Response to the comments from Anonymous Referee #2

We thank Referee #2 for his/her efforts and provide very constructive comments that greatly helped us correct errors and improve the quality of our manuscript. We have responded (in blue fonts) to the comments point by point and revised the manuscript accordingly.

Shao and Luo attempt to better constrain the environmental drivers behind the observed biogeography of gamma A, a cosmopolitan marine non-cyanobacterial diazotroph group, using a metadata analysis of previously published gamma A abundances (estimated using qPCR targeting the gamma A nifH gene along with a suite of environmental parameters derived from the world ocean atlas, MODIS and several model outputs. On this whole, this represents a valid and interesting approach to gain insight into gamma A, but there are many items I feel need to be addressed prior to being considered for publication.

In general, I am concerned with ignoring all the apparent 0s in the compiled qPCR dataset. Better justification is needed for doing this – including why we should assume that abundance data would be normally distributed (in my experience with this type of data, it certainly isn't always), and better justification for the stated assumption that these undetects are not true 0s due to primer specificity. Is there any precedent for ignoring 0s in other published work that uses GAMs or other similar analyses?

Response: We thank the referee to bring up this very import issue. After carefully considering the referee's comments, we now agree that a large fraction of the zero-value data of Gamma A nifH copies were true zeros: The non-zero Gamma A abundance data were approximately log-normally distributed as shown in Fig. S1. Because the detection limits for nifH abundance usually ranges from 10^1 to 10^2 copies L^{-1}, the number of data that were not true zero but were below detection was very likely no more than 72, assuming the detection limit was 10^2 copies L^{-1} (Fig. S1). Therefore, the fact that there were far more zero data points (682) in our dataset indicated a large fraction of zero data could represent true absence of Gamma A.

Additionally, based on the above analyses that many zeros represent true absence, we now also agree with the referee that the Gamma A is patch in space and time. The patchiness of diazotrophs, as suggested by the study recommended by the referee (Robidart et al., 2014), can be a consequence of lateral transport and mixing of water masses. The patchiness of Gamma A was also supported by the facts that many non-zero and zero Gamma A data were spatially close to each other (Fig. 1) and by our new analyses in the revised manuscript (new Fig. S2), showing that the environmental conditions of the non-zero and the zero Gamma A data largely overlapped.

In the revised manuscript, we still decided not to include zero-value data in the statistical analyses. The first reason was the patchiness of Gamma A distribution, which implicated that Gamma A can be either present or absent even when the environmental conditions we analyzed in this study were suitable for

Gamma A. That is, the presence of Gamma A needs a suitable environment, but a suitable environment does not necessarily guarantee the presence of Gamma A. If the zero-value data were included otherwise, similar environmental conditions could associate with both substantial and zero abundance, which would bias the response function of our statistical analyses, particularly as the fraction of zero-value data was large (~1/3) in our Gamma A dataset. Another reason was that we cannot identify true or false zeros of the Gamma A data, particularly considering the accuracy of qPCR that was highly sensitive to sample preservation, extraction protocol and the reliance of the standard curve (Smith and Osborn, 2009).

We found several marine ecological data analyses also removed zero-value abundance data and only used presence data (Irwin et al., 2012; Xiao et al., 2019). Their main reasons are similar: (1) reliability of zero data highly depends on the difficulty in species detection, and (2) large fraction of zeros would bias the response function of commonly used statistical analysis.

Therefore, we have revised the reasoning why the zero-value Gamma A abundance data were not included in the GAM (Method Section 2.1), added the description of the zero values and compared the environmental conditions associated to zero and non-zero data (Results Section 3.1), and revised discussion on the reliability of Gamma A *nifH* data (Section 3.6).

Method (Section 2.1):

