Open access • Posted Content • DOI:10.1101/249441

Genetic Diversity and Distributional Pattern of Ammonia Oxidizing Archaea Lineages in the Global Oceans — Source link

Shun Yan Cheung, Wingkwan Mak, Xiaomin Xia, Yanhong Lu ...+1 more authors

Institutions: Hong Kong University of Science and Technology, Zhejiang Ocean University

Published on: 17 Jan 2018 - bioRxiv (Cold Spring Harbor Laboratory)

Genetic Diversity and Distributional Pattern of Ammonia Oxidizing Archaea Lineages in the Global Oceans

Shunyan Cheung[1], Wingkwan Mak[1,2], Xiaomin Xia[1,3], Yanhong Lu[1], Hongbin Liu[1]*

[1] Division of Life Science, The Hong Kong University of Science and Technology, Hong Kong, China.

[2] Department of Ocean Sciences, University of California, Santa Cruz, CA 95064, USA.

[3] Zhejiang Ocean University, Zhoushan, Zhejiang, China.

*Corresponding author

E-mail: liuhb@ust.hk

Summary

In the study, we used miTAG approach to analyse the distributional pattern of the ammonium oxidizing archaea (AOA) lineages in the global oceans using the metagenomics datasets of the *Tara* Oceans global expedition (2009-2013). Using ammonium monooxygenase alpha subunit gene as biomarker, the AOA communities were obviously segregated with water depth, except the upwelling regions. Besides, the AOA communities in the euphotic zones are more heterogeneous than in the mesopelagic zones (MPZs). Overall, water column A clade (WCA) distributes more evenly and widely in the euphotic zone and MPZs, while water column B clade (WCB) and SCM-like clade mainly distribute in MPZ and high latitude waters, respectively. At fine-scale genetic diversity, SCM1-like and 2 WCA subclades showed distinctive niche separation of distributional pattern. The AOA subclades were further divided into ecological significant taxonomic units (ESTUs), which were delineated from the distribution pattern of their corresponding subclades. For examples, ESTUs of WCA have different correlation with depth, nitrate to silicate ratio and salinity; SCM1-like-A was negatively correlated with irradiation; the other SCM-like ESTUs preferred low temperature and high nutrient conditions, etc. Our study provides new insight to the genetic diversity of AOA in global scale and its connections with environmental factors.

Introduction

Nitrification is an important and well-concerned pathway in the marine nitrogen cycle, which can supply approximately 25-36 % of bioavailable nitrogen to phytoplankton (Wankel et al., 2007; Yool et al., 2007). Besides that, it was also estimated as dominant source of nitrous oxide (Freing et al., 2012), a strong greenhouse gas influencing the Earth climate (Solomon, 2007). As the first and rate-limiting step of nitrification, ammonia oxidation was formerly only discovered within bacterial domain (AOB) (Ward, 2002). However, unique archaeal-associated ammonia monooxygenase gene recovered from environmental metagenomics data over a decade ago, suggested the metabolic potential of archaeal ammonia oxidation (Venter et al., 2004; Treusch et al., 2005). Using ammonia monooxygenase alpha subunit (*amoA*) as a biomarker, archaeal *amoA* genes were detected throughout water column and marine sediments (Francis et al., 2005). With the first isolation of marine ammonia oxidizing archaea (AOA) (*Nitrosopumilis maritimus* SCM-1), AOA are known to be chemo-lithoautotrophor, aerobically oxidizing ammonia to nitrite while fixing inorganic carbon (Könneke et al., 2005). Subsequent field studies revealed AOA outnumber AOB in the oceanic waters (Wuchter et al., 2006; Mincer et al., 2007; Beman et al., 2008; Beman et al., 2010; Santoro et al., 2010), and they play a dominant role in determining the activity and distribution of nitrification process (Wuchter et al., 2006; Beman et al., 2008; Smith et al., 2014). In contrast to AOB, AOA (*N. maritimus* SCM-1) exhibited exceptionally high ammonia affinity and low thresholds, explaining why AOA are dominant in the oceanic waters (Martens-Habbena et al., 2009).

In the first phylogenetic analysis of AOA *amoA* gene, two groups of AOA were commonly detected in the water column, and they were named water column A (WCA) and water column B (WCB), respectively (Francis et al., 2005). WCA and WCB are considered as the major ecotypes of AOA in the oceanic waters, predominating in upper (<

200m) and deep (> 200m) layers, respectively (Francis et al., 2005; Beman et al., 2008; Santoro et al., 2017). So far, there is only one culture of WCA (*Nitrosopelagicus brevis* CN25) reported (Santoro and Casciotti, 2011; Santoro et al., 2015) and the WCB is still uncultivated, therefore, the physiology of these two ecotypes are still unclear (Santoro et al., 2017). In spite of that, several factors have been hypothesized as potentially causing the vertical succession of WCA and WCB, including light inhibition, differential affinity and utilization of nitrogen sources, and different requirement of micro-nutrients (Amin et al., 2013; Sintes et al., 2013; Luo et al., 2014; Qin et al., 2014; Sintes et al., 2016; Smith et al., 2016; Santoro et al., 2017).

With development of DNA sequencing technique, our knowledge of the genetic diversity of AOA is improving (Sintes et al., 2016; Jing et al., 2017). Using next generation sequencing to samples from western subarctic Pacific Ocean, diverse subclades of ecotype WCA and WCB were defined; and some subclades of same ecotypes showed differential distributional patterns (Jing et al., 2017). In addition to that, recently global scale study has found that the ecological significant taxonomic units (ESTUs) of *Synechococcus* were delineated from their corresponding ecotypes, in which OTUs of same ecotypes were grouped according to their distributional patterns (Farrant et al., 2016). Therefore, although most of the previous studies were targeting the whole clade of WCA and WCB (Francis et al., 2005; Beman et al., 2008; Santoro et al., 2010; Smith et al., 2014; Smith et al., 2016; Santoro et al., 2017), the subclades and ESTUs of WCA and WCB may show distinctive distributional patterns in the oceans. Moreover, analysing the subclades and ESTUs of AOA in diverse marine ecosystems may provide better and more insights on the specific connection between genetic diversity of AOA and environmental conditions.

In this study, we conducted detailed phylogenetic analysis of AOA in the global oceans using metagenomics datasets of *Tara* Oceans expedition (2009-2013) (Karsenti et al., 2011; Armbrust and Palumbi, 2015). We have built a database of *amoA* gene sequences and recovered the AOA communities from the metagenomics datasets using miTAG approach (Logares et al., 2014). miTAG is an approach mapping short reads of metagenomics dataset to long sequences of known identity, which has been proven as a promising and PCR amplification biases free method for recovering community of bacteria and picocyanobacteria from the *Tara* Ocean datasets (Logares et al., 2014; Farrant et al., 2016). Furthermore, we have further divided the subclades of AOA into ecological significant taxonomic units (ESTUs), in which OTUs of same subclades were grouped according to their distributional patterns (Farrant et al., 2016). Through all these analyses, we described the distributional patterns of AOA and characterized the AOA ESTUs with environmental conditions.

Methods

Constructing database of *amoA* gene

In order to retrieve and identify the *amoA* affiliated sequences (AARs) in the *Tara* Oceans metagenomics dataset (Karsenti et al., 2011; Armbrust and Palumbi, 2015), the reference sequences and environmental sequences of *amoA* gene were downloaded from National Centre for Biotechnology Information (NCBI) nucleotide database (Acland et al., 2014). These sequences were aligned with MUSCLE (Edgar, 2004) and the majority of sequences have overlapping at the region targeted by the Arch-amoAF/Arch-amoAR primer set (Francis et al., 2005). The sequences that have overlapping in this region were chosen and clipped into an alignment with the same length, and then were used to calculated OTUs at 97 % DNA similarity cut off using Mothur (Schloss et al., 2009).

Besides that, amino acid sequences of these selected sequences were also downloaded from the NCBI (Acland et al., 2014).

