# Controlling factors on the global distribution of a representative marine non-cyanobacterial diazotroph phylotype (Gamma A)

Zhibo Shao[1], Ya-Wei Luo[1]

[1]State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, 361102 Xiamen, Fujian, China

*Correspondence to*: Ya-Wei Luo (ywluo@xmu.edu.cn)

**Abstract.** Non-cyanobacterial diazotrophs may be contributors to global marine $N_2$ fixation, while the factors controlling their distribution are unclear. Here, we explored what controls the distribution of the most sampled non-cyanobacterial diazotroph phylotype, Gamma A, in the global ocean. First, we represented Gamma A abundance by its *nifH* qPCR copies reported in the literature and analyzed its relationship to climatological biological and environmental conditions. There was a positive correlation between the Gamma A abundance and local net primary production (NPP), and the maximal observed Gamma A abundance increased with NPP and became saturated when NPP reached ~400 mg C $m^{-2}$ $d^{-1}$. Additionally, an analysis using a multivariate generalized additive model (GAM) revealed that the Gamma A abundance increased with light intensity but decreased with increasing iron concentration. The GAM also showed a weak but significant positive relationship between Gamma A abundance and silicate concentration, as well as a substantial elevation of Gamma A abundance when the nitrate concentration was very high (> ~10 μM). Using the GAM, these climatological factors together explained 43% of the variance in the Gamma A abundance. Second, in addition to the climatological background, we found that Gamma A abundance was elevated in mesoscale cyclonic eddies in high-productivity (climatological NPP > 400 mg $m^{-2}$ $d^{-1}$) regions, implying that Gamma A can respond to mesoscale features and benefit from nutrient inputs. Overall, our results suggest that Gamma A tends to inhabit ocean environments with high productivity and low iron concentrations, and therefore provide insight into the niche differentiation of Gamma A from cyanobacterial diazotrophs, which are generally most active in oligotrophic ocean regions and need a sufficient iron supply, although both groups prefer well-lit surface waters. More sampling on Gamma A and other non-cyanobacterial diazotroph phylotypes is needed to reveal the controlling mechanisms of heterotrophic $N_2$ fixation in the ocean.

## 1 Introduction

Dinitrogen ($N_2$) fixation, mostly conducted by prokaryotic bacteria (termed "diazotrophs"), is an important bioavailable nitrogen (N) source to the ocean (Moore et al., 2018; Karl et al., 2002). Although autotrophic cyanobacteria have been recognized as important diazotrophs in the ocean (Zehr, 2011), non-cyanobacteria diazotrophs (NCDs) that are presumably heterotrophic (probably including photoheterotrophic) bacteria (Bombar et al., 2016) have been widely detected (e.g.,

Moisander et al., 2008; Langlois et al., 2008; Halm et al., 2012; Moisander et al., 2014; Shiozaki et al., 2014) and sometimes even found to dominate the diazotrophic gene pools in surface oceans (Farnelid et al., 2011). For example, NCD *nifH* (a gene encoding subunit of nitrogenase enzyme) amplicons had higher relative abundances than autotrophic diazotrophs at some sampling sites in the South Pacific Ocean (Halm et al., 2012; Moisander et al., 2014), Indian Ocean (Shiozaki et al., 2014; Wu et al., 2019) and South China Sea (Ding et al., 2021). Metagenomic studies also revealed the abundant presence of diverse $N_2$-fixing proteobacteria in ocean genomic databases (Delmont et al., 2018; Delmont et al., 2021). Additionally, *nifH* of NCDs was also detected in subphotic seawaters (Benavides et al., 2018) and oxygen deficient zones (Jayakumar and Ward, 2020; Loescher et al., 2014) where nitrogen loss was considered significant (Lam and Kuypers, 2011). Although the $N_2$ fixed by NCDs has not been quantified, substantial $N_2$ fixation found in aphotic zones (Rahav et al., 2013; Bonnet et al., 2013) and in experiments with photosynthetic inhibitors (Rahav et al., 2015; Geisler et al., 2020), as well as recovered transcripts of the NCD *nifH* gene (Fernandez et al., 2011; Gradoville et al., 2017), provided a line of indirect evidence of heterotrophic $N_2$ fixation in the ocean.

The major known NCD classes include bacteria such as Alphaproteobacteria, Gammaproteobacteria, Epsilonproteobacteria, Betaproteobacteria, and Firmicutes belonging to Cluster I of *nifH* clusters, and some obligate anaerobic bacteria and archaea belonging to Cluster III of *nifH* clusters (Zehr et al., 2003; Riemann et al., 2010). Among them, Gamma A is the most sampled and studied phylotype. Gamma A represents a part of uncultured Gammaproteobacterial sequences isolated from the open ocean, and its cluster is distantly related to cultured NCD (Langlois et al., 2015). Gamma A *nifH* gene expression has been widely found in the global ocean, suggesting its important role in marine $N_2$ fixation (Bird et al., 2005; Moisander et al., 2014; Langlois et al., 2015; Shiozaki et al., 2017).

It is unclear what controls the growth and distribution of NCDs, as most of them, including Gamma A, are uncultivated (Bombar et al., 2016). NCDs can be different from their autotrophic counterparts in, if they are heterotrophic or photoheterotrophic, depending on organic matter as their carbon and energy sources, which can be supported by experimental evidence that $N_2$ fixation is stimulated by adding dissolved organic matter (DOM) (Rahav et al., 2016; Rahav et al., 2015; Bonnet et al., 2013; Bentzon-Tilia et al., 2015). However, DOM addition sometimes did not stimulate *nifH* expression of Gamma A even when its DNA copies were ambient (Benavides et al., 2018b), implying that DOM may not always stimulate the activity of Gamma A. Due to sensitivity to $O_2$ and high energy requirements of $N_2$ fixation (Bombar et al., 2016), abundant NCDs were found to associate with particles that supposably provide diazotrophs with a microenvironment with depleted oxygen and rich organic matter (Riemann et al., 2010; Farnelid et al., 2010; Scavotto et al., 2015; Pedersen et al., 2018; Geisler et al., 2019). NCDs were also detected in diatom mats (Martínez et al., 1983), implying another novel habitat for NCDs. An isolated strain of diazotrophic Alphabacteria from the Baltic Sea found to contain photosynthetic genes (Bentzon-Tilia et al., 2015) may complicate this issue, doubting whether NCDs can be mixotrophic and also depend on light.

Although dissolved inorganic nitrogen (DIN) is generally considered to inhibit marine $N_2$ fixation (Karl et al., 2002; Zehr and Kudela, 2011), substantial presence of NCDs is found in DIN-replete environments such as estuaries (Geisler et al., 2020),

coastal zones (Li et al., 2020), upwelling regions (Geisler et al., 2020; Moreira-Coello et al., 2017; Dekaezemacker et al., 2013) and other eutrophic seas (Bird and Wyman, 2013). Culture experiments showed that the DIN inhibition effect on NCDs can be strain specific (Bentzon-Tilia et al., 2015; Martínez-Pérez et al., 2018). Temperature could be another factor controlling NCDs which may prefer warm oligotrophic surface oceans (Langlois et al., 2015; Shiozaki et al., 2017), the same region where the majority of autotrophic $N_2$ fixation occurs (Wang et al., 2019; Luo et al., 2014). Similar to cyanobacterial diazotrophs, phosphate can also limit the growth of NCDs in oligotrophic environments (Rahav et al., 2015). Regarding other important factors that control autotrophic diazotrophs, iron (Fe) may potentially impact NCDs if they also depend on the high Fe-containing nitrogenases to fix $N_2$ (Bombar et al. 2016), although, as discussed above, the $N_2$ fixation by NCDs is still not quantified. Strong stratification may also benefit NCDs by accumulating organic matter in the upper water column (Langlois et al. 2015). However, to our knowledge, no studies have analyzed the effects of Fe or stratification on NCDs.

Mesoscale eddies may also impact NCD abundance. Although anticyclonic eddies were generally considered to benefit autotrophic diazotrophs by inhibiting vertical DIN input from deep waters (Liu et al., 2020; Fong et al., 2008; Church et al., 2009), a class of NCDs, Gammaproteobacteria, were found to dominate diazotrophic communities inside cyclonic eddies in the South China Sea (Zhang et al., 2011; Liu et al., 2020). Different types of mesoscale eddies may have discrepancies in impacting the ecophysiology of NCDs and their autotrophic counterparts (Benavides and Robidart, 2020).

Langlois et al. (2015) analyzed the distribution of the Gamma A phylotype in the Pacific and Atlantic Oceans and suggested that Gamma A prefers warm and oligotrophic surface oceans. With more data becoming available in recent years, we collected, to the best of our knowledge, all the reported in situ measurements of Gamma A *nifH* copies using qPCR assays, compiling a dataset with 80% more data than those used in Langlois et al. (2015). We then analyzed the relationship between this *nifH*-based Gamma A abundance and the long-term background of ecological and environmental factors by using their climatological monthly averages. In addition to temperature and concentrations of nitrate, phosphate and silicate that have been used in Langlois et al. (2015), we included 5 more variables (primary production, Fe, DOC concentrations, solar radiation and mixed layer depth) to more thoroughly analyze potential controlling factors on Gamma A. We further explored the influence of mesoscale eddies on Gamma A abundance. Our analyses suggested that local primary productivity, temperature, dissolved Fe concentration and the occurrence of cyclonic eddies can be the main factors impacting the distribution of Gamma A in the global ocean.

## 2 Methods

### 2.1 Data summary and quality control of Gamma A abundance

A total of 1861 *in situ* measurements of *nifH* copies of Gamma A in the Pacific, Atlantic and Indian Oceans were collected from 21 published papers (Table 1), and are available in a data repository (https://doi.org/10.6084/m9.figshare.17284517) (Shao and Luo, 2021). Gamma A was sometimes also named γ-24774A11 in the collected papers (Moisander et al., 2008). All

these data were measured using quantitative polymerase chain reaction (qPCR). Note that the primer of Gamma A used by Langlois et al. (2015) in the North Atlantic was slightly different from other studies. Most samples (88%) were collected in the upper 100 m of the water column. In the following analyses, we represented Gamma A abundance using its *nifH* copies, although we noted that variations in *nifH* copies in different cyanobacterial diazotrophic cells have been reported (White et al., 2018; Sargent et al., 2016) and *nifH* copy numbers in Gamma A genome remain unknown.