"The non-zero abundance data was approximately log-normal distribution (Fig. S1). There were 682 data points reporting zero *nifH* copies which theoretically could indicate that Gamma A was either true absent or its abundance was below the detection limit in the samples. As the reported detection limit of qPCR usually ranges from 10^1 to 10^2 copies L^{-1}, the number of data that were below detection, according to log-normal distribution of observed non-zero data, was very likely no more than 72 even assuming a large detection limit of 10^2 copies L^{-1} (Fig. S1). The fact that there were far more zero data (682) in our dataset indicated a large fraction of zero data could represent true absence of Gamma A. Therefore, the distribution of Gamma A could be patchy, which was also confirmed by the mixed spatial distribution of the zero and non-zero data (see Results). The patchiness of diazotrophs in a small temporal and spatial scale has been widely found 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. That is, the presence of Gamma A needs a suitable environment, but a suitable environment does not necessarily guarantee the presence of Gamma A. Partly for this reason, the zero-value abundance data of Gamma A were not included in our further analyses. If the zero-value data were included otherwise, similar environmental conditions could associate with both substantial and zero abundance (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) in all the Gamma A data. Another reason why the zero-abundance data were not included was that, considering the accuracy of qPCR was highly sensitive to sample preservation, extraction protocol and the reliance of the standard curve (Smith and Osborn, 2009), it was difficult to identify whether the zero values represented true absence or below-detection abundance of Gamma A."

Results (Section 3.1):

"Although high Gamma A abundance over 10^6 *nifH* copies L^{-1} was observed in surface North Pacific Ocean, zero-value data were also massive (215 in total 608 data points) and even located close to those high-abundance data (Cheung et al., 2020) (Fig.1), indicating the patchy distribution of Gamma A. As discussed already, 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 true absence of Gamma A."

Results (Section 3.4.7):

"It was interesting that although Gamma A was undetected in the all samples in the South Pacific Gyre (Fig. 1) and all these zero-value data were not included in our GAM analyses, the prediction still showed the lowest Gamma A in this region (Fig. 6a), partly supporting the robustness of our prediction on Gamma A. However, another study suggested that NCDs were major players of N_2 fixation in SPG (Halm et al., 2012), which could reflect a possibility that Gamma A may not always be the dominant NCD phylotype in the ocean. For example, Gamma 4 was suggested as a more versatile NCD phylotype in north Pacific Ocean (Cheung et al., 2021)."

Results (Section 3.6):

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 quantification can be affected by methodological factors. For example, even highly reproducible standard curves may result in significant variations in quantities of the same template in separated qPCR assays (Smith et al., 2006) due to the log nature of the curve. Extraction method of nucleic acid, sample preparation, variations in the efficiencies of the qPCR, differences in the qPCR platform can also impact the quantitative results (Smith et al., 2009). In addition, the copy numbers of *nifH* gene in Gamma A's genome remains unknown. There existed a large uncertainty that to what extend *nifH* gene copies can represent Gamma A abundance, especially in contrast to its autotrophic counterparts. All these problems will need better technology to resolve in the future.

Reference

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.

Cheung, S., Zehr, J. P., Xia, X., Tsurumoto, C., Endo, H., Nakaoka, S.-i., Mak, W., Suzuki, K., and Liu, H.: Gamma4: a genetically 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.

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.

Irwin, A. J., Nelles, A. M., and Finkel, Z. V.: Phytoplankton niches estimated from field data, Limnol Oceanogr, 57, 787-797, https://doi.org/10.4319/lo.2012.57.3.0787, 2012.

Robidart, J. C., Church, M. J., Ryan, J. P., Ascani, F., Wilson, S. T., Bombar, D., Marin, R., Richards, K. J., Karl, D. M., Scholin, C. A., and Zehr, J. P.: Ecogenomic sensor reveals controls on N2-fixing microorganisms in the North Pacific Ocean, ISME J, 8, 1175-1185, 10.1038/ismej.2013.244, 2014.

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, 2006.

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

Xiao, W. P., Wang, L., Laws, E., Xie, Y. Y., Chen, J. X., Liu, X., Chen, B. Z., and Huang, B. Q.: Realized niches explain spatial gradients in seasonal abundance of phytoplankton groups in the South China Sea, Prog. Oceanogr., 162, 223-239, https://doi.org/10.1016/j.pocean.2018.03.008, 2018.