Recruiting AARs from the *Tara* Oceans dataset

The clean dataset of *Tara* Ocean metagenomics dataset (processed nucleotide reads, 100-200 bases per reads) were downloaded from the European Bioinformatics Institute (EMBL-EMI) (https://www.ebi.ac.uk/metagenomics/projects/ERP001736) and aligned against the amino acid sequences of our *amoA* gene database using DIAMOND (e-value = $1e^{-5}$) (Buchfink et al., 2015). DIAMOND is a BLASTX like algorithm 2000 times faster than BLASTX on short reads (Buchfink et al., 2015), which reduces the demand of computational resources when recruiting the *amoA* from large metagenomics dataset of the *Tara* Oceans. The sequences recruited with DIAMOND were then blasted against the DNA sequences of our *amoA* database using BLASTN (McGinnis and Madden, 2004). The reads which have 99-100 % similarity and coverage with the DNA sequences in our *amoA* database were chosen as AARs. There was less than 1 % of reads showed lower than 99 % similarity with the sequences in our dataset, which were neglected in downstream analysis. AARs were recruited from 35 bacterial size fraction metagenomics datasets that contained significant number of AARs. Except 3 datasets, which contains 35, 43 and 51 AARs, the rest datasets contained more than 100 AARs each (111-10,604 reads) (S 1 Table). These datasets contain 22 and 13 samples from euphotic zone and MPZ, covering Pacific Ocean, Atlantic Ocean, Indian Ocean, Southern Ocean, Red Sea, Mediterranean Sea and Gulf of Mexico (Fig. 1).

Phylogenetic analysis

The number of AARs which affiliated to the DNA sequences of the same OTUs were merged into the sequence number of each OTUs; and the relative abundance of each OTU

141 in each sample were calculated by dividing sequence number of the OTU in particular
142 sample by total clear reads of that sample. Because the number of AARs in different
143 samples are highly varied, ordering of the OTUs were normalized by their relative
144 abundance, instead of sequence numbers. Top 202 OTUs were selected for downstream
145 analysis, which accounts for more than 90 % of the whole dataset.

146 To identify the top 202 OTUs, DNA sequences of these OTUs, species of AOA and
147 previously defined subclades of AOA (Jing et al., 2017) were codon aligned and used to
148 construct a Maximum-Likelihood (ML) phylogenetic tree with MEGA 7.0 (Kumar et al.,
149 2016). According to the result of model test (Bayesian Information Criterion calculation),
150 the Tamura 3-parameter model, using discrete Gamma distribution with assumption that a
151 certain portion of sites are evolutionarily invariables (T92+G+I), was used to construct the
152 ML phylogenetic tree. The tree was further edited with iTOL (Letunic and Bork, 2016),
153 with the relative abundances of top 202 OTUs displayed. The OTUs fell into the same
154 subclades were further hierarchically clustered into ESTUs using SIMPROF test
155 (method.distance=Euclidean) of the package "clustsig" in R (Whitaker and Christman,
156 2010), in which the OTUs having significantly similar distributional patterns (P value <
157 0.05) were grouped into same clusters (i.e. ESTUs).

158 **Statistical analyses**

159 Non-metric multidimensional scaling (NMDS) plots were plotted using Primer 5
160 (Clarke and Gorley, 2001), to visualise the relationship between the AOA communities
161 recovered in the 35 metagenomics datasets, based on the Bray-Curtis dissimilarity matrix.
162 To characterize the ESTUs with environmental factors, correlation between the relative
163 abundances of the ESTUs and environmental factors were analysed with Pearson test using
164 package "Hmisc" in R; and the correlation plots were generated with the package "Corrplot"

in R (Wei and Simko, 2013). The environmental factors were downloaded from the www.pangaea.de/.

Results and Discussions

High diversity of AOA in the metagenomics datasets

The top 202 OTUs were all affiliated to WCA, WCB and SCM-like clades, except two unclassified OTUs (OTU 53 and 154) (Fig. 2). Using the sequences of previously defined subclades of WCA and WCB (Jing et al., 2017) as references, all the subclades were detected in the *Tara* Oceans metagenomics datasets. In addition to that, new subclades of WCA (WCAIV) and WCB (WCAV) were identified (Fig. 2). The previously defined WCBII (Jing et al., 2017) was further divided into WCBII and WCBIII; and WCBIII in previous study (Jing et al., 2017) was renamed as WCBIV in this study (Fig. 2). Although there were 10 subclades of AOA commonly detected in the oceans (Fig. 2), only WCAI and SCM1-like has their representative cultures, *N. brevis* CN 25 (Santoro et al., 2015) and *N. maritimus* SCM-1 (Könneke et al., 2005) (Fig. 2). Based on the result of SIMPROF test, the OTUs of most subclades were grouped into ESTUs, except WCAIV, WCBII and WCBV, in which the OTUs showed homogenous distributional patterns (Fig. 2). Besides that, the genetically similar OTUs were not necessary to be included in the same ESTUs. For examples, within WCAI, the OTUs within WCAI-C were genetically similar, while the OTUs of WCAI-B were not closely clustering together in the phylogenetic tree (Fig. 2).

Considering the high genetic diversity of WCA and WCB and that the previous studies quantified the whole group of WCA and WCB (Mosier and Francis, 2011; Smith et al., 2014; Smith et al., 2016; Jing et al., 2017; Santoro et al., 2017), the sequences of qPCR primer sets of WCA and WCB (Mosier and Francis, 2011) were compared with

representative sequences of the top OTUs of all the subclades (S2 Table). The result showed that both primer sets were well designed and can cover all the subclades, except WCAII, which has two bases mismatched with the WCA-amoA-P (Taqman probe). It has been reported that mismatch between primer set and template may influence the efficiency of amplification (Liu et al., 2006; Stadhouders et al., 2010). Therefore, the current primer set of WCA (Mosier and Francis, 2011) may underestimate the abundance of WCA, when WCAII predominates the AOA community. The actual influence of the mismatches to detection of WCAII is waiting for validation in future study.

Global distributional pattern of AOA clades / subclades

The AOA communities in the euphotic zones were predominated by WCA and SCM1-like, while the communities in MPZs were mainly predominated by WCB (Fig. 3). The general vertical distributional pattern of WCA and WCB in this study agreed with the previous findings (Beman et al., 2008; Sintes et al., 2013; Sintes et al., 2016; Jing et al., 2017; Santoro et al., 2017). At the level of subclade, WCAI was the most dominant subclade in the euphotic layers of open oceans, followed by WCAII and WCAIII (Fig. 3). However, in the Gulf of Mexico and Red sea, WCAII became more dominant than WCAI. Previous study in the western North Pacific Ocean suggested that WCAII was specific to the Western Pacific Ocean (Jing et al., 2017), because the WCAII affiliated sequences in the NCBI nucleotide database were all recovered from the East China Sea (Hu et al., 2011). Our result suggested that WCAII is indeed globally distributed, however, it is predominant in the ecosystems of marginal seas. SCM1-like was another major member of the AOA community in the euphotic zone, which mainly predominated the waters above the latitude of 35˚ in both hemispheres. This agreed with previous study that *N. maritimus* SCM-1 affiliated sequences were detected in Arctic Ocean and Antarctic coastal waters (Kalanetra et al., 2009). Our results suggested

niche separation of WCAI, WCAII and SCM1-like in the euphotic zones of the oceans, which are predominant in open oceans, marginal seas and high latitude waters, respectively.

