cabrita, M. J., Aires-De-Sousa, J., Gomes Da Silva, M. D. R., Rei, F. & Costa Freitas, A. M. Multivariate statistical approaches for wine classification based on low molecular weight phenolic compounds. *Austr. J. Grape Wine Res.* **18**(2), 138–146. <https://doi.org/10.1111/j.1755-0238.2012.00182.x> (2012).
51. Hubbell, S. P. *A Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, 2001).
52. Elzhov, T. V., Mullen, K. M., Spiess, A. & Bolker, B. R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK. Plus support for bounds, 1–2 (2010).
53. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks, in International AAAI conference on web and social media. San Jose, California. (2009).

## Acknowledgements

The authors are highly grateful to the editor Dr. Yanhui Dong and two anonymous reviewers for their constructive and valuable comments that have substantially improved the manuscript. This work received financial supports from the following projects: the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB42000000), Laoshan Laboratory (LSKJ202204200), the National Key R&D Program of China (Grant No.2022YFC2803800).

## Author contributions

L.Q.X., L.Y.L. and L.T.G. designed the study and participated in technical discussions; L.T.G. collected the sediment samples; L.Q.X. performed experiment, analyzed dataset and wrote the manuscript. All authors reviewed and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-57214-8>.

**Correspondence** and requests for materials should be addressed to Y.L. or T.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024