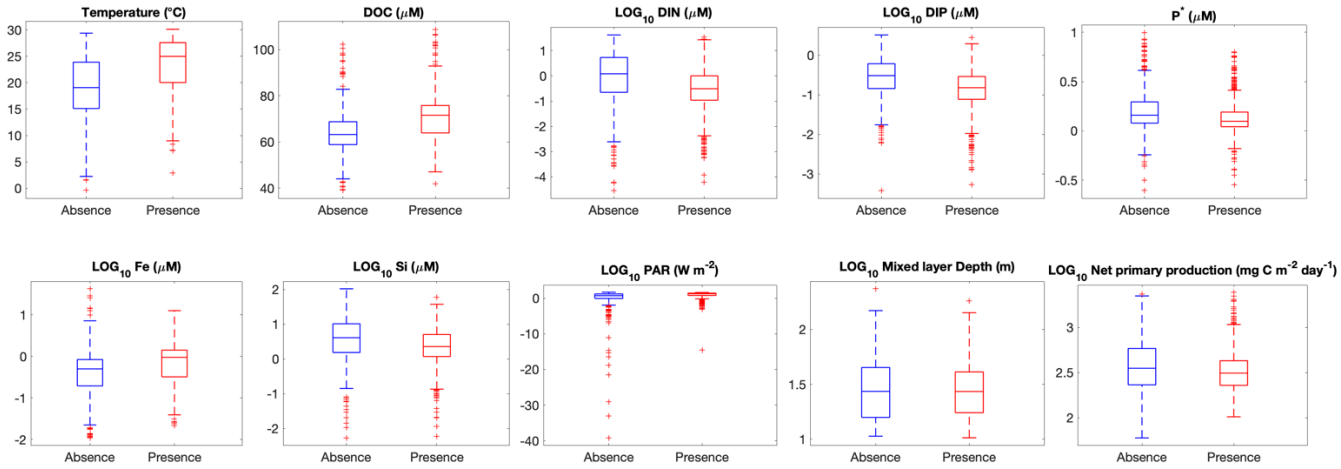


Figure S2. Environmental conditions of observed Gamma A absence and presence data. Absence consists of zero data, UD (under detection) and ND (no detected) in dataset.