In the MPZs, WCBIV is the most dominant subclades, followed by WCBI, WCBII and WCBIII. However, it should be noticed that WCA and SCM1-like was also detected in the MPZs. Although the relative abundances of these two clades in the MPZ AOA communities were low (Fig. 3), their significant relative abundances in the MPZ metagenomics datasets suggested that they are more evenly and widely distributed than WCB (Fig. 2). The vertical profiles of WCA abundance have been analysed in several studies (Sintes et al., 2016; Smith et al., 2016; Jing et al., 2017; Santoro et al., 2017), which showed different results. In the equatorial Pacific and the northeast Pacific Ocean, the abundance of WCA decreased dramatically in the upper MPZ (near zero) (Santoro et al., 2017) and lower MPZ (10^2 gene copy per liter) (Smith et al., 2016), respectively. However, at some stations in Atlantic Ocean and northwest Pacific Ocean, the abundances of WCA remained significant in the MPZs (10^3 – 10^4 gene copies per liter) and were comparable with that in the euphotic zones (Sintes et al., 2016; Jing et al., 2017). Together with our results (Fig. 2), it is not necessary for the distribution of WCA to be restricted in the euphotic zone and upper MPZs, and the significance of WCA and SCM1-like in the MPZs and even deeper water (Jing et al., 2017) is worthy for further exploration.

Vertical segregation of AOA communities

At 7 geographic locations, AOA communities were recovered in both euphotic zones and MPZs (Fig. 4). The communities in the euphotic layers (IO6, IO7, SAO3, SAO6, SAO9, SPO6 and SO1) were more heterogeneous than those in the MPZs (IO5, IO8, SAO2, SAO7, SAO10, SPO5 and SO2) (Fig. 4). Especially in the Southern Ocean, the community in the euphotic zone (SO1) was distant from all other AOA communities in the dataset, while community in the MPZ (SO2) still clustered closely with other MPZ communities (Fig. 4).

Similar patterns of bacteria and archaea communities have been reported in Pacific Ocean, in which the communities in surface waters are more heterogeneous than in the deep water (Jing et al., 2013; Xia et al., 2017). It was explained that the communities are less stable under the fluctuating physiochemical factors in the surface waters (Jing et al., 2013; Bryant et al., 2016). Except NAO5, which was sampled in upper MPZ (246 m in depth), all the MPZs datasets were originated from the samples collected in lower MPZs (> 500 m in depth) (S1 Table). Although the AOA communities were located in the lower MPZs in different oceans, they were all clustered together closely (Fig. 4). This is because the relative abundances of WCB subclades in the AOA communities were similar among the lower MPZs of different Oceans (Fig. 3), which can be explained by that the relatively stable environmental conditions in deep waters select the similar community of prokaryotes (Jing et al., 2013; Xia et al., 2017) and support the theory that "everything is in everywhere, but, the environment selects" (Baas-Becking, 1934).

The vertical segregation of AOA communities was disrupted in the upwelling regions of northeast tropical Pacific Ocean (NPO3), Arabian Sea (IO1) and Equatorial Pacific Ocean (EPO) (Fig. 3, 4) (Wyrtki, 1967; Ulloa et al., 2012), where WCB were dominant in the euphotic layers (Fig. 3). Significant abundance of WCB was also detected in the shallow water of upwelling regions in Monterey Bay (Smith et al., 2014) and southeast Pacific (Molina et al., 2010). Our result showed that the vertical segregation of AOA communities is stable and predictable in the oceans, and physical force that transporting WCB to euphotic layers in upwelling region is the main factor of disrupting the segregation.

Characterizing the ESTUs of AOA with environmental factors

For finer resolution of the connection between AOA genetic diversity and environmental factors, instead of subclades, the correlation between ESTUs of different subclades and environmental factors were examined using Pearson test. As the result, the

263 ESTUs within same subclades could be correlated to different environmental factors (Fig. 5).
264 Within WCAI subclade, three dominant ESTUs (WCAI-A, WCAI-B and WCAI-C) showed
265 different correlations with depth. The WCAI-A prefers shallower water with low
266 concentrations of organic nutrients WCAI-C prefers deeper waters with low temperature,
267 while WCAI-B has no significant correlation with depth (Fig. 5). This agreed with the
268 distribution patterns of these ESTUs that the relative abundance of OTU1 (WCAI-A) was
269 higher in the euphotic zone, OTU46 of WCAI-B did not show obvious distributional pattern
270 and OTU63 of WCAI-C was mainly distributed in the MPZs (Fig. 2). In previous studies
271 which quantified the whole group of WCA, WCA was most abundant in the euphotic zone
272 and remained significant abundance in the MPZs (Sintes et al., 2016; Jing et al., 2017). It can
273 now be explained that the WCAI-A (top ESTUs of WCA) is responsible for the high
274 abundance of WCA in the euphotic zone, while WCA in MPZ is mainly contributed by
275 WCAI-C. The vertical succession between WCA and WCB along the water column has been
276 well documented (Beman et al., 2008; Sintes et al., 2016; Smith et al., 2016; Jing et al., 2017;
277 Santoro et al., 2017). In addition to that, our result suggested that vertical succession is also
278 existing among the ESTUs of WCA. The ESTUs of WCAII have positive correlations with
279 salinity and temperature (Fig. 5), because they were predominant in the low latitude marginal
280 seas (Fig, 3), where the salinity and temperature were high (S1 Table). Similar to WCAI-C,
281 WCAIII-A and WCAIII-C were also positively correlated with water depth. Besides the
282 physical factors, a number of ESTUs of WCAI and WCAIII were positively correlated with
283 nitrate : silicate ratio, which agreed with the recent finding that the abundance of WCA was
284 positively correlated with this ratio in Equatorial Pacific upwelling region (Santoro et al.,
285 2017). The nitrate : silicate ratio was treated as indicator of remineralization in their studied
286 region (Raimbault et al., 1999; Jiang et al., 2003; Buesseler et al., 2008) , and the positive
287 correlation implied that WCA is a major player of nitrification and the trace metal and

ammonium released during remineralization process are important requirement of WCA (Santoro et al., 2017). Our result showed significant and positive correlation of WCAI and WCAIII with nitrate : silicate ratio in the larger geographic scale. However, it should be noticed that the nitrate : silicate ratio may have other implications, which are depended on phytoplankton community structure and the geographic locations (Koike et al., 2001; Bibby and Moore, 2011). Therefore, further explorations are needed for verifying the relationship between WCA (or even WCA subclades), nitrate : silicate ratio and remineralization of organic matter.

The SCM1-like-A was negatively correlated with irradiation, while the rest ESTUs of SCM1-like clade did not (Fig. 5). SCM1-like-B, D and E were negatively correlated with temperature, agreeing with the general pattern that SCM1-like clade prefers high latitude waters (Fig. 3). Besides that, similar to WCAI-C and WCAIII, SCM1-like-C has a strong and positive correlation with depth (Fig. 5), which explains their main distribution in the MPZs (Fig. 2). Culture based study has discovered that different marine strains of *Nitrosopumilus* (SCM-1, PS0 and HCA1) had different sensitivity to photoinhibition (Qin et al., 2014), which supports our result about SCM1-like-A. Moreover, the positive correlations of SCM1-like-B and D with coloured dissolve organic matter (Fig. 5) also agreed with previous finding that some strains of *Nitrosopumilus* were obligate mixotrophs (Qin et al., 2014). The similar results of culture based study and our field study suggested that, the correlations between AOA ESTUs and environmental factors provides useful knowledge about the ecophysiology of AOA, especially the uncultured subclades.

WCB was rare and nearly undetected in the euphotic zone (Fig. 2), therefore, only the samples from lower MPZs (594 – 989 m in depth) were included in the correlation analysis. Some ESTUs of WCBI and IV showed preferences to deeper waters (Fig. 5). The WCBIV-K, L and Q showed similar pattern of correlation with salinity, temperature and macro-nutrients,

313 which is because they were only found in rare number in the MPZ of the Southern Ocean
314 (SO2), where was low in temperature and high in macronutrients (S1 table). Besides that,
315 WCBII and WCIV-H showed positive correlations with nitrate : silicate ratio. Because the
316 environmental condition in the MPZs is less variable, the ESTUs of WCB in the MPZs did
317 not show significant correlations with most of the environmental factors (Fig. 5). Moreover,
318 since the number of datasets from MPZs are limited, the correlation results of WCB may not
319 be as significant as that of WCA and SCM1-like.