The non-zero *nifH*-based abundance data of Gamma A were approximately log-normally distributed (Fig. S1). There were 748 data points reporting zero *nifH* copies which theoretically could indicate that Gamma A in the samples was either true absent or its abundance was below the detection limit. As the reported detection limit of qPCR usually ranges from $10^1$ to $10^2$ copies $L^{-1}$, the number of the Gamma A *nifH* data that could be below detection in our dataset, according to the log-normal distribution of observed non-zero data, was very likely less than 72 even assuming a high detection limit of $10^2$ copies $L^{-1}$ (Fig. S1). The fact that there were far more zero-value data (748) in our dataset indicated that a high fraction of the zero-value data could represent true absence of Gamma A.

The zero-value abundance data of Gamma A were not included in our further analyses, mainly because of two reasons. First, the fact that Gamma A was absent in many samples, as well as the spatially mixed distribution of the zero-value and non-zero Gamma A abundance data (see Results), indicated the patchy distribution of Gamma A, which was also widely found for other diazotrophs as a consequence of lateral transport and mixing of water masses (Robidart et al., 2014).

The patchiness of Gamma A implicated that it could be either present or absent even when the environmental conditions were suitable for its growth. It can also indicate our limited understanding of environmental conditions: The currently available environmental data do not include all the controlling factors of Gamma A. Nevertheless, if the zero-value data were included, similar environmental conditions could possibly be associated with both high abundance and zero abundance of Gamma A (Fig. S2), which would bias the response function of our statistical analyses, particularly as the fraction of the zero-abundance data was large (~1/3). Second, it is difficult to identify whether the zero-value data represented true absence or below-detection abundance of Gamma A, considering that the accuracy of qPCR was highly sensitive to sample preservation, extraction protocol and the reliance of the standard curve (Smith and Osborn, 2009).

Chauvenet's criterion was used to identify outliers in the non-zero Gamma A abundance data by first log-transforming all the data (Glover et al., 2011). Two outliers (0.22 copies $L^{-1}$ and 0.33 copies $L^{-1}$) were removed because their probability of deviation from the mean was smaller than $1/(2n)$, where *n* was the number of data. Even though they can be reliable, we excluded them from the analyses to avoid possible biases.

**Table 1. Data source of complied Gamma A *nifH* samples.**

| Reference | Location | Latitude | Longitude | Depth (m) |
|---|---|---|---|---|
| *Pacific Ocean* | | | | |
| Moisander et al. (2008) | South China Sea | 9°-12°N | 107°- 110°W | 0-1700 |
| Bombar et al. (2011) | Mekong River Plume | 9°-11°N | 106°- 107°W | 0 |
| Hamersley et al. (2011) | Southern California Bight | 33°N | 118°W | 5-885 |
| Halm et al. (2012) | South Pacific Gyre | 27°- 42°S | 117°- 153°W | 0-100 |
| Turk-Kubo et al. (2014) | Eastern Tropical South Pacific | 10°- 20°S | 80°- 110°W | 5-225 |
| Moisander et al. (2014) | South Pacific Ocean | 15°- 30°S | 177°E- 155°W | 4-175 |
| Shiozaki et al. (2015) | Northwest Pacific | 38°- 39°N | 141°- 143°W | 0-119 |
| Berthelot et al. (2017) | Western Pacific Ocean | 3°- 12°S | 140°- 160°W | 5-70 |
| Shiozaki et al. (2017) | North Pacific Ocean | 0°- 68°N | 168°- 170°E | 0-157 |
| Shiozaki et al. (2018a) | South Pacific Ocean | 0°- 40°S | 170°- 100°W | 0-215 |
| Shiozaki et al. (2018b) | Kuroshio | 25°-33°N | 124°- 139°W | 0-5 |
| Chen et al. (2019a) | Western Pacific Ocean | 0° - 21°N | 110°- 159°W | 0-150 |
| Chen et al. (2019b) | South China Sea | 19° - 22°N | 116°- 121°W | 5-1000 |
| Cheung et al. (2020) | North Pacific Ocean | 7°- 54°N | 139°E- 80°W | 5 |
| *Atlantic Ocean* | | | | |
| Langlois et al. (2008) | North Atlantic Ocean | 0° - 40°N | 10°- 70°W | 5-100 |
| Benavides et al. (2016) | North Atlantic Ocean | 0°-21°N | 15°-75°W | 0-150 |
| Martinez-Perez et al. (2016) | Tropical North Atlantic Ocean | 11°-15°N | 21°-60°E | 5-200 |
| Moreira-Coello et al. (2017) | Upwelling Region off NW Iberia | 6°-18°N | 18°-54°E | 0-157 |
| Moore et al. (2018) | Tropical Atlantic Ocean | 0°-21°N | 15°-55°W | 0 |
| *Indian Ocean* | | | | |
| Shiozaki et al. (2014) | Arabian Sea | 4°S–20°N | 65° –70° E | 0-86 |
| Wu et al. (2019) | Bay of Bengal | 4°S–10°N | 84° –96° E | 0-200 |

## 2.2 Environmental and ecological parameters

Monthly climatological environmental and ecological parameters were used as predictors for Gamma A abundance (Table 2).
Temperature and concentrations of nitrate, phosphate and silicate were the products of the World Ocean Atlas (WOA) 2018 (www.nodc.noaa.gov)(Locarnini et al., 2018; Garcia et al., 2019), and excess phosphate ($P^*$) was derived from concentrations of phosphate and nitrate based on the Redfield ratio ($P^*$ = [phosphate] - [nitrate]/16). Dissolved iron (Fe) concentrations were obtained from the Community Earth System Model – Biogeochemistry module (Misumi et al., 2014). Dissolved organic carbon concentration used a product estimated by artificial neural network (Roshan and Devries, 2017). Mixed layer depth (MLD) was downloaded from Ifremer (http://www.ifremer.fr/) using the criterion that the potential density of water parcels at the depth was 0.03 kg m$^{-3}$ higher than that at the surface (De Boyer Montégut et al., 2004). Net primary production used a satellite data based on the VGPM algorithm (Behrenfeld and Falkowski, 1997) (http://sites.science.oregonstate.edu/ocean.productivity/). Surface photosynthetically active radiation (PAR) was downloaded from MODIS-Aqua program (http://oceancolor.gsfc.noaa.gov/). To estimate vertical profile of PAR, we first obtained the estimated euphotic zone depth $Z_e$ (https://oceancolor.gsfc.nasa.gov/) at 1% surface PAR based on an inherent optical property (IOP)-centered approach (Lee et al., 2005), and used it to estimate the attenuation coefficient:

$$k_d = \frac{\ln(0.01)}{Z_e}. \tag{1}$$

The PAR at depth $z$ can be calculated while we assumed organisms in the mixed layer were exposed to PAR homogenously:

$$PAR(z) = \begin{cases} PAR_0 e^{-k_d z} & (z > MLD) \\ \frac{1}{MLD} \int_0^{MLD} PAR_0 e^{-k_d z} dz & (z < MLD) \end{cases}, \tag{2}$$

where $PAR_0$ is the surface PAR.

To identify whether the Gamma A abundance was sampled in cyclonic or anticyclonic eddies, we extracted from AVISO program (www.aviso.altimetry.fr) the satellites-merged daily sea level anomaly (SLA) for the sampling days of the Gamma A data. The cores of mesoscale eddies were identified by the outermost closed contour lines of the SLA field. Only those sampling points located in the cores of cyclonic (negative SLA) or anticyclonic (positive SLA) eddies were recorded. Otherwise, data points were recorded as "outside eddy".

All the variables used in the analyses are available in a data repository. (https://doi.org/10.6084/m9.figshare.17284517) (Shao and Luo, 2021)

**Table 2. Environmental and ecological parameters.**

| Environmental parameters | Symbol | Source | Spatial resolution | Log-normally distributed and log-transformed |
|---|---|---|---|---|
| Temperature (°C) | T | | 1° | No |
| Nitrate (μM) | $NO_3$ | World Ocean Atlas 2018 | 1° | Yes |
| Phosphate (μM) | $PO_4$ | | 1° | Yes |
| Silicate (μM) | Si | | 1° | Yes |
| Excess P (μM) | $P^*$ | $PO_4$-$NO_3$/16 | 1° | No |
| Fe (nM) | Fe | CESM | 1° | Yes |
| Mixed layer depth (m) | MLD | IFREMER | 2° | Yes |
| Net primary production (mg C $m^{-2}$ $d^{-1}$) | NPP | VGPM | 1/6° | Yes |
| Photosynthetic active radiation (E $m^{-2}$ $d^{-1}$) | PAR | MODIS-Aqua | 1/6° | Yes |
| Dissolved organic carbon (μM) | DOC | Model | 1° | No |
| Sea level anomaly | SLA | AVISO | 1/6° | No |

## 2.3 Statistical analyses

For Gamma A data points sampled in the same months and the same depth bins (defined in WOA), they were binned to 2° × 2° grids to help eliminate possible biases caused by concentrated samplings in specific regions, resulting in 939 binned means of log-10 based Gamma A *nifH* abundance. The corresponding environmental and ecological parameters were also averaged to the same bins when necessary.

We used the generalized additive model (GAM) using R package 'mgcv' (Wood, , 2017) to demonstrate non-linear relationships between the multiple predictors and the Gamma A abundance:

$$y = \alpha + \sum_{i=1}^{n} s(x_i) + \varepsilon, \tag{3}$$

where y is response variable (Gamma A abundance), $x_i$ is the $i$th predictor (i.e., the environmental or ecological variable), $\alpha$ is the intercept, $s(x_i)$ is a linear combination of smooth functions of predictor $x_i$, $n$ is the number of predictors and $\varepsilon$ is the standard error. To avoid over-fitting to noise, the Restricted Maximum Likelihood (REML) method was selected for the GAM smoothing parameters of every predictor with the basis function number (k) set to 9 (Wood et al., 2016). In the model selection of GAM, a double penalization approach was used to identify and remove those insignificant predictors with large smoothing parameters and set them to zero functions (Marra and Wood, 2011).

The scientific color maps are used in several figures to prevent visual distortion of the data and exclusion of readers with color-vision deficiencies (Crameri et al., 2020).

## 3 Results and Discussion

### 3.1 Global distribution of Gamma A *nifH* abundance

The *nifH* gene abundance ranges from 1 to $10^7$ copies $L^{-1}$ in the global ocean and shows an approximately log-normal distribution (Fig. S1). High abundance of Gamma A *nifH* abundance over $10^5$ copies $L^{-1}$ is prevalent in the subpolar North Pacific, tropical Atlantic and Bay of Bengal (Indian Ocean) (Fig. 1). Most Gamma A abundance data were sampled above 100 m, particularly in the upper 25 m. The deepest datum with detectable Gamma A *nifH* was sampled at 885 m in Southern California Bight (Hamersley et al., 2011). Available data showed that *nifH* abundance decreased with depth in the Southwestern Pacific Ocean, the Indian Ocean and the South China Sea, but did not have an apparent trend from the surface down to 200 m in the tropical Atlantic Ocean (Fig. S3). More data particularly in deep waters are needed to better and more reliably reveal the vertical pattern of Gamma A abundance.