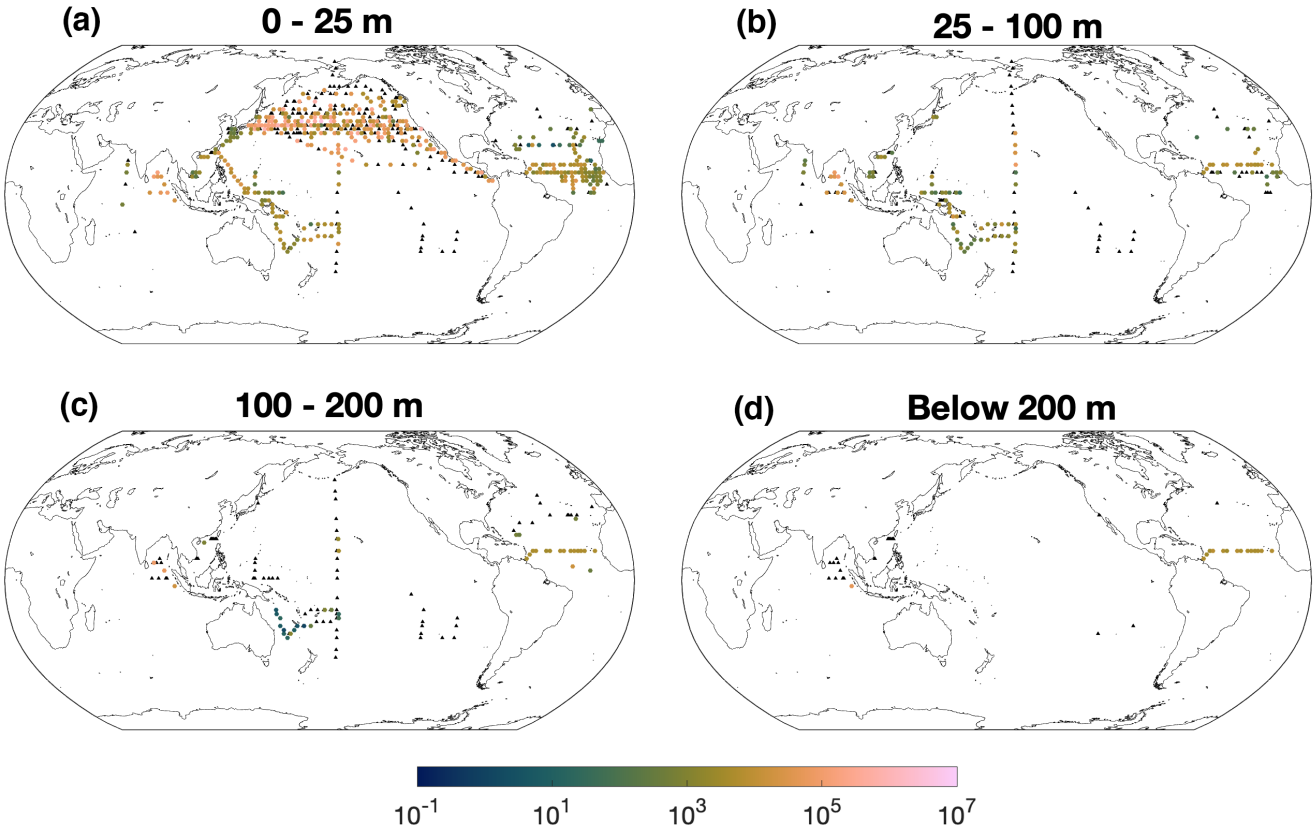


Figure 1. Gamma A abundance including zero-value points (*nifH* copies L-1). The panels show data in depth ranges of (a) 0-25 m, (b) 25-100 m, (c) 100-200 m and (d) below 200 m. For clear demonstration, data are binned to 2° × 2° and geometric means in each bin are shown. Zero data were denoted as black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

I also find much of the discussion to be speculative, esp. when trying to relate these findings to the broader group of gamma proteobacterial diazotrophs, or NCDs in their entirety.

Response: Thanks for your comment. We revised our manuscript and tried to remove speculative discussions (see more details below).

I suggest sticking with non-cyanobacterial diazotrophs throughout in place of “heterotrophic diazotrophs”.

Response: Replacement has been done.

Specific Comments

Line 7 – First sentence is awkward. Perhaps "non-cyanobacterial diazotrophs (NCDs) may be contributors to global marine...."

Response: We have rephrased the sentence as "Non-cyanobacterial diazotrophs (NCDs) may be contributors to global marine N_2 fixation, "

Line 10 – This needs definition since this is not a commonly used term for this sort of data. Is this even the right term to be used here and throughout? Aren't you really talking simply about abundance?

Response: Thanks for your comments. We represented Gamma A abundance using its *nifH* copies in this study. We will define this term in the abstract as "First, we represented Gamma A abundance by its *nifH* qPCR copies reported in literature, and analyzed its relationship to climatological biological and environmental conditions."

Line 15 – because the GLMs only explain some of the variance in gamma A abundances, I suggest using less definite terms here and throughout, e.g. in line 18 "most likely determined by" to "influenced by", etc.

Response: Thanks for your suggestion. We have found some place with definite term and revised them accordingly.

Line 75 "Our analyses revealed 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." to "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."

Line 158 "**Primary production determines the maximal Gamma A abundance" to "Primary production supports the maximal Gamma A abundance"**

Line 170 "These results indicated that local NPP could largely determine the carrying capacity of Gamma A abundance, …" to "These results indicated that local NPP could largely support the maximal observed Gamma A abundance,…"

Line 262 "Our GAM results also revealed a positive relationship between silicate and $\Delta_{Gamma\text{-}A}$ in both the low- and the high-NPP groups (Figs. 4f and 4n)" to "Our GAM results also suggested a positive relationship between silicate and $\Delta_{Gamma\text{-}A}$ in both the low- and the high-NPP groups (Figs. 4f and 4n)"

Line 371 "In addition, our analyses also revealed that Gamma A was more abundant in Fe-depleted areas, possibly to avoid competition with autotrophic diazotrophs in high-Fe environments" to "In addition, our analyses also suggested that Gamma A was more abundant in Fe-depleted areas, possibly to avoid competition with autotrophic diazotrophs in high-Fe environments"

Line 17 – "mesoscale" in place of "short-term"

Response: Corrected.

Line 18 – "matter" in place of "matters

Response: Corrected.

19 – "provide insight into" in place of "insight a"

Response: Corrected.

Line 25 – remove heterotrophic here

Response: Corrected.

Line 26 – these aren't the best papers to cite here

Response: we have updated the citations here.

"non-cyanobacterial diazotrophs (NCDs) 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)"

Reference:

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.

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.

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.

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.

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.

Line 28 – "had higher relative abundances than" in place of "were far superior in number to"

Response: Corrected.

Line 30 – remove "dominant" or find a way to rephrase

Response: we have changed dominant to abundant. "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)."

Line 33 – rephrase "heavy"

Response: We have changed "heavy" to "significant"

Line 33-36 – Marine N2 fixation by NCDs is not quantified at all – please rephrase and make it clear that there is only indirect evidence, including nifH transcription which does not "support" active N2 fixation by NCDs at all, it only provides another line of indirect evidence.