320 Based on our analysis, both subclades of the same clades and ESTUs of the same
321 subclades could have different distributional patterns and correlation with different
322 environmental factors (Fig. 2; 3 and 5), indicating the micro-diversity of AOA in the oceans
323 should not be overlooked. In addition to studying the distribution of the whole group of WCA
324 or WCB (Sintes et al., 2016; Smith et al., 2016; Jing et al., 2017; Santoro et al., 2017),
325 analysing micro-diversity of WCA and WCB in the global scale refined our understanding to
326 the ecophysiology of uncultivated AOA groups in marine ecosystems. Recent high-resolution
327 genetic analysis of *Synechococcus* and unicellular cyanobacteria diazotroph (UCYN-A) also
328 showed that the sublineages or ESTUs have different distributional patterns due to their
329 different preference and sensitivity to certain environmental factors, causing niche separation
330 in the oceans (Farrant et al., 2016; Cheung et al., 2017; Turk-Kubo et al., 2017).

331 Our study provided the first insight to the micro-diversity of AOA in the global ocean,
332 and demonstrated the importance of high resolution genetic analysis of previous defined
333 ecotypes. In addition to large geographic scale, high resolution studies in mesoscale regions
334 with steep environmental gradients may provide more insights to the correlations between
335 environmental factors and microorganisms (Robidart et al., 2014; Cheung et al., 2017). In
336 addition to macro-nutrients, trace elements have also been suggested to influence the
337 distribution of WCA and WCB (Santoro et al., 2017). For better understanding of the eco-

338 physiology of uncultivated marine AOA, the abundance and high resolution distributional
339 patterns of AOA subclades can be analysed using subclade specific qPCR primer sets in
340 future studies.

341 ## Acknowledgements

342 We thank the *Tara* Oceans for the huge open accessed metagenomics datasets of global
343 oceans. This work was funded by the National Key Scientific Research Project
344 (2015CB954003) sponsored by the Ministry of Science and Technology of the PRC and RGC
345 GRF grant 16101917.

346

References

Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C. et al. (2014) Database resources of the national center for biotechnology information. *Nucleic acids res* **42**: D7.

Amin, S.A., Moffett, J.W., Martens-Habbena, W., Jacquot, J.E., Han, Y., Devol, A. et al. (2013) Copper requirements of the ammonia-oxidizing archaeon *Nitrosopumilus maritimus* SCM1 and implications for nitrification in the marine environment. *Limnol Oceanogr* **58**: 2037-2045.

Armbrust, E.V., and Palumbi, S.R. (2015) Uncovering hidden worlds of ocean biodiversity. *Science* **348**: 865-867.

Baas-Becking, L.G.M. (1934) *Geobiologie; of inleiding tot de milieukunde*: WP Van Stockum & Zoon NV.

Beman, J.M., Popp, B.N., and Francis, C.A. (2008) Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. *ISME J* **2**: 429-441.

Beman, J.M., Sachdeva, R., and Fuhrman, J.A. (2010) Population ecology of nitrifying Archaea and Bacteria in the Southern California Bight. *Environ microbiol* **12**: 1282-1292.

Bibby, T., and Moore, C. (2011) Silicate: nitrate ratios of upwelled waters control the phytoplankton community sustained by mesoscale eddies in sub-tropical North Atlantic and Pacific. *Biogeosciences* **8**: 657-666.

Bryant, J.A., Aylward, F.O., Eppley, J.M., Karl, D.M., Church, M.J., and DeLong, E.F. (2016) Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. *ISME J* **10**: 1308-1322.

Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat methods* **12**: 59-60.

372 Buesseler, K.O., Trull, T.W., Steinberg, D.K., Silver, M.W., Siegel, D.A., Saitoh, S.-I. et al.
373 (2008) VERTIGO (VERtical Transport In the Global Ocean): a study of particle sources and
374 flux attenuation in the North Pacific. *Deep Sea Res Part 2 Tol Stud Oceanogr* **55**: 1522-
375 1539.
376 Cheung, S., Suzuki, K., Saito, H., Umezawa, Y., Xia, X., and Liu, H. (2017) Highly
377 heterogeneous diazotroph communities in the Kuroshio Current and the Tokara Strait, Japan.
378 *PloS one* **12**: e0186875.
379 Clarke, K., and Gorley, R. (2001) User manual tutorial PRIMER 5.0. In: Primer.
380 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
381 throughput. *Nucleic acids res* **32**: 1792-1797.
382 Farrant, G.K., Doré, H., Cornejo-Castillo, F.M., Partensky, F., Ratin, M., Ostrowski, M. et al.
383 (2016) Delineating ecologically significant taxonomic units from global patterns of marine
384 picocyanobacteria. *Proc Natl Acad Sci U S A* **113**: E3365-E3374.
385 Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005) Ubiquity
386 and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean.
387 *Proc Natl Acad Sci U S A* **102**: 14683-14688.
388 Freing, A., Wallace, D.W., and Bange, H.W. (2012) Global oceanic production of nitrous
389 oxide. *Philos Tran R Soc Lond B Biol Sci* **367**: 1245-1255.
390 Hu, A., Jiao, N., Zhang, R., and Yang, Z. (2011) Niche partitioning of marine group I
391 Crenarchaeota in the euphotic and upper mesopelagic zones of the East China Sea. *Appl*
392 *environ microbiol* **77**: 7469-7478.
393 Jiang, M.-S., Chai, F., Dugdale, R., Wilkerson, F., Peng, T.-H., and Barber, R. (2003) A
394 nitrate and silicate budget in the equatorial Pacific Ocean: a coupled physical–biological
395 model study. *Deep Sea Res Part 2 Top Stud Oceanogr* **50**: 2971-2996.

396 Jing, H., Xia, X., Suzuki, K., and Liu, H. (2013) Vertical profiles of bacteria in the tropical
397 and subarctic oceans revealed by pyrosequencing. *PloS one* **8**: e79423.
398 Jing, H., Cheung, S., Xia, X., Suzuki, K., Nishioka, J., and Liu, H. (2017) Geographic
399 Distribution of Ammonia-Oxidizing Archaea along the Kuril Islands in the Western Subarctic
400 Pacific. *Front in microbiol* **8**: 1247.
401 Kalanetra, K.M., Bano, N., and Hollibaugh, J.T. (2009) Ammonia-oxidizing Archaea in the
402 Arctic Ocean and Antarctic coastal waters. *Environ microbiol* **11**: 2434-2445.
403 Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J. et al. (2011) A
404 holistic approach to marine eco-systems biology. *PLoS biol* **9**: e1001177.
405 Koike, I., Ogawa, H., Nagata, T., Fukuda, R., and Fukuda, H. (2001) Silicate to nitrate ratio
406 of the upper sub-arctic Pacific and the Bering Sea Basin in summer: its implication for
407 phytoplankton dynamics. *J Oceanogr* **57**: 253-260.
408 Könneke, M., Bernhard, A.E., José, R., Walker, C.B., Waterbury, J.B., and Stahl, D.A.
409 (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543-
410 546.
411 Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics
412 Analysis version 7.0 for bigger datasets. *Mol biol evol* **33**: 1870-1874.
413 Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the
414 display and annotation of phylogenetic and other trees. *Nucleic acids res* **44**: W242-W245.
415 Liu, H., Wang, H., Shi, Z., Wang, H., Yang, C., Silke, S. et al. (2006) TaqMan probe array
416 for quantitative detection of DNA targets. *Nucleic Acids Res* **34**: e4-e4.
417 Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F.M., Ferrera, I., Sarmento, H. et al.
418 (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon
419 sequencing to explore diversity and structure of microbial communities. *Environ microbiol*
420 **16**: 2659-2671.