Although high Gamma A abundance over $10^6$ *nifH* copies $L^{-1}$ was observed in the surface North Pacific Ocean, zero-value data were also massive (215 in a total of 608 data points) and even located closly to those high-abundance data (Cheung et al., 2020) (Fig.1), indicating the patchy distribution of Gamma A. As discussed already (Section 2.1), zero-abundance data were not included in the further analyses due to the patchiness of Gamma A and the limitations of qPCR method in detecting the true absence of Gamma A.

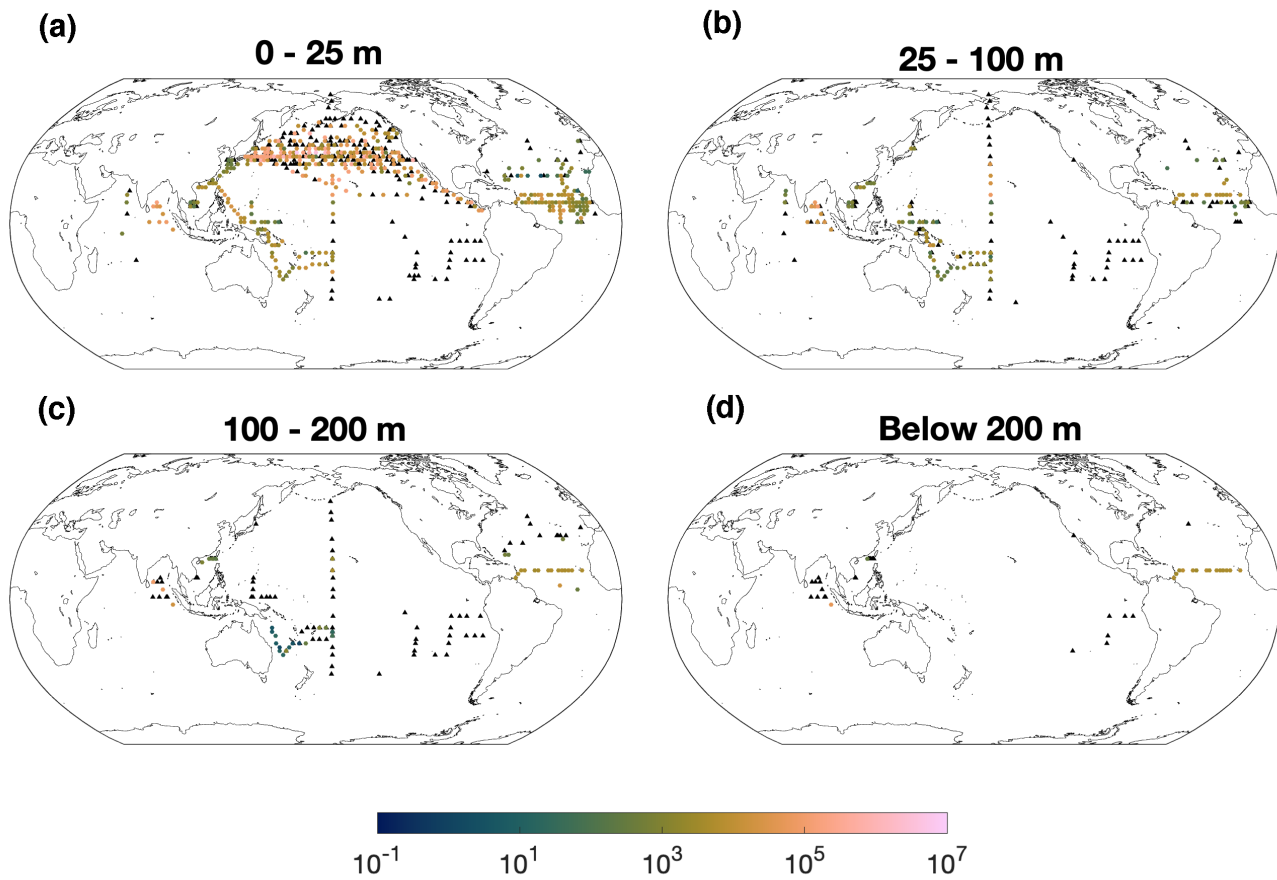



**Figure 1. Gamma A abundance (*nifH* copies L⁻¹).** The panels show data in depth ranges of (a) $0 - 25$ m, (b) $25 - 100$ m, (c) $100 - 200$ m and (d) below 200 m. For clear demonstration, data are binned to $2° \times 2°$ and geometric means in each bin are shown. Zero-value data were denoted as black triangles.

195

**3.2 Relationship between primary production and Gamma A abundance**

If the presumption that Gamma A is heterotrophic or photoheterotrophic bacteria (Bombar et al., 2016; Zehr and Capone, 2020) is true, a positive relationship between the Gamma A abundance and net primary production (NPP) can be expected because its energetically intensive $N_2$ fixation can benefit from a sufficient supply of organic matter from primary producers. The significant positive correlation between the logarithm of Gamma A *nifH* abundance and the logarithm of NPP in our data

(correlation $= 0.21$, $p < 0.01$) (Fig. 2) was consistent with this presumption. However, this positive correlation could just reveal a fact that Gamma A and primary producers share certain common controlling factors. For example, even if Gamma A would be autotrophic or mixotrophic and can harvest energy from solar radiation, it could also positively correlate with NPP, as both of them would be supported by high light intensity. Although the capability of Gamma A to fix $N_2$ has not been quantified, it could also be possible that the fixed N by Gamma A, if it occurred, could in turn support NPP. It is also interesting that the upper bound of the observed Gamma A abundance increased with the NPP [$\log_{10}$(Gamma A) = 3.7 $\log_{10}$NPP - 3.2] when NPP $< 10^{2.6}$ ($\approx$400) mg C m$^{-2}$ d$^{-1}$, above which the upper bound of Gamma A abundance saturated at ~$10^7$ *nifH* copies L$^{-1}$ (Fig. 2). This pattern indicated that the maximal potential of Gamma A abundance was positively associated with NPP, although it is unclear why the maximal Gamma A abundance ceased to increase when NPP increased above 400 mg C m$^{-2}$ d$^{-1}$. The observed reduction of Gamma A from its maximal potential could indicate that Gamma A was limited by other environmental factors and/or simply reflect the vertical distribution of Gamma A in the water column (see more analysis below), noting that the NPP used here is a depth-integrated rate of the euphotic zone.

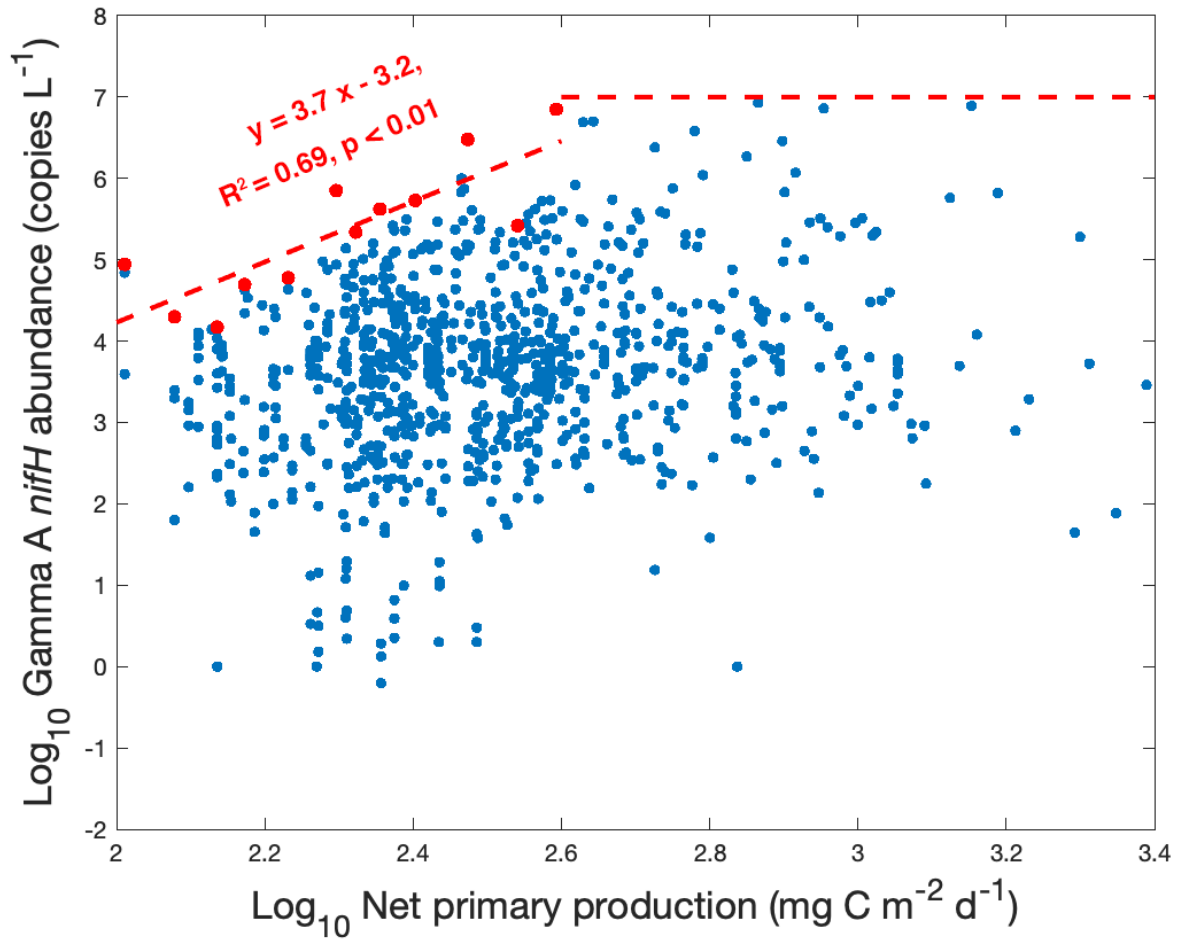
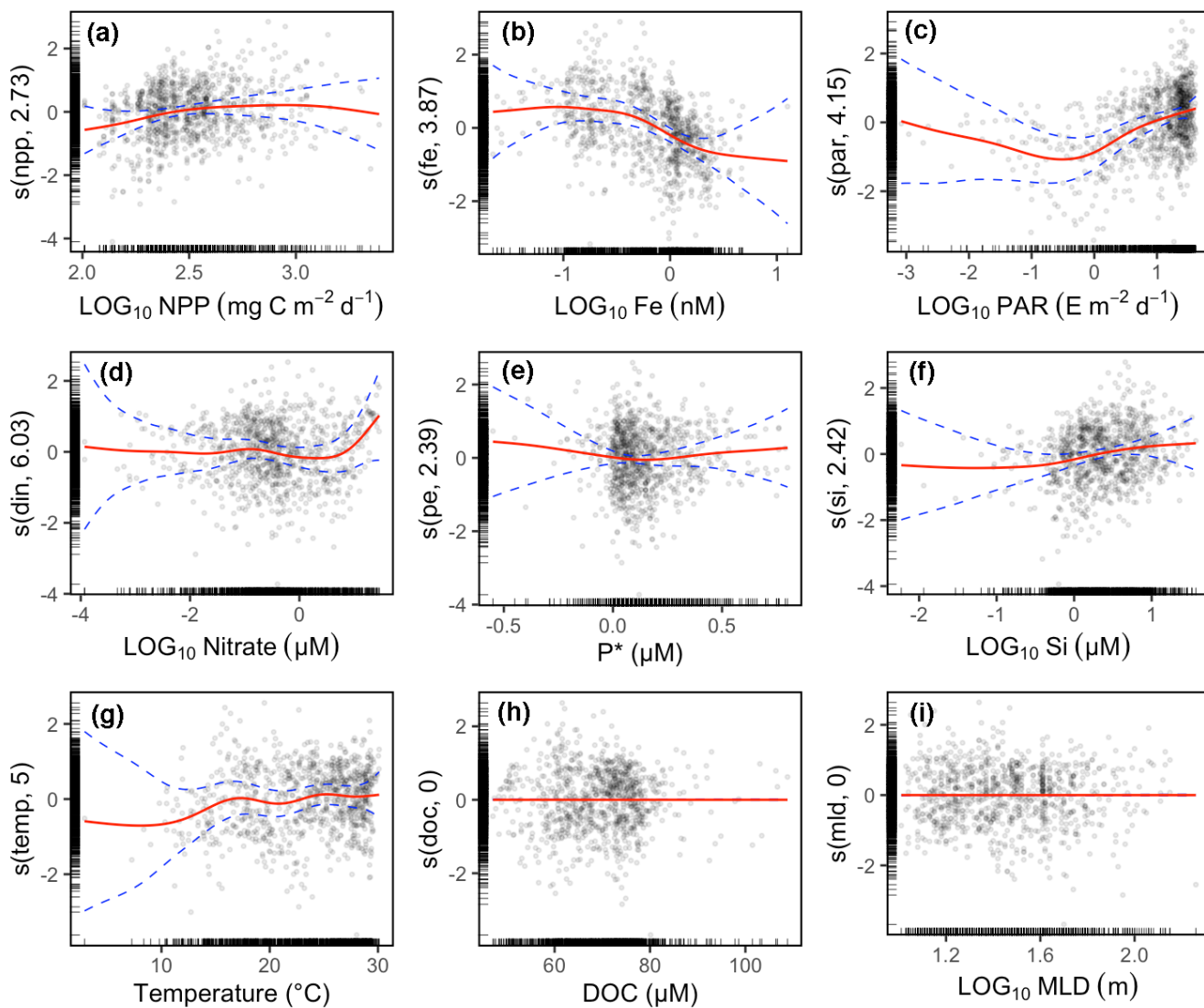
Nevertheless, our results show that Gamma A tends to inhabit high-productivity waters. Our finding contradicted the hypothesis mentioned above that Gamma A preferred oligotrophic waters based on samples mainly in tropical and subtropical Pacific and Atlantic Oceans, in which Gamma A reached $8 \times 10^4$ *nifH* copies L$^{-1}$ (Shiozaki et al., 2018a; Langlois et al., 2015). However, the new dataset (Cheung et al., 2020) included in the present study showed even higher (over $10^5$ nifH copies L$^{-1}$) Gamma A abundance in the subarctic North Pacific (Fig. 1), where nutrient concentrations and NPP are generally high.