Response: Thanks for your comments. We have rephrased the sentence as "Although the N_2 fixed by non-cyanobacterial diazotrophs 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 non-cyanobacterial N_2 fixation in the ocean."

Line 36-37 – in addition to being uncultivated there are likely diverse niches and metabolic strategies used by this broad group. I might move the paragraph beginning at line 64 up, so that you can make this point and introduce gamma A earlier.

Response: Thanks for your comments. We agree and have revised.

Line 38 – "Apparently" is awkward here – remove. Add "presumably" before "depending"

Response: Corrected.

Line 41 – this is misleading – there was no gamma A in Benavides et al. 2018, and they did not assess other NCDs in this study

Response: Thank you for your comment. In Benavides et al. 2018b, abundant Gamma A *nifH* DNA copies were detected（Table 1of the paper). It was the expression of Gamma A *nifH* gene that was not detected in any DOM addition experiment (including controlled group). The authors also mentioned that the expression of Gamma A nifH was not detected "Despite being the most abundant ambient group as determined by DNA qPCR counts" (page 6).
Therefore, we have rephrased this sentence as "However, DOM addition sometimes did not stimulate *nifH* expression of Gamma A even when its DNA copies was ambient (Benavides et al., 2018b), implying DOM may not always stimulate the activity of Gamma A.

Line 46 – Bonnet et al., citation makes no sense here.

Response: We meant that NCD may be like other cyanobacterial diazotrophs that they can provide N to diatoms. But we agree with the referee that this statement is too speculative. We then has revised the sentence as: "NCDs were also detected in diatom mats (Martínez et al., 1983), implying another novel habitat for NCDs".

Line 46 – "equip" is awkward

Response: Thanks for this comment. We have changed "equip with" to "contain".

Line 55 – is "supposably" needed here?

Response: Thanks for this comment. We have deleted "supposably" here.

Line 69 – as above, gamma A nifH transcription doesn't "reveal" it's important role in marine N2 fix

Response: Thanks for this comment. We have revised to "suggesting its role in marine N_2 fixation".

Line 71 – state here that this data is compiled from nifH-based qPCR studies.

Response: Thanks for this comment. In this study, we collected, to our best knowledge, all the reported in situ measurements of Gamma A nifH copies. We would restate this sentence as "With more data becoming available in the recent years, we collected, to our best knowledge, all the reported in situ measurements of Gamma A *nifH* copies using qPCR assays, …"

Line 85 – zero nifH copies can also be true zeros

Response: Thanks for this comment. Please see our response to the general comments. We agree and have made necessary revisions.

Line 87 – all studies have different detection limits based on filter volumes, extraction volumes, the amount of template used in the qPCR, etc. This is misleading.

Response: Thanks for this comment. Yes, We agree there is no common qPCR detection limit. Usually the detection limit ranges from 10^1 to 10^2 copies L^{-1}. We have revised to:

As the reported detection limit of qPCR usually ranges from 10^1 to 10^2 copies L^{-1}, the number of data that were below detection, according to log-normal distribution of observed non-zero data, was very likely no more than 72 even assuming a large detection limit of 10^2 copies L^{-1} (Fig. S1). The fact that there were far more zero data (682) in our dataset indicated a large fraction of zero data could represent true absence of Gamma A.

Please see our response to the general comments for more related details.

Line 95 – replace "were" with "have been" and you should note that these studies are specific to cyanobacterial diazos, and we do not know gene copy #s in gamma A.

Response: Thanks for this comment. We have revised the texts as follows:

"In the following analyses, we represented Gamma A abundance using its *nifH* copies, although we noted that variations in *nifH* copies in different cyanobacterial diazotroph cells have been reported (White et al., 2018; Sargent et al., 2016) and *nifH* copy numbers in Gamma A genome remain unknown.

Table 1 – There are additional studies represented in the Figure S5, it seems? These should be listed in Supp.

Response: Thanks for your comment. All the studies we used in our manuscripts have been listed in Table 1. Fig. S5 only included zero-value data comparing to Fig. 1. (In the revised manuscript, Fig.1 has also included zero-value data and Fig. S5 has been removed.)

Line 118 – More description needed about how cyclonic and anticyclonic eddies were called. What does a "clear shape" mean? Why is SLA missing from Table 2?