421 Luo, H., Tolar, B.B., Swan, B.K., Zhang, C.L., Stepanauskas, R., Moran, M.A., and
422 Hollibaugh, J.T. (2014) Single-cell genomics shedding light on marine Thaumarchaeota
423 diversification. *ISME J* **8**: 732-736.
424 Martens-Habbena, W., Berube, P.M., Urakawa, H., José, R., and Stahl, D.A. (2009)
425 Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria.
426 *Nature* **461**: 976-979.
427 McGinnis, S., and Madden, T.L. (2004) BLAST: at the core of a powerful and diverse set of
428 sequence analysis tools. *Nucleic acids res* **32**: W20-W25.
429 Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., and DeLong, E.F. (2007)
430 Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and
431 the North Pacific Subtropical Gyre. *Environ microbiol* **9**: 1162-1175.
432 Molina, V., Belmar, L., and Ulloa, O. (2010) High diversity of ammonia-oxidizing archaea in
433 permanent and seasonal oxygen-deficient waters of the eastern South Pacific. *Environ*
434 *microbiol* **12**: 2450-2465.
435 Mosier, A.C., and Francis, C.A. (2011) 9 Determining the Distribution of Marine and Coastal
436 Ammonia-Oxidizing Archaea and Bacteria Using a Quantitative Approach. *Methods enzymol*
437 **486**: 205.
438 Qin, W., Amin, S.A., Martens-Habbena, W., Walker, C.B., Urakawa, H., Devol, A.H. et al.
439 (2014) Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide
440 ecotypic variation. *Proc Natl Acad Sci U S A* **111**: 12504-12509.
441 Raimbault, P., Slawyk, G., Boudjellal, B., Coatanoan, C., Conan, P., Coste, B. et al. (1999)
442 Carbon and nitrogen uptake and export in the equatorial Pacific at 150 W: Evidence of an
443 efficient regenerated production cycle. *J Geophys Res Oceans* **104**: 3341-3356.

444 Robidart, J.C., Church, M.J., Ryan, J.P., Ascani, F., Wilson, S.T., Bombar, D. et al. (2014)
445 Ecogenomic sensor reveals controls on N_2-fixing microorganisms in the North Pacific Ocean.
446 *ISME J* **8**: 1175-1185.
447 Santoro, A.E., and Casciotti, K.L. (2011) Enrichment and characterization of ammonia-
448 oxidizing archaea from the open ocean: phylogeny, physiology and stable isotope
449 fractionation. *ISME J* **5**: 1796-1808.
450 Santoro, A.E., Casciotti, K.L., and Francis, C.A. (2010) Activity, abundance and diversity of
451 nitrifying archaea and bacteria in the central California Current. *Environ Microbiol* **12**: 1989-
452 2006.
453 Santoro, A.E., Saito, M.A., Goepfert, T.J., Lamborg, C.H., Dupont, C.L., and DiTullio, G.R.
454 (2017) Thaumarchaeal ecotype distributions across the equatorial Pacific Ocean and their
455 potential roles in nitrification and sinking flux attenuation. *Limnol Oceanogr* doi:
456 10.1002/lno.10547.
457 Santoro, A.E., Dupont, C.L., Richter, R.A., Craig, M.T., Carini, P., McIlvin, M.R. et al.
458 (2015) Genomic and proteomic characterization of "*Candidatus Nitrosopelagicus brevis*": An
459 ammonia-oxidizing archaeon from the open ocean. *Proceedings of the National Academy of*
460 *Sciences* **112**: 1173-1178.
461 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al.
462 (2009) Introducing mothur: open-source, platform-independent, community-supported
463 software for describing and comparing microbial communities. *Appl environ microbiol* **75**:
464 7537-7541.
465 Sintes, E., De Corte, D., Haberleitner, E., and Herndl, G.J. (2016) Geographic distribution of
466 archaeal ammonia oxidizing ecotypes in the Atlantic Ocean. *Front microbiol* **7**: 77.

467 Sintes, E., Bergauer, K., De Corte, D., Yokokawa, T., and Herndl, G.J. (2013) Archaeal
468 *amoA* gene diversity points to distinct biogeography of ammonia-oxidizing Crenarchaeota in
469 the ocean. *Environ microbiol* **15**: 1647-1658.
470 Smith, J.M., Casciotti, K.L., Chavez, F.P., and Francis, C.A. (2014) Differential contributions
471 of archaeal ammonia oxidizer ecotypes to nitrification in coastal surface waters. *ISME J* **8**:
472 1704-1714.
473 Smith, J.M., Damashek, J., Chavez, F.P., and Francis, C.A. (2016) Factors influencing
474 nitrification rates and the abundance and transcriptional activity of ammonia-oxidizing
475 microorganisms in the dark northeast Pacific Ocean. *Limnol Oceanogr* **61**: 596-609.
476 Solomon, S. (2007) *Climate change 2007-the physical science basis: Working group I*
477 *contribution to the fourth assessment report of the IPCC*: Cambridge University Press.
478 Stadhouders, R., Pas, S.D., Anber, J., Voermans, J., Mes, T.H., and Schutten, M. (2010) The
479 effect of primer-template mismatches on the detection and quantification of nucleic acids
480 using the 5′ nuclease assay. *J Mol Diagn* **12**: 109-117.

481 Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.P., and Schleper, C.
482 (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of
483 uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ microbiol* **7**: 1985-1995.
484 Turk-Kubo, K.A., Farnelid, H.M., Shilova, I.N., Henke, B., and Zehr, J.P. (2017) Distinct
485 ecological niches of marine symbiotic N2-fixing cyanobacterium *Candidatus*
486 *Atelocyanobacterium thalassa* sublineages. *J phycol* **53**: 451-461.
487 Ulloa, O., Canfield, D.E., DeLong, E.F., Letelier, R.M., and Stewart, F.J. (2012) Microbial
488 oceanography of anoxic oxygen minimum zones. *Proc Natl Acad Sci U S A* **109**: 15996-
489 16003.

490 Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A. et al.
491 (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.
492 Wankel, S.D., Kendall, C., Pennington, J.T., Chavez, F.P., and Paytan, A. (2007)
493 Nitrification in the euphotic zone as evidenced by nitrate dual isotopic composition:
494 Observations from Monterey Bay, California. *Global Biogeochem Cycles* **21**: GB2009.
495 Ward, B. (2002) Nitrification in aquatic systems. *Encyclopedia of Environmental*
496 *Microbiology*.
497 Wei, T., and Simko, V. (2013) corrplot: Visualization of a correlation matrix. *R package*
498 *version 073* **230**: 11.
499 Whitaker, D., and Christman, M. (2010) clustsig: Significant Cluster Analysis. R package
500 version 1.0. In.
501 Wuchter, C., Abbas, B., Coolen, M.J., Herfort, L., van Bleijswijk, J., Timmers, P. et al.
502 (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci* **103**: 12317-12322.
503 Wyrtki, K. (1967) Equatorial Pacific Ocean1. *Int J Oceanol & Limnol Vol* **1**: 117-147.
504 Xia, X., Guo, W., and Liu, H. (2017) Basin scale variation on the composition and diversity
505 of Archaea in the Pacific Ocean. *Front Microbiol* **8**: 2057.
506 Yool, A., Martin, A.P., Fernández, C., and Clark, D.R. (2007) The significance of
507 nitrification for oceanic new production. *Nature* **447**: 999-1002.
508
509

510 **Legends**

511 Fig. 1. Sampling locations of metagenomics datasets selected in this study. The relative
512 abundances of AOA (AARs divided by total clean reads) in the metagenomics datasets were
513 shown with color gradient.

514

515 Fig.2. Maximum likelihood phylogenetic tree constructed using DNA sequences of top 202
516 OTUs, reference sequences of marine AOA cultures and previous defined subclades of WCA
517 and WCB. The bootstraps test was conducted for 1000 times, and the values which higher
518 than 60 % were displayed as black circles. The OTU abundances referred to that OTU
519 affiliated reads divided by total clean reads in particular metagenomics datasets. The OTUs
520 were labelled with their corresponding ESTUs (e.g. WCAI-A, WCAI-B, etc).