225

**Figure 2. The relationship between Gamma A abundance and net primary production.** Both Gamma A abundance and net primary production (NPP) were $\log_{10}$-transformed. The data with NPP of $10^{2.0}$–$10^{2.6}$ mg C m$^{-2}$ d$^{-1}$ are divided into 12 groups with equal log-NPP intervals of 0.05, and the highest Gamma A abundance in each group is identified (red dots). The upper bound of Gamma A abundance (red line) is estimated by linearly fitting the red dots in the low NPP range and saturates at

230  ~$10^{7.0}$ *nifH* copies L$^{-1}$ for NPP > $10^{2.6}$ mg C m$^{-2}$ d$^{-1}$.

### 3.3. Multivariate nonlinear relationships between environmental factors and Gamma A abundance using GAM

We then analyzed which environmental factors can be the predictors for Gamma A abundance. Multivariate GAM analysis was used to obtain a nonlinear relationship between Gamma A abundance and environmental parameters. Note that phosphate was not included in the GAM because its variance can be partly represented by $P^*$.

The GAM results also showed that Gamma A abundance increased with NPP at low or intermediate levels of NPP (Fig. 3a), while the Fe concentration and PAR even contributed more than NPP to the variance in Gamma A abundance (Figs. 3b and 3c). The prediction of Gamma A abundance is also contributed by nitrate (only when its concentration is high) and silicate (Figs. 3d and 3e). $P^*$ and temperature, however, did not show a clear pattern with Gamma A particularly in the ranges where most data were located, although Gamma A abundance was reduced at temperature lower than 15°C (Figs. 3f and 3g). The GAM did not identify a significant relationship between Gamma A abundance and DOC concentration or MLD (Figs. 3h and 3i). In the following, we discuss the relationship of Gamma A abundance to the substantial predictors, including Fe, PAR, nitrate (with reference to $P^*$) and silicate.

250 **Figure 3. Partial effects of environmental variables on Gamma A abundance using GAM multivariate analysis.** The analyses show the anomaly of Gamma A abundance contributed by the smooth function (red line) and its 95% confidence interval (dashed blue line) of each environmental variable. Data (circles) are shown as partial residuals after all other partial effects have been considered. The numbers in the parentheses of y-axis labels are the degree of freedom of the smooth functions. A degree of freedom smaller than 1 is equivalent to a linear line, and higher degrees of freedom represent more wiggly curves.

255 The blue ticks on the x-axis and y-axis also indicate the location of the data.

### 3.3.1 Iron

The Gamma A abundance generally showed a decreasing trend with increasing dissolved Fe (Figs. 3b). Our dataset showed that a high abundance of Gamma A was prevalently observed in the North Pacific Ocean (Fig. 1a), where Fe was considered the dominant limiting factor for $N_2$ fixation (Sohm et al., 2011). Other Gammaproteobacterial phylotypes, such as Gamma 3 and Gamma ETSP2, were also found to dominate the diazotrophic community in the eastern South Pacific (Turk-Kubo et al., 2014; Halm et al., 2012) where Fe heavily limited primary production (Knapp et al., 2016; Bonnet et al., 2008). It has also been suggested that Gamma A and unicellular cyanobacterial diazotroph UCYN-B share niches in Fe-depleted western and southern Pacific Oceans (Moisander et al., 2014; Chen et al., 2019a), possibly to avoid competing with other Fe-demanding diazotrophs. Gammaproteobacterial diazotrophs may be equipped with siderophore releasing genes, such as those already reported in another versatile phylotype Gamma 4 (Cheung et al., 2021), and the released siderophores are an efficient tool for scavenging low-level Fe in the ocean (Boyd and Ellwood, 2010). Although more studies are certainly needed to further explore the ecological and physiological mechanisms and evolutionary reasons, the good survival of Gamma A in a low-Fe environment is an intriguing finding that may expand our recognized space of active $N_2$ fixation in the ocean.

### 3.3.2 Light

In the range that most PAR data spanned ($> \sim 1$ E m$^{-2}$ d$^{-1}$), Gamma A abundance increased with PAR (Fig. 3c). This result represented a general pattern in which the Gamma A abundance decreased with depth (Fig. S3a, c and d) (Moisander et al., 2008; Langlois et al., 2015; Chen et al., 2019b; Shiozaki et al., 2014; Wu et al., 2019). This can be because that Gamma A was photoheterotrophic, as hypothesized by previous studies (Moisander et al., 2014), but can also be because it took the advantage of using photosynthetic products in the surface ocean (Langlois et al., 2015). The limited amount of data showing the substantial presence of Gamma A under very low PAR (Fig. 3c), as well as the nearly constant Gamma A abundance with depth in the Tropical Atlantic Ocean (Figs. S3b), raises the question of whether the active transport of organic matter from the surface can support the growth of Gamma A in the dark deeper ocean.

### 3.3.3 Nitrate and P*

Neither nitrate nor P* generally had a substantial effect on Gamma A abundance, except that Gamma A abundance increased substantially when the nitrate concentration was higher than 10 µM (Fig. 3d). This is consistent with a previous review showing that nitrate did not show an immediate inhibition of Gamma A (Moisander et al., 2017) and the findings of abundant Gamma A in oceans with high nitrate concentrations (Bird and Wyman, 2013) or shallow nitracline (Shiozaki et al., 2014). How and to what extent Gamma A, including the NCDs in general, is inhibited by nitrate remain unknown (Bombar et al., 2016). The hypothesis that low-nitrate and high-P* environments favor autotrophic diazotrophs is based on the competition of inorganic nutrients between diazotrophs and other phytoplankton (Karl and Letelier, 2008; Deutsch et al., 2007), while our results

tentatively suggested that competition may not occur strongly between NCDs and phytoplankton, although it is still unclear whether NCDs use inorganic or organic P sources. Nevertheless, our results did not suggest that high inorganic nutrients can be an inhibiting factor on Gamma A.

290

### 3.3.4 Silicate

Most Gamma A abundance data were associated with silicate concentrations higher than $10^{0.5}$ (~3.2) µM, and a positive relationship between them was revealed by the GAM (Fig. 3f), suggesting a possible association between Gamma A and diatoms. NCDs have been found on the surface of diatoms or on the diatom mats (Martínez et al., 1983), as discussed above.

295 Diatom-dominant ecosystems tend to produce abundant large particles either from dead diatoms and their aggregates or the fecal pellets generated by zooplankton (Tréguer et al., 2018). Large particles can be a good habitat for NCDs, as already discussed. Our results then provide indirect evidence for the association between Gamma A and diatoms.