Response: Thanks for your comment. We define the core of mesoscale eddy as where the outermost closed contour line of the SLA field is. If a sampling point located in the eddy core, we recorded it as within anticyclonic eddy (positive SLA) or cyclonic eddy (negative SLA). We have revised this paragraph as:

"To identify if 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 cyclonic (negative SLA) and

anticyclonic (positive SLA) eddies cores were recorded. Otherwise, data points were recorded as 'outside the eddy'."

We have also added data source of SLA in Table 2.

Line 148 – maybe specify it was undetected in this SPOT sample?

Response: Thanks for your comment. We have checked the reported data, Gamma A was detected but not quantified in this SPOT sample. The deepest datum was sampled at 1700m in South China Sea, but Gamma A nifH was undetected. Therefore, we have changed this sentence into:
"The deepest datum with detectable Gamma A nifH was sampled at 885 m in Southern California Bight (Hamersley et al., 2011)."

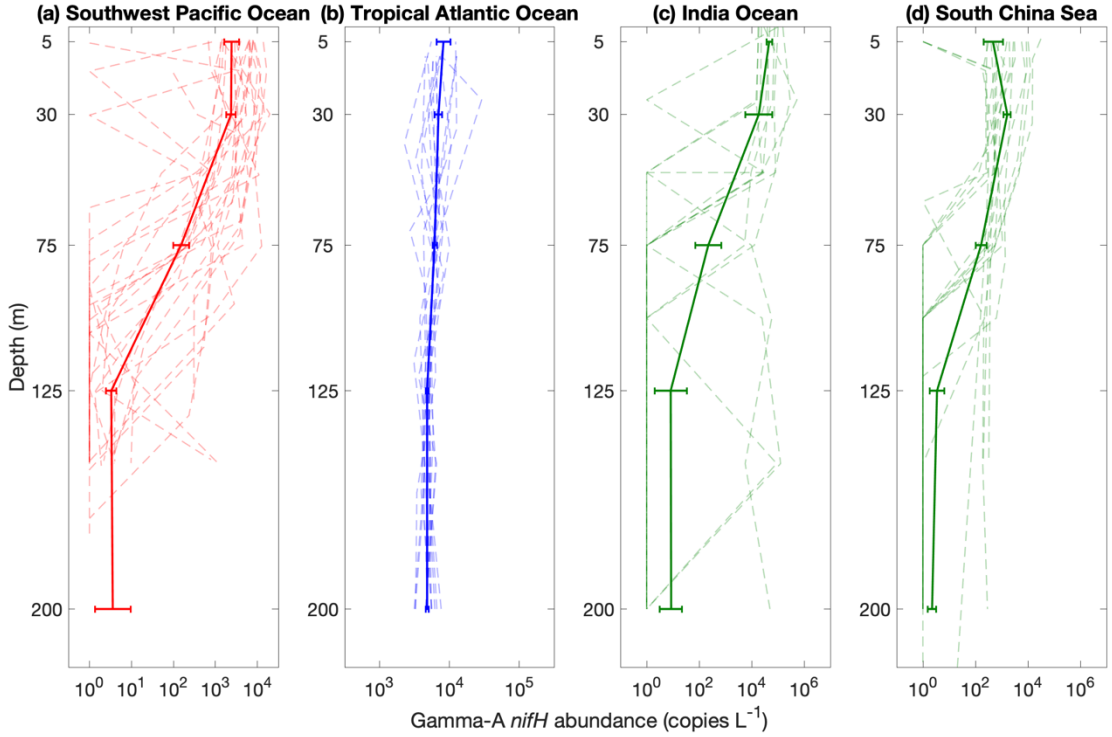
Line 150 – there are other studies that describe the depth distribution patterns of gamma A, eg. Chen et al., 2019, which seems to be missing from your list of studies???

Chen, Tien-Yi, et al. "Community and abundance of heterotrophic diazotrophs in the northern South China Sea: revealing the potential importance of a new alphaproteobacterium in N2 fixation." *Deep Sea Research Part I: Oceanographic Research Papers* 143 (2019): 104-114.

Response: Thanks for your comment. We have added the data from Chen et al. (2019) in our dataset and in fig. S2 (becoming Fig. S3 in revised manuscript). Also, we update this figure by including the zero-abundance data, which made us to revise the sentence here:

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).