521

522 Fig. 3. Distributional pattern of AOA communities in the global oceans. The community
523 structures were consisted of the relative abundances of top 202 OTUs.

524

525 Fig. 4. NMDS plot showed relationship between the AOA communities recovered from
526 metagenomics datasets. The datasets that have corresponding datasets in euphotic zones or
527 MPZs (same geographic locations) were bolded.

528

529 Fig. 5. Correlation of relative abundances of total AOA and AOA ESTUs with environmental
530 factors using Pearson test. The significant correlations (p-value < 0.05) were displayed. The
531 ESTUs of different subclades were distinguished with different colors, which are identical to
532 the coloring in Fig. 2 and 3.

533 **Supplementary information**

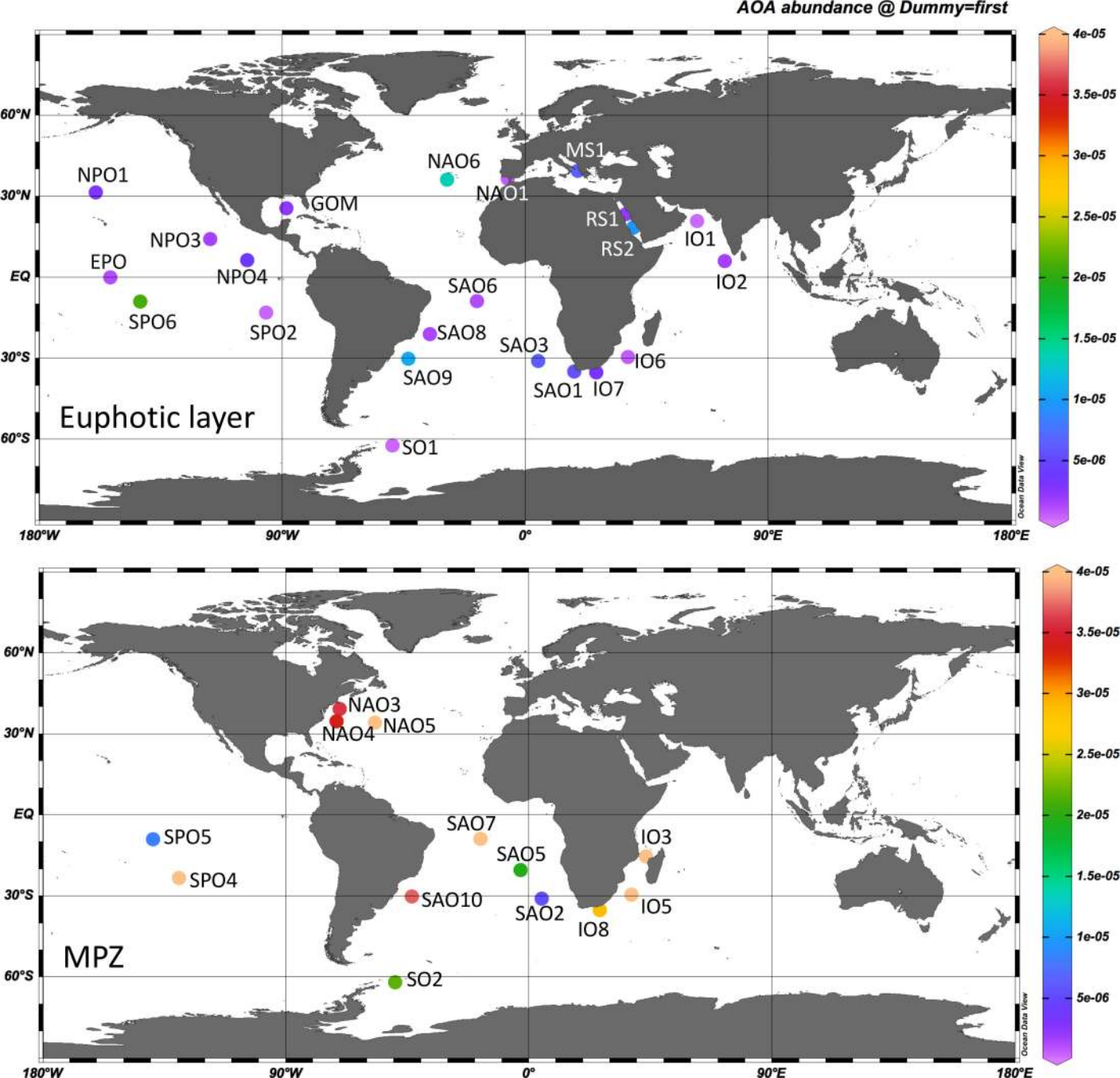
534 S1 Table. Information of the metagenomics datasets used in this study.