### 3.3.5 Predictions based on GAM

300 Overall, the multivariate GAM model explained 43% of the variance in observed Gamma A abundance (Fig. 4). The predicted Gamma A abundance generally followed the observed values, although it tended to underestimate the observed high Gamma A abundance (> $10^5$ copies $L^{-1}$) or overestimate the low Gamma A abundance (< $10^2$ copies $L^{-1}$) (Fig. 4). The moderate explained variance indicated that although the tested environmental factors can substantially influence Gamma A abundance, there must be other untested factors and unknown mechanisms that can also substantially impact the Gamma A distribution.

305

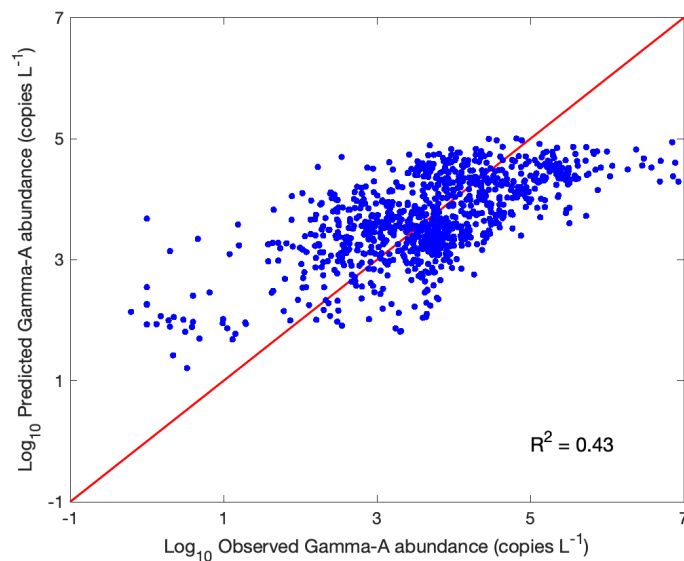



**Figure 4. Predictivity of GAM.** Comparison of predicted versus observed Gamma A *nifH* abundance. The red lines are 1:1
ratio of prediction to observation.

Although the overall $R^2$ was at a moderate level of 43%, we applied this model to give a first-order estimate of Gamma A
abundance in the surface ocean (Fig. 5a) from climatological NPP and environmental factors (Fig. S4), admitting that this
demonstration did not fully cover the observed spatial variance in Gamma A abundance. The results suggested that Gamma A
was most abundant in the upwelling region of the eastern tropical South Pacific (Fig. 5a). High Gamma A abundance was also
predicted in certain regions of the Southern Ocean (Fig. 5a), which was mostly caused by its high nitrate concentration (Figs.
S4g–h). However, the largest uncertainties for the predictions also exist in the Southern Ocean (Fig. 5b) as there were no
Gamma A samples in this high-nitrate area (Fig. 1). Future sampling in the Southern Ocean can then test our predictions and
reduce the uncertainties. The lowest predicted Gamma A abundance was in the subtropical North Atlantic and in the
southwestern Pacific near Australia (Fig. 5a), which was mostly caused by high Fe concentrations (Figs. S4c–d).

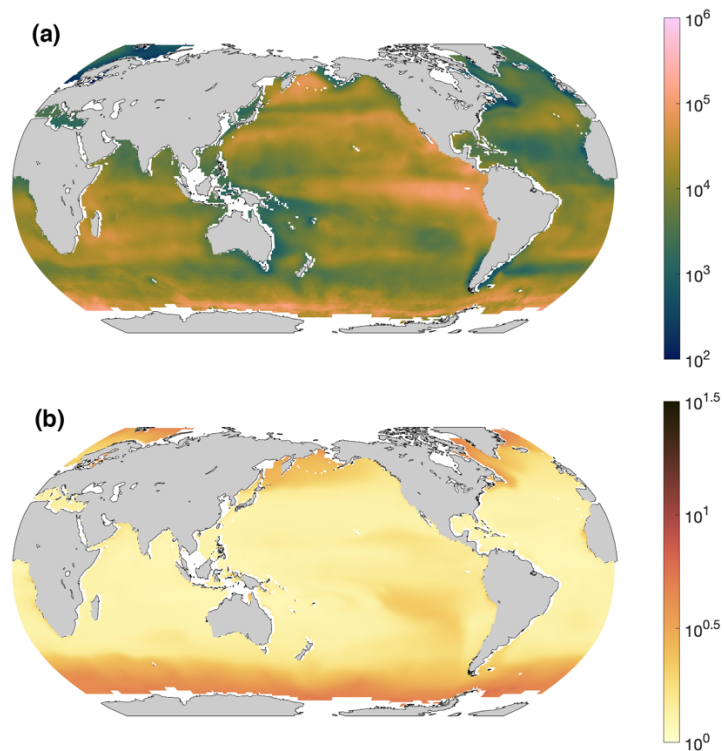



**Figure 5. Prediction of Gamma A abundance.** (a) Predicted annual mean surface (0–25 m) Gamma A abundance (*nifH* copies $L^{-1}$) and (b) the standard errors estimated by the GAM.

## 3.4 Impact of mesoscale eddies on Gamma A

The root-mean-square error (RMSE) of 0.84 and an $R^2$ of 43% in the prediction model (Fig. 4) indicated that there was still substantial unexplained variance in Gamma A abundance. One possible reason was that we used the climatological monthly means for the environmental factors, while the *in situ* conditions can differ greatly from the climatological values. For example, oceanic mesoscale eddies can influence biogeochemical processes not only by advective transport but also by variations in the biological and chemical environments (McGillicuddy, 2016). Particularly, as discussed above, some regional studies have suggested that mesoscale eddies may influence the distribution of autotrophic diazotrophs and/or NCDs. We then explored whether the occurrence of mesoscale eddies can impact Gamma A abundance. We identified 59 data points of Gamma A abundance that were sampled in cyclonic eddies, while more (268) were sampled in anticyclonic eddies. This is consistent with the fact that eddies are more likely anticyclonic in the Northern Hemisphere, where our most (74%) sampling points were located (Chelton et al., 2011).

The results showed that the average Gamma A abundance within cyclonic eddies was substantially higher than that in anticyclonic eddies or outside eddies (t-test, $p < 0.01$)(Fig. 6a). We further separated the data into a low-NPP group and a high-

NPP group at an NPP of 400 mg m$^{-2}$ d$^{-1}$, a value selected because our data showed that Gamma A abundance saturated when NPP was higher than this level (Fig. 2). Interestingly, the phenomenon that Gamma A was more abundant in cyclonic eddies

345    was mostly contributed by the data in the high-NPP group (Fig. 6c), but not by the low-NPP group (Fig. 6b). To avoid the possible biases caused by different climatological conditions at the locations of cyclonic eddies, we then checked the residuals of the predicted Gamma A abundance using climatological factors (i.e., Fig. 4) and still found that the Gamma A abundance in cyclonic eddies in the high-NPP group was significantly higher (one-tail t-test, $p < 0.05$) than the climatology-based predictions by on average nearly an order of magnitude, while this was not the case for samples in anticyclonic eddies and

350    outside eddies in the high-NPP group or for all the samples in the low-NPP group (Figs. 6d-f).

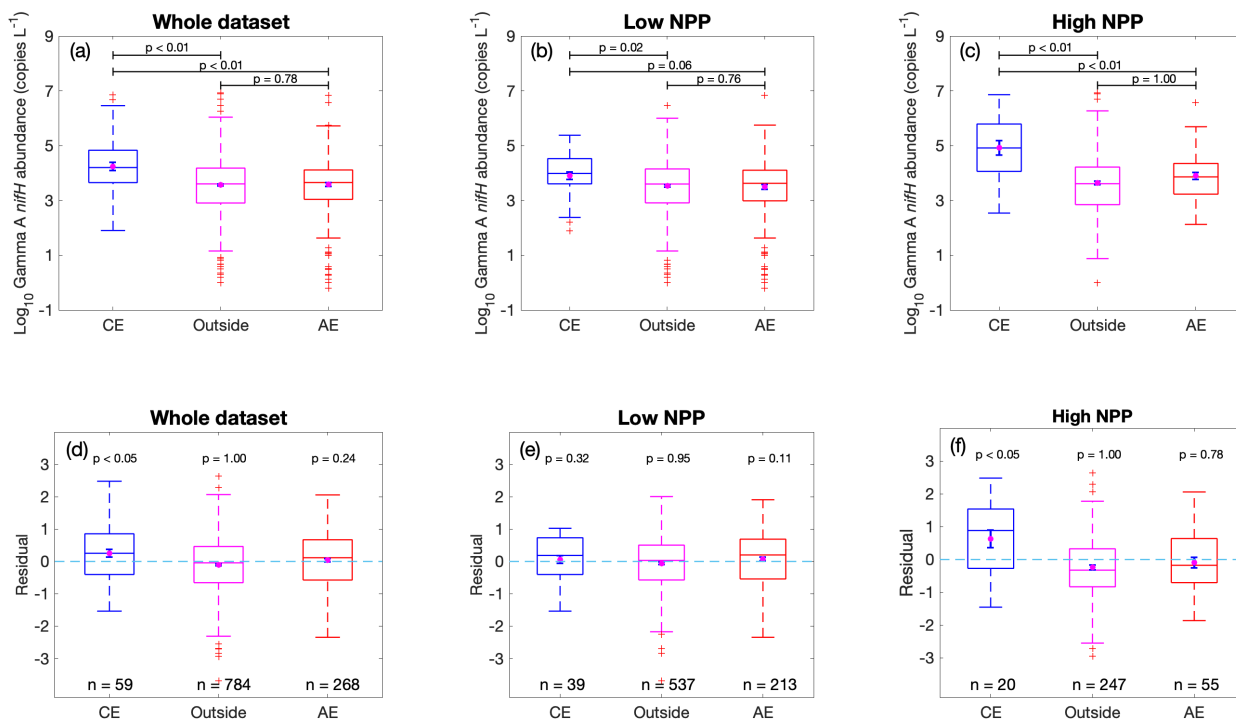



**Figure 6. Influence of mesoscale eddies on observed Gamma A abundance.** (a-c) Gamma A abundance and (d-f) residuals of predicted Gamma A abundance using climatological NPP and environmental factors in Fig. 4, grouped according to the data were sampled in cyclonic eddies (CE), anticyclonic eddies (AE) or outside eddies. The box plots show the median (central

355    line), 25$^{th}$ and 75$^{th}$ percentile of data (upper and lower edges of box), 5$^{th}$ and 95$^{th}$ percentile (error lines) and outliers (red crosses). The error bars within boxes show the mean value (purple dots) and its standard error. Values above brackets are p-values of two tailed t-test whether the means of observed Gamma A abundance are equal (a–c) or one-tailed t-test whether the residuals are greater than zero (d-e). The analyses are conducted for the whole dataset (a, d), for the data with local

climatological NPP lower than 400 mg m$^{-2}$ d$^{-1}$ (b, e) and for the data with local climatological NPP higher than 400 mg m$^{-2}$ d$^{-1}$ (c, f).