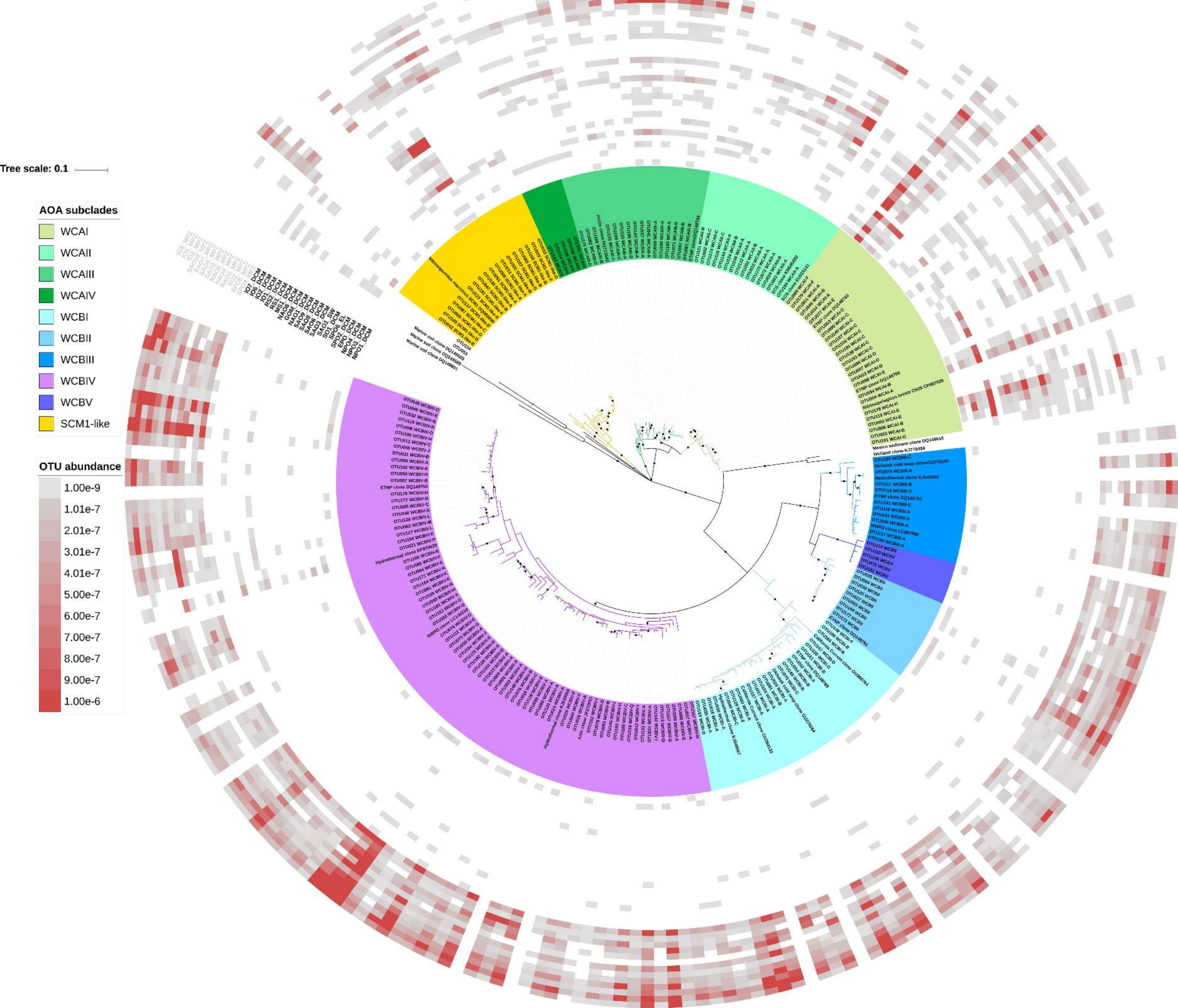
535

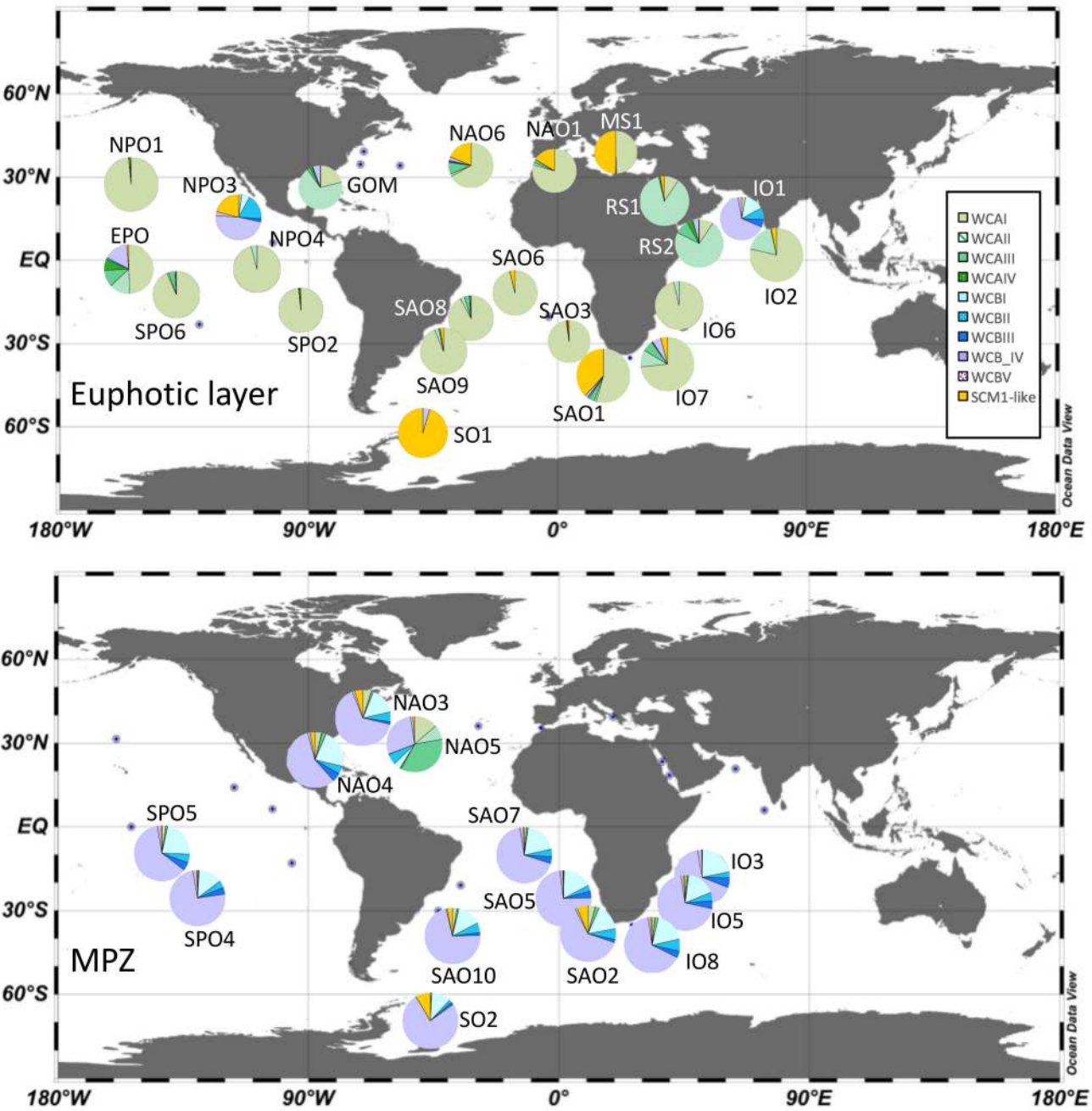
536 S2 Table. Mismatch between the qPCR primer sets that targeted WCA and WCB clades

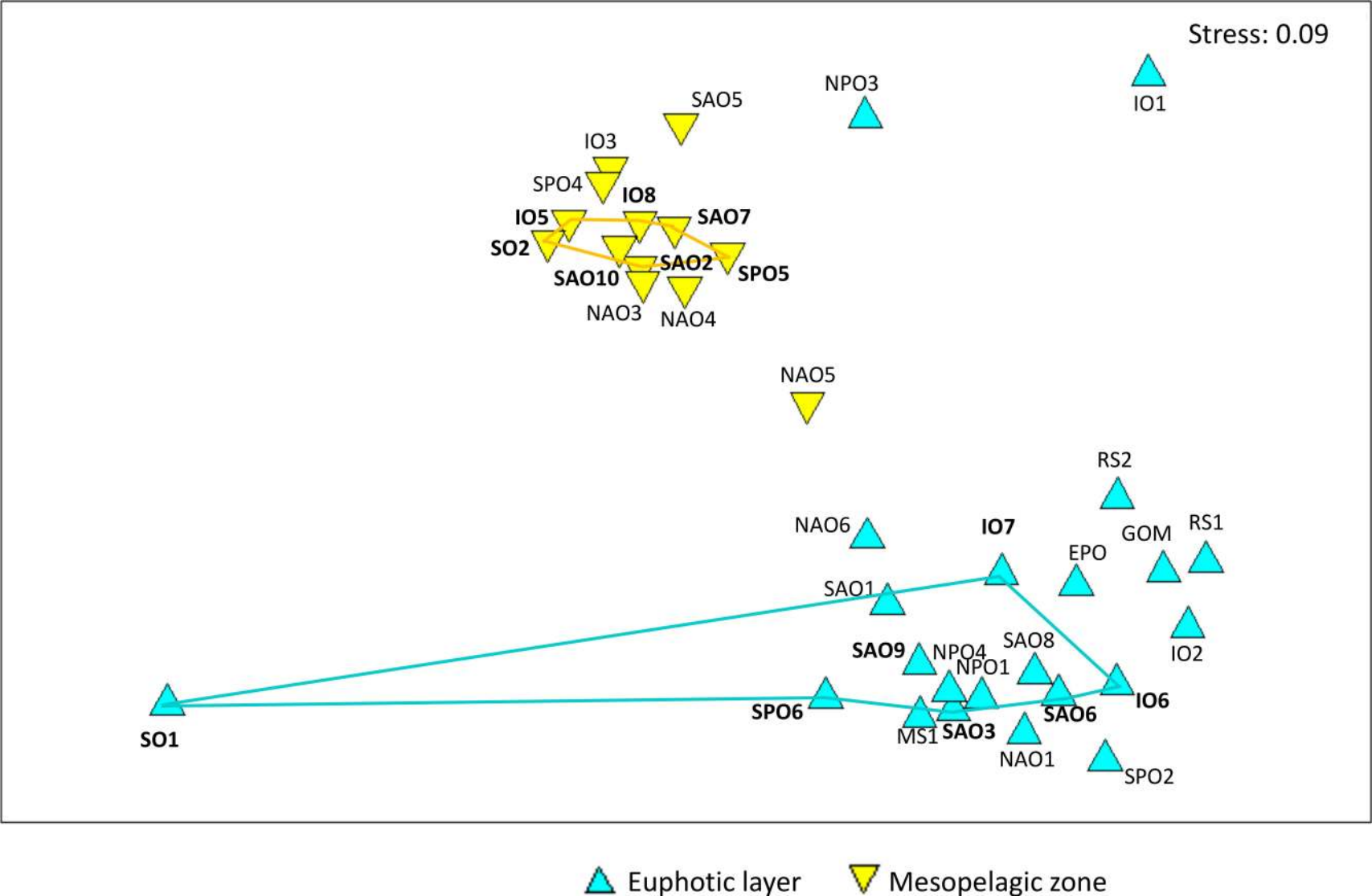
537 (Mosier and Francis, 2011) with the representative sequences of the AOA subclades.