These results indicated that cyclonic eddies could stimulate Gamma A abundance, but only in the high productivity oceans. This finding is opposite to a previous hypothesis on autotrophic diazotrophs that anticyclonic eddies form a nitrate-depleted and well-lit environment favorable to N$_2$ fixation (Davis and Mcgillicuddy, 2006; Fong et al., 2008; Church et al., 2009; Liu et al., 2020). However, our results suggested that Gamma A can benefit from the vertical pumping of nutrition-rich water driven by cyclonic eddies (Mcgillicuddy et al., 1998). Nevertheless, the biogeochemical consequences of mesoscale eddies can be complex (Gaube et al., 2014; Mcgillicuddy Jr, 2016). For example, in addition to vertical pumping, the eddy stirring and trapping generated by mesoscale eddies can also have spatial effects on phytoplankton (Abraham, 1998; Wiebe and Joyce, 1992). Further sampling and studies are still needed to improve our mechanistic understanding of the effects of mesoscale eddies on both autotrophic cyanobacterial diazotrophs and NCDs.

### 3.5 Reliability of Gamma A *nifH* data

It is questionable whether the *nifH* copies measured using qPCR and collected in this study can reliably represent the abundance of Gamma A or even NCDs in general. When metadata are used, the reliability of comparison among absolute quantifications of *nifH* copies can be affected by methodological factors of qPCR assays. For example, even highly reproducible standard curves may result in significant variations in quantities of the same template in separated qPCR assays due to the log nature of the curve (Smith et al., 2006). The extraction method of nucleic acids, sample preparation, variations in the efficiencies of qPCR, and differences in the qPCR platform can also impact the quantitative results (Smith and Osborn, 2009). In addition, the copy numbers of the *nifH* gene in Gamma A's genome remain unknown. There exists a large uncertainty regarding the extent to which *nifH* gene copies can represent Gamma A abundance, especially in contrast to its autotrophic counterparts. All these problems will need better technology to be resolved in the future.

### 4. Summary and outlook

With more measurements becoming available, we explored in this study what factors controlled the distribution of a representative phylotype of non-cyanobacterial diazotrophs, Gamma A, in the global ocean. The results of our study did not fully agree with the conclusion of a previous study that Gamma A preferred warm oligotrophic oceans (Langlois et al. 2015). Instead, our study suggests that Gamma A tended to inhabit high-productivity waters. In addition, our analyses also found that Gamma A was more abundant in Fe-depleted areas, possibly to avoid competition with autotrophic diazotrophs in high-Fe environments. Overall, our study suggests that productivity and Fe can be factors differentiating niches between non-cyanobacterial and cyanobacterial diazotrophs in the ocean, with the former favoring a high productivity and low-Fe niche, while the latter occupies the opposite. Our study is also consistent with previous findings that Gamma A prefers well-lit surface

oceans. The phenomenon of elevated Gamma A abundance in cyclonic eddies found in this study is worth of further exploration for mechanistic understanding.

However, the moderate explanatory power of our prediction model indicates that there must be other unknown factors and mechanisms that also impact non-cyanobacterial diazotrophs. For instance, non-cyanobacterial diazotrophs found in the guts of copepods (Scavotto et al., 2015) imply that they are subject to top-down controls, which was also suggested for marine autotrophic diazotrophs (Landolfi et al., 2021; Wang et al., 2019; Wang and Luo, in press). Future studies should consider qPCR primer and probe sets targeting other NCDs such as Alphaproteobacteria and Cluster III phylotypes, which can also be important diazotrophs particularly in previously unrecognized regions for marine $N_2$ fixation (Wu et al., 2019; Langlois et al., 2008; Martínez-Pérez et al., 2018; Chen et al., 2019b). The combination of PCR amplification and metagenomic data can identify a broader NCD community (Delmont et al., 2018) and may help us design a better universal primer targeting major NCDs. Last, the uneven spatial samplings of Gamma A, particularly the relatively scarce samples in the Southern Hemisphere, may also introduce biases into our analyses. More samples and studies are needed in the future to improve our understanding of the controlling factors, niches and distributions for non-cyanobacterial diazotrophs, so that their contribution to global marine $N_2$ fixation can be better evaluated.

**Data availability**

All the data used in this study are available in a data repository (https://doi.org/10.6084/m9.figshare.17284517) (Shao and Luo, 2021).

**Author contributions**

Y.-W.L. conceived and supervised the study. Y.-W.L. and Z.S. designed the study. Z.S. collected and analyzed the data and drafted the first version of the manuscript. Y.-W.L. and Z.S. contributed to the discussion of the results and revised the manuscript.

**Competing interests**

The authors declare that they have no conflicts of interest.

**Acknowledgements**

415  The authors would like to thank the scientists and crew to sample the data that used in this study. The authors also thank Hua Wang, Yuhong Huang and Xiaoli Lu from XMU for their efforts in data collection and computation support. The authors would also thank Wupeng Xiao, Kuanbo Zhou and Xin Lin for the very useful discussions with them and three anonymous referees for their very constructive comments in improving the manuscript. This study was funded by the National Natural Science Foundation of China (42076153 and 41890802).

## References

Abraham, E. R.: The generation of plankton patchiness by turbulent stirring, Nature, 391, 577-580, https://doi.org/10.1038/35361, 1998.

Behrenfeld, M. J. and Falkowski, P. G.: A consumer's guide to phytoplankton primary productivity models, Limnol Oceanogr, 42, 1479-1491, https://doi.org/10.4319/lo.1997.42.7.1479, 1997.

Benavides, M. and Robidart, J.: Bridging the spatiotemporal gap in diazotroph activity and diversity with high-resolution measurements, 425  Front. Mar. Sci., 7, https://doi.org/10.3389/fmars.2020.568876, 2020.

Benavides, M., Bonnet, S., Berman-Frank, I., and Riemann, L.: Deep into oceanic $N_2$ fixation, Front. Mar. Sci., 5, 108, https://doi.org/10.3389/fmars.2018.00108, 2018.

Benavides, M., Moisander, P. H., Daley, M. C., Bode, A., and Aristegui, J.: Longitudinal variability of diazotroph abundances in the subtropical North Atlantic Ocean, J. Plankton Res., 38, 662-672, https://doi.org/10.1093/plankt/fbv121, 2016.

430  Bentzon-Tilia, M., Severin, I., Hansen, L. H., and Riemann, L.: Genomics and ecophysiology of heterotrophic nitrogen-fixing bacteria isolated from estuarine surface water, mBio, 6, 4, https://doi.org/10.1128/mBio.00929-15, 2015.

Berthelot, H., Benavides, M., Moisander, P. H., Grosso, O., and Bonnet, S.: High-nitrogen fixation rates in the particulate and dissolved pools in the Western Tropical Pacific (Solomon and Bismarck Seas), Geophys. Res. Lett., 44, 8414-8423, https://doi.org/10.1002/2017gl073856, 2017.

435  Bird, C. and Wyman, M.: Transcriptionally active heterotrophic diazotrophs are widespread in the upper water column of the Arabian Sea, FEMS Microbiology Ecology, 84, 189-200, https://doi.org/10.1111/1574-6941.12049, 2013.

Bird, C., Martinez, M. J., O'Donnell, A. G., and Wyman, M.: Spatial distribution and transcriptional activity of an uncultured clade of planktonic diazotrophic γ-proteobacteria in the Arabian Sea, Applied and Environmental Microbiology, 71, 2079-2085, https://doi.org/10.1128/AEM.71.4.2079-2085.2005, 2005.

440  Bombar, D., Paerl, R. W., and Riemann, L.: Marine non-cyanobacterial diazotrophs: moving beyond molecular detection, Trends Microbiol., 24, 916-927, https://doi.org/10.1016/j.tim.2016.07.002, 2016.

Bombar, D., Moisander, P. H., Dippner, J. W., Foster, R. A., Voss, M., Karfeld, B., and Zehr, J. P.: Distribution of diazotrophic microorganisms and *nifH* gene expression in the Mekong River plume during intermonsoon, Mar. Ecol. Prog. Ser., 424, 39-55, https://doi.org/10.3354/meps08976, 2011.

445 Bonnet, S., Dekaezemacker, J., Turk-Kubo, K. A., Moutin, T., Hamersley, R. M., Grosso, O., Zehr, J. P., and Capone, D. G.: Aphotic $N_2$ fixation in the Eastern Tropical South Pacific Ocean, PLoS One, 8, 14, https://doi.org/10.1371/journal.pone.0081265, 2013.

Bonnet, S., Guieu, C., Bruyant, F., Prášil, O., Van Wambeke, F., Raimbault, P., Moutin, T., Grob, C., Gorbunov, M. Y., Zehr, J. P., Masquelier, S. M., Garczarek, L., and Claustre, H.: Nutrient limitation of primary productivity in the Southeast Pacific (BIOSOPE cruise), Biogeosciences, 5, 215-225, https://doi.org/10.5194/bg-5-215-2008, 2008.

450 Boyd, P. W. and Ellwood, M. J.: The biogeochemical cycle of iron in the ocean, Nat Geosci, 3, 675-682, https://doi.org/10.1038/ngeo964, 2010.

Chelton, D. B., Schlax, M. G., and Samelson, R. M.: Global observations of nonlinear mesoscale eddies, Prog. Oceanogr., 91, 167-216, https://doi.org/10.1016/j.pocean.2011.01.002, 2011.

Chen, M. M., Lu, Y. Y., Jiao, N. Z., Tian, J. W., Kao, S. J., and Zhang, Y.: Biogeographic drivers of diazotrophs in the western Pacific
455 Ocean, Limnol Oceanogr, 64, 1403-1421, https://doi.org/10.1002/lno.11123, 2019a.

Chen, T. Y., Chen, Y. L. L., Sheu, D. S., Chen, H. Y., Lin, Y. H., and Shiozaki, T.: Community and abundance of heterotrophic diazotrophs in the northern South China Sea: Revealing the potential importance of a new alphaproteobacterium in $N_2$ fixation, Deep-Sea Res Pt I, 143, 104-114, https://doi.org/10.1016/j.dsr.2018.11.006, 2019b.

Cheung, S., Zehr, J. P., Xia, X., Tsurumoto, C., Endo, H., Nakaoka, S.-i., Mak, W., Suzuki, K., and Liu, H.: Gamma4: a genetically
460 versatile Gammaproteobacterial *nifH* phylotype that is widely distributed in the North Pacific Ocean, Environ. Microbiol., 23, 4246-4259, https://doi.org/10.1111/1462-2920.15604, 2021.