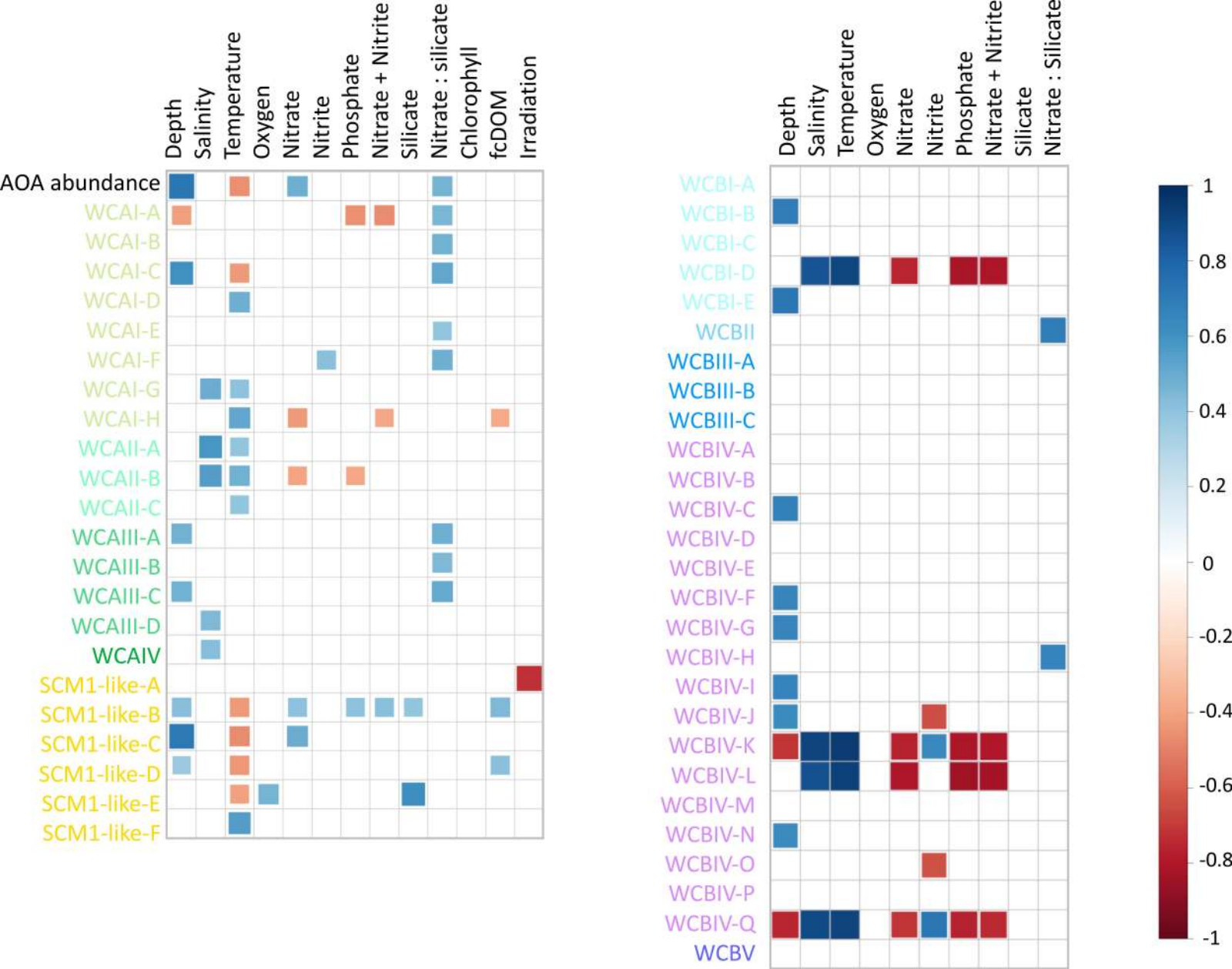
538


AOA abundance @ Dummy=first
4e-05
3.5e-05
3e-05
2.5e-05
2e-05
1.5e-05
1e-05
5e-06
60°N
30°N
EQ
30°S
60°S
180°W
90°W
0°
90°E
180°E
NPO1
GOM
NAO6
MS1
NAO1
RS1
RS2
IO1
IO2
NPO3
EPO
NPO4
SAO6
SPO6
SPO2
SAO8
SAO3
IO6
SAO9
SAO1
IO7
Euphotic layer
SO1
Ocean Data View
NAO3
NAO4
NAO5
SPO5
SAO7
IO3
SAO5
SPO4
SAO10
SAO2
IO5
IO8
MPZ
SO2




60°N
30°N
EQ
30°S
60°S
180°W
90°W
0°
90°E
180°E
NPO1
NPO3
GOM
NAO6
NAO1
MS1
IO1
RS1
RS2
EPO
NPO4
SAO6
IO2
SAO8
SAO3
IO6
SPO6
SPO2
SAO9
SAO1
IO7
SO1
Euphotic layer
WCAI
WCAII
WCAIII
WCAIV
WCBI
WCBII
WCBIII
WCB_IV
WCBV
SCM1-like
Ocean Data View
NAO3
NAO5
NAO4
SPO5
SAO7
IO3
SAO5
IO5
SPO4
IO8
SAO10
SAO2
SO2
MPZ
Ocean Data View


Stress: 0.09
NPO3
IO1
SAO5
IO3
SPO4
IO8
IO5
SAO7
SO2
SAO2
SPO5
SAO10
NAO3
NAO4
NAO5
RS2
NAO6
IO7
GOM
RS1
EPO
SAO1
IO2
SAO9
NPO4
SAO8
NPO1
IO6
SPO6
SAO6
SO1
MS1
SAO3
NAO1
SPO2
Euphotic layer
Mesopelagic zone


Depth
Salinity
Temperature
Oxygen
Nitrate
Nitrite
Phosphate
Nitrate + Nitrite
Silicate
Nitrate : silicate
Chlorophyll
fcDOM
Irradiation
AOA abundance
WCAI-A
WCAI-B
WCAI-C
WCAI-D
WCAI-E
WCAI-F
WCAI-G
WCAI-H
WCAII-A
WCAII-B
WCAII-C
WCAIII-A
WCAIII-B
WCAIII-C
WCAIII-D
WCAIV
SCM1-like-A
SCM1-like-B
SCM1-like-C
SCM1-like-D
SCM1-like-E
SCM1-like-F
Depth
Salinity
Temperature
Oxygen
Nitrate
Nitrite
Phosphate
Nitrate + Nitrite
Silicate
Nitrate : Silicate
WCBI-A
WCBI-B
WCBI-C
WCBI-D
WCBI-E
WCBII
WCBIII-A
WCBIII-B
WCBIII-C
WCBIV-A
WCBIV-B
WCBIV-C
WCBIV-D
WCBIV-E
WCBIV-F
WCBIV-G
WCBIV-H
WCBIV-I
WCBIV-J
WCBIV-K
WCBIV-L
WCBIV-M
WCBIV-N
WCBIV-O
WCBIV-P
WCBIV-Q
WCBV
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1