Cheung, S. Y., Nitanai, R., Tsurumoto, C., Endo, H., Nakaoka, S., Cheah, W., Lorda, J. F., Xia, X. M., Liu, H. B., and Suzuki, K.: Physical forcing controls the basin-scale occurrence of nitrogen-fixing organisms in the North Pacific Ocean, Global Biogeochem Cy, 34, 9, https://doi.org/10.1029/2019GB006452, 2020.

465 Church, M. J., Mahaffey, C., Letelier, R. M., Lukas, R., Zehr, J. P., and Karl, D. M.: Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre, Global Biogeochem Cy, 23, 2, https://doi.org/10.1029/2008gb003418, 2009.

Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nat. Commun., 11, 5444, https://doi.org/10.1038/s41467-020-19160-7, 2020.

Davis, C. S. and McGillicuddy, D. J., Jr.: Transatlantic abundance of the $N_2$-fixing colonial cyanobacterium *Trichodesmium*, Science, 312,
470 1517-1520, https://doi.org/10.1126/science.1123570, 2006.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, J Geophys Res-Oceans, 109, C12, https://doi.org/10.1029/2004JC002378, 2004.

Dekaezemacker, J., Bonnet, S., Grosso, O., Moutin, T., Bressac, M., and Capone, D. G.: Evidence of active dinitrogen fixation in surface waters of the eastern tropical South Pacific during El Nino and La Nina events and evaluation of its potential nutrient controls, Global
475    Biogeochem Cy, 27, 768-779, https://doi.org/10.1002/gbc.20063, 2013.

Delmont, T. O., Pierella Karlusich, J. J., Veseli, I., Fuessel, J., Eren, A. M., Foster, R. A., Bowler, C., Wincker, P., and Pelletier, E.: Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean, The ISME Journal, https://doi.org/10.1038/s41396-021-01135-1, 2021.

Delmont, T. O., Quince, C., Shaiber, A., Esen, O. C., Lee, S. T. M., Rappe, M. S., MacLellan, S. L., Lucker, S., and Eren, A. M.:
480    Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes, Nat. Microbiol, 3, 804-813, https://doi.org/10.1038/s41564-018-0176-9, 2018.

Deutsch, C., Sarmiento, J. L., Sigman, D. M., Gruber, N., and Dunne, J. P.: Spatial coupling of nitrogen inputs and losses in the ocean, Nature, 445, 163-167, https://doi.org/10.1038/nature05392, 2007.

Ding, C., Wu, C., Li, L., Pujari, L., Zhang, G., and Sun, J.: Comparison of diazotrophic composition and distribution in the South China
485    Sea and the Western Pacific Ocean, Biology, 10, 555, https://doi.org/10.3390/biology 10060555, 2021.

Farnelid, H., Tarangkoon, W., Hansen, G., Hansen, P. J., and Riemann, L.: Putative $N_2$-fixing heterotrophic bacteria associated with dinoflagellate–cyanobacteria consortia in the low-nitrogen Indian Ocean, Aquat Microb Ecol, 61, 105-117, https://doi.org/10.3354/ame01440, 2010.

Farnelid, H., Andersson, A. F., Bertilsson, S., Abu Al-Soud, W., Hansen, L. H., Sorensen, S., Steward, G. F., Hagstrom, A., and Riemann,
490    L.: Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria, PLoS One, 6, 9, https://doi.org/10.1371/journal.pone.0019223, 2011.

Fernandez, C., Farias, L., and Ulloa, O.: Nitrogen fixation in denitrified marine waters, PLoS One, 6, 6, https://doi.org/10.1371/journal.pone.0020539, 2011.

Fong, A. A., Karl, D. M., Lukas, R., Letelier, R. M., Zehr, J. P., and Church, M. J.: Nitrogen fixation in an anticyclonic eddy in the
495    oligotrophic North Pacific Ocean, ISME J, 2, 663-676, https://doi.org/10.1038/ismej.2008.22, 2008.

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, R., Zweng, M., Mishonov, A., Baranova, O., Seidov, D., and Reagan, J.: World Ocean Atlas 2018, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate). A. Mishonov Technical, NOAA Atlas NESDIS 84, 35 pp. [dataset], 2019.

Gaube, P., McGillicuddy Jr., D. J., Chelton, D. B., Behrenfeld, M. J., and Strutton, P. G.: Regional variations in the influence of mesoscale eddies on near-surface chlorophyll, J Geophys Res-Oceans, 119, 8195-8220, https://doi.org/10.1002/2014JC010111, 2014.

Geisler, E., Bogler, A., Bar-Zeev, E., and Rahav, E.: Heterotrophic nitrogen fixation at the hyper-eutrophic qshon river and estuary system, Front Microbiol, 11, 1370, https://doi.org/10.3389/fmicb.2020.01370, 2020.

Geisler, E., Bogler, A., Rahav, E., and Bar-Zeev, E.: Direct detection of heterotrophic diazotrophs associated with planktonic aggregates, Sci Rep-Uk, 9, 1-9, https://doi.org/10.1038/s41598-019-45505-4, 2019.

Glover, D. M., Jenkins, W. J., and Doney, S. C.: Modeling methods for marine science, Cambridge University Press, Cambridge, UK, 2011.

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

Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., D'Hondt, S., and Kuypers, M. M. M.: Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre, ISME J, 6, 1238-1249, https://doi.org/10.1038/ismej.2011.182, 2012.

Hamersley, M. R., Turk, K. A., Leinweber, A., Gruber, N., Zehr, J. P., Gunderson, T., and Capone, D. G.: Nitrogen fixation within the water column associated with two hypoxic basins in the Southern California Bight, Aquat Microb Ecol, 63, 193-205, https://doi.org/10.3354/ame01494, 2011.

Jayakumar, A. and Ward, B. B.: Diversity and distribution of nitrogen fixation genes in the oxygen minimum zones of the world oceans, Biogeosciences, 17, 5953-5966, https://doi.org/10.5194/bg-17-5953-2020, 2020.

Karl, D., Michaels, A., Bergman, B., Capone, D., Carpenter, E., Letelier, R., Lipschultz, F., Paerl, H., Sigman, D., and Stal, L.: Dinitrogen fixation in the world's oceans, Biogeochemistry, 57, 47-98, https://doi.org/10.1023/a:1015798105851, 2002.

Karl, D. M. and Letelier, R. M.: Nitrogen fixation-enhanced carbon sequestration in low nitrate, low chlorophyll seascapes, Mar. Ecol. Prog. Ser., 364, 257-268, https://doi.org/10.3354/meps07547, 2008.

Knapp, A. N., Casciotti, K. L., Berelson, W. M., Prokopenko, M. G., and Capone, D. G.: Low rates of nitrogen fixation in eastern tropical South Pacific surface waters, P Natl Acad Sci USA, 113, 4398-4403, https://doi.org/10.1073/pnas.1515641113, 2016.

Lam, P. and Kuypers, M. M. M.: Microbial nitrogen cycling processes in oxygen minimum zones, Ann Rev Mar Sci, 3, 317-345, https://doi.org/10.1146/annurev-marine-120709-142814, 2011.

Landolfi, A., Prowe, A. E. F., Pahlow, M., Somes, C. J., Chien, C.-T., Schartau, M., Koeve, W., and Oschlies, A.: Can top-down controls expand the ecological niche of marine $N_2$ fixers?, Front Microbiol, 12, 690200, https://doi.org/10.3389/fmicb.2021.690200, 2021.

Langlois, R., Grokopf, T., Mills, M., Takeda, S., and LaRoche, J.: Widespread distribution and expression of Gamma A (UMB), an uncultured, diazotrophic, gamma-proteobacterial *nifH* phylotype, PLoS One, 10, 17, https://doi.org/10.1371/journal.pone.0128912, 2015.

Langlois, R. J., Hummer, D., and LaRoche, J.: Abundances and distributions of the dominant *nifH* phylotypes in the Northern Atlantic Ocean, Appl Environ Microbiol, 74, 1922-1931, https://doi.org/10.1128/AEM.01720-07, 2008.

530  Lee, Z. P., Du, K. P., Arnone, R., Liew, S. C., and Penta, B.: Penetration of solar radiation in the upper ocean: A numerical model for oceanic and coastal waters, J Geophys Res-Oceans, 110, C09019, https://doi.org/10.1029/2004jc002780, 2005.

Li, D., Jing, H., Zhang, R., Yang, W., Chen, M., Wang, B., Zheng, M., and Qiu, Y.: Heterotrophic diazotrophs in a eutrophic temperate bay (Jiaozhou Bay) broadens the domain of $N_2$ fixation in China's coastal waters, Estuar Coast Mar Sci, 242, 106778, https://doi.org/10.1016/j.ecss.2020.106778, 2020.

535  Liu, J. X., Zhou, L. B., Li, J. J., Lin, Y. Y., Ke, Z. X., Zhao, C. Y., Liu, H. J., Jiang, X., He, Y. H., and Tan, Y. H.: Effect of mesoscale eddies on diazotroph community structure and nitrogen fixation rates in the South China Sea, Reg Stud Mar Sci, 35, 14, https://doi.org/10.1016/j.rsma.2020.101106, 2020.

Locarnini, M., Mishonov, A., Baranova, O., Boyer, T., Zweng, M., Garcia, H., Reagan, J., Seidov, D., Weathers, K., Paver, C., and Smolyar, I.: World Ocean Atlas 2018, Volume 1: Temperature.A. Mishonov Technical Ed., NOAA Atlas NESDIS 81, 52pp. [dataset],
540  2018.

Loescher, C. R., Großkopf, T., Desai, F. D., Gill, D., Schunck, H., Croot, P. L., Schlosser, C., Neulinger, S. C., Pinnow, N., Lavik, G., Kuypers, M. M. M., LaRoche, J., and Schmitz, R. A.: Facets of diazotrophy in the oxygen minimum zone waters off Peru, ISME J, 8, 2180-2192, https://doi.org/10.1038/ismej.2014.71, 2014.

Luo, Y. W., Lima, I. D., Karl, D. M., Deutsch, C. A., and Doney, S. C.: Data-based assessment of environmental controls on global marine
545  nitrogen fixation, Biogeosciences, 11, 691-708, https://doi.org/10.5194/bg-11-691-2014, 2014.

Marra, G. and Wood, S. N.: Practical variable selection for generalized additive models, Comput Stat Data An, 55, 2372-2387, https://doi.org/10.1016/j.csda.2011.02.004, 2011.

Martínez, L., Silver, M. W., King, J. M., and Alldredge, A. L.: Nitrogen fixation by floating diatom mats: A source of new nitrogen to oligotrophic ocean waters, Science, 221, 152-154, https://doi.org/10.1126/science.221.4606.152, 1983.

550  Martínez-Pérez, C., Mohr, W., Schwedt, A., Dürschlag, J., Callbeck, C. M., Schunck, H., Dekaezemacker, J., Buckner, C. R., Lavik, G., and Fuchs, B. M.: Metabolic versatility of a novel $N_2$-fixing Alphaproteobacterium isolated from a marine oxygen minimum zone, Environ. Microbiol., 20, 755-768, https://doi.org/10.1111/1462-2920.14008, 2018.

Martínez-Pérez, C., Mohr, W., Loscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B. M., Lavik, G., Schmitz, R. A., LaRoche, J., and Kuypers, M. M.: The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle, Nat Microbiol, 1, 16163, https://doi.org/10.1038/nmicrobiol.2016.163, 2016.

McGillicuddy, D. J., Robinson, A. R., Siegel, D. A., Jannasch, H. W., Johnson, R., Dickey, T. D., McNeil, J., Michaels, A. F., and Knap, A. H.: Influence of mesoscale eddies on new production in the Sargasso Sea, Nature, 394, 263-266, https://doi.org/10.1038/28367, 1998.

McGillicuddy Jr, D. J.: Mechanisms of physical-biological-biogeochemical interaction at the oceanic mesoscale, Ann Rev Mar Sci, 8, 125-159, https://doi.org/10.1146/annurev-marine-010814-015606, 2016.

Misumi, K., Lindsay, K., Moore, J. K., Doney, S. C., Bryan, F. O., Tsumune, D., and Yoshida, Y.: The iron budget in ocean surface waters in the 20th and 21st centuries: projections by the Community Earth System Model version 1, Biogeosciences, 11, 33-55, https://doi.org/10.5194/bg-11-33-2014, 2014.

Moisander, P. H., Beinart, R. A., Voss, M., and Zehr, J. P.: Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon, ISME J, 2, 954-967, https://doi.org/10.1038/ismej.2008.51, 2008.

Moisander, P. H., Serros, T., Paerl, R. W., Beinart, R. A., and Zehr, J. P.: Gammaproteobacterial diazotrophs and *nifH* gene expression in surface waters of the South Pacific Ocean, ISME J, 8, 1962-1973, https://doi.org/10.1038/ismej.2014.49, 2014.

Moisander, P. H., Benavides, M., Bonnet, S., Berman-Frank, I., White, A. E., and Riemann, L.: Chasing after non-cyanobacterial nitrogen fixation in marine pelagic environments, Front Microbiol, 8, 1736, https://doi.org/10.3389/fmicb.2017.01736, 2017.

Moore, R. M., Grefe, I., Zorz, J., Shan, S., Thompson, K., Ratten, J., and LaRoche, J.: On the relationship between hydrogen saturation in the tropical Atlantic Ocean and nitrogen fixation by the symbiotic diazotroph UCYN-A, J Geophys Res-Oceans, 123, 2353-2362, https://doi.org/10.1002/2017jc013047, 2018.

Moreira-Coello, V., Mourino-Carballido, B., Maranon, E., Fernandez-Carrera, A., Bode, A., and Varela, M. M.: Biological $N_2$ fixation in the upwelling region off NW Iberia: magnitude, relevance, and players, Front. Mar. Sci., 4, 303, https://doi.org/10.3389/fmars.2017.00303, 2017.

Pedersen, J. N., Bombar, D., Paerl, R. W., and Riemann, L.: Diazotrophs and $N_2$-fixation associated with particles in coastal estuarine waters, Front Microbiol, 9, 2759, https://doi.org/10.3389/fmicb.2018.02759, 2018.

Rahav, E., Giannetto, M. J., and Bar-Zeev, E.: Contribution of mono and polysaccharides to heterotrophic $N_2$ fixation at the eastern Mediterranean coastline, Sci Rep-Uk, 6, 27858, https://doi.org/10.1038/srep27858, 2016.

Rahav, E., Herut, B., Mulholland, M. R., Belkin, N., Elifantz, H., and Berman-Frank, I.: Heterotrophic and autotrophic contribution to dinitrogen fixation in the Gulf of Aqaba, Mar. Ecol. Prog. Ser., 522, 67-77, https://doi.org/10.3354/meps11143, 2015.

Rahav, E., Bar-Zeev, E., Ohayon, S., Elifantz, H., Belkin, N., Herut, B., Mulholland, M. R., and Berman-Frank, I.: Dinitrogen fixation in aphotic oxygenated marine environments, Front Microbiol, 4, 11, https://doi.org/10.3389/fmicb.2013.00227, 2013.

Riemann, L., Farnelid, H., and Steward, G. F.: Nitrogenase genes in non-cyanobacterial plankton: prevalence, diversity and regulation in marine waters, Aquat Microb Ecol, 61, 225-237, https://doi.org/10.3354/ame01431, 2010.

585 Roshan, S. and DeVries, T.: Efficient dissolved organic carbon production and export in the oligotrophic ocean, Nat. Commun., 8, 2036, https://doi.org/10.1038/s41467-017-02227-3, 2017.

Sargent, E. C., Hitchcock, A., Johansson, S. A., Langlois, R., Moore, C. M., LaRoche, J., Poulton, A. J., and Bibby, T. S.: Evidence for polyploidy in the globally important diazotroph *Trichodesmium*, FEMS microbiology letters, 363, fnw244, https://doi.org/10.1093/femsle/fnw244, 2016.

590 Scavotto, R. E., Dziallas, C., Bentzon-Tilia, M., Riemann, L., and Moisander, P. H.: Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean, Environ. Microbiol., 17, 3754-3765, https://doi.org/10.1111/1462-2920.12777, 2015.

Shao, Z. and Luo, Y.-W.: Gamma A *nifH* abundance in the global ocean, figshare [data set], https://doi.org/10.6084/m9.figshare.17284517, 2021.

Shiozaki, T., Kondo, Y., Yuasa, D., and Takeda, S.: Distribution of major diazotrophs in the surface water of the Kuroshio from 595 northeastern Taiwan to south of mainland Japan, J. Plankton Res., 40, 407-419, https://doi.org/10.1093/plankt/fby027, 2018a.

Shiozaki, T., Ijichi, M., Kodama, T., Takeda, S., and Furuya, K.: Heterotrophic bacteria as major nitrogen fixers in the euphotic zone of the Indian Ocean, Global Biogeochem Cy, 28, 1096-1110, https://doi.org/10.1002/2014gb004886, 2014.

Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., Ehama, M., Hamasaki, K., and Furuya, K.: Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea, Global 600 Biogeochem Cy, 31, 996-1009, https://doi.org/10.1002/2017gb005681, 2017.

Shiozaki, T., Bombar, D., Riemann, L., Sato, M., Hashihama, F., Kodama, T., Tanita, I., Takeda, S., Saito, H., Hamasaki, K., and Furuya, K.: Linkage between dinitrogen fixation and primary production in the oligotrophic South Pacific Ocean, Global Biogeochem Cy, 32, 1028-1044, https://doi.org/10.1029/2017gb005869, 2018b.

Smith, C. J. and Osborn, A. M.: Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology, FEMS 605 Microbiology Ecology, 67, 6-20, https://doi.org/10.1111/j.1574-6941.2008.00629.x, 2009.

Smith, C. J., Nedwell, D. B., Dong, L. F., and Osborn, A. M.: Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples, Environ. Microbiol., 8, 804-815, https://doi.org/10.1111/j.1462-2920.2005.00963.x, 2006.

Sohm, J. A., Webb, E. A., and Capone, D. G.: Emerging patterns of marine nitrogen fixation, Nat. Rev. Microbiol., 9, 499-508, https://doi.org/10.1038/nrmicro2594, 2011.

Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L., Dugdale, R., Finkel, Z., Iudicone, D., Jahn, O., Guidi, L., Lasbleiz, M., Leblanc, K., Levy, M., and Pondaven, P.: Influence of diatom diversity on the ocean biological carbon pump, Nat Geosci, 11, 27-37, https://doi.org/10.1038/s41561-017-0028-x, 2018.

Turk-Kubo, K. A., Karamchandani, M., Capone, D. G., and Zehr, J. P.: The paradox of marine heterotrophic nitrogen fixation: abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific, Environ Microbiol, 16, 3095-3114, https://doi.org/10.1111/1462-2920.12346, 2014.

Wang, H. and Luo, Y.-W.: Top-down control on major groups of global marine diazotrophs, Acta Oceanol. Sin., in press.

Wang, W. L., Moore, J. K., Martiny, A. C., and Primeau, F. W.: Convergent estimates of marine nitrogen fixation, Nature, 566, 205-213, https://doi.org/10.1038/s41586-019-0911-2, 2019.

White, A. E., Watkins-Brandt, K. S., and Church, M. J.: Temporal variability of *Trichodesmium* spp. and diatom-diazotroph assemblages in the North Pacific subtropical gyre, Front. Mar. Sci., 5, 27, https://doi.org/10.3389/fmars.2018.00027, 2018.

Wiebe, P. H. and Joyce, T. M.: Introduction to interdisciplinary studies of Kuroshio and Gulf Stream rings, Deep Sea Research A, 39, v-vi, https://doi.org/10.1016/S0198-0149(11)80001-4, 1992.

Wood, S. N.: Generalized additive models: an introduction with R, CRC press, 2017.

Wood, S. N., Pya, N., and Säfken, B.: Smoothing parameter and model selection for general smooth models, J Am Stat Assoc, 111, 1548-1563, https://doi.org/10.1080/01621459.2016.1180986, 2016.

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

Zehr, J. P.: Nitrogen fixation by marine cyanobacteria, Trends Microbiol., 19, 162-173, https://doi.org/10.1016/j.tim.2010.12.004, 2011.

Zehr, J. P. and Capone, D. G.: Changing perspectives in marine nitrogen fixation, Science, 368, 729-+, https://doi.org/10.1126/science.aay9514, 2020.

Zehr, J. P. and Kudela, R. M.: Nitrogen cycle of the open ocean: from genes to ecosystems, Ann Rev Mar Sci, 3, 197-225, https://doi.org/10.1146/annurev-marine-120709-142819, 2011.

635     Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F.: Nitrogenase gene diversity and microbial community structure: a cross-system comparison, Environ. Microbiol., 5, 539-554, https://doi.org/10.1046/j.1462-2920.2003.00451.x, 2003.

Zhang, Y., Zhao, Z., Sun, J., and Jiao, N.: Diversity and distribution of diazotrophic communities in the South China Sea deep basin with mesoscale cyclonic eddy perturbations, FEMS Microbiology Ecology, 78, 417-427, https://doi.org/10.1111/j.1574-6941.2011.01174.x, 2011.

640