Also revised accordingly in Section 3.4.6: "The decrease in Gamma A abundance with depth (Fig. S3 a, c and d; Moisander et al., 2008; Langlois et al., 2015; Chen et al., 2019b; Shiozakiet al., 2014; Wu et al., 2019) may therefore be attributed to" … "The nearly constant Gamma A abundance with depth in the Tropical Atlantic Ocean (Figs. S2b) can be the results of"…



340

Figure S3. Vertical profiles of Gamma A abundance in (a) Southwest Pacific Ocean, (b) tropical Atlantic Ocean and (c) India Ocean, (d) South China Sea. Dashed lines show all the sampled profiles, and solid lines and error bars are the mean and standard error in depth ranges of 0–10 m, 10–50 m, 50–100 m, 100–150 m and 150–250 m. Zero data was presented by 1 copy L^{-1} in this figure.

345

Figure 2 caption – it's not clear why some data was singled out as "highest" and shown with red dots, while other high datapoints were left out – much better description needed.

Response: The caption was revised:

350 "**Figure 2. The relationship between Gamma A abundance and net primary production.** Both Gamma A abundance and net primary production (NPP) are log10-transformed. The data with NPP of $10^{2.0}$–$10^{2.6}$ mg C m^{-2} d^{-1} (the "low" NPP range) are divided into 6 groups with equal log-NPP intervals (i.e., divided at NPP of $10^{2.1}$, $10^{2.2}$, $10^{2.3}$, $10^{2.4}$ and $10^{2.5}$ mg C m^{-2} d^{-1}), and the highest Gamma A abundance is identified in each group (red dots). The NPP-supported maximal Gamma A abundance
355 (red line) is estimated by linearly fitting the red dots in the low NPP range, and saturates at $10^{7.0}$ *nifH* copies L^{-1} for NPP > $10^{2.6}$ mg C m^{-2} d^{-1} (the "high" NPP range).

Section 3.2 and elsewhere – as above, I wonder whether carrying capacity is a needed term – at minimum it needs to be better defined, especially since the term has ecological ramifications that I am not sure are relevant here.

Response: The term "carrying capacity" was used in the previous manuscript to represent the maximal observed Gamma A abundance at given level of local net primary production. However, (as another referee also commented), the term "carrying capacity" has a strict ecological meaning. We decided to replace this term with "'NPP-supported maximal Gamma A abundance" in the revised manuscript.

Line 171 – "gamma A is expected to require a sufficient..."

Response: Corrected.

Line 208 – not clear why linear correlations are needed if the GAM is more reliable.

Response: The term "carrying capacity" was used in the previous manuscript to represent the maximal observed Gamma A abundance at given level of local net primary production. However, (as another referee also commented), the term "carrying capacity" has a strict ecological meaning. We decided to replace this term with "'NPP-supported maximal Gamma A abundance" in the revised manuscript.

Line 219 – "is presumed to be" in place of "was supposably"

Response: Corrected.

Line 235 – too speculative

Response: Thanks for your comment. We accepted that this conclusion is too speculative. We have deleted this sentence in our manuscript.

Section 3.4.5 – although this relationship is interesting, this discussion is speculative, thus needs to be better phrased – e.g. interpreting this as "indirect" evidence supporting the hypothesis that some NCDs are motile is misleading.

Response: Thanks for your comment. Our main hypothesis is Gamma A may benefit from the association with diatom. Swimming motility gene was suggested as a potential mechanism to find favorable niche and probably an indication of particle-attached lifestyle in Delmont et al. (2018). We have deleted this misleading message and revised this paragraph as:

"Our GAM results also suggested a positive relationship between silicate and $\Delta_{Gamma\text{-}A}$ in both the low- and the high-NPP groups (Figs. 4f and 4n), indicating 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. 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). The 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 diatom."

Line 275 – Abundance does not equal active N2 fixation. No evidence that gamma A fixes anywhere, including the mesopelagic. Needs rewording.

Response: Thanks for your comment. Line 275 did not have the relevant message, and we guessed you were talking about the last sentence of this paragraph (line 278): …"can be the results of active transport of organic matter from the surface that fuels heterotrophic N_2 fixation in the dark deeper ocean."

We then changed "fuels heterotrophic N_2 fixation" to "supports the growth of Gamma A".

Line 292 – was this described in the methods? More details would be helpful.

Response: Thanks for your comment. This was not described in method section because we used this relationship based on the observations that the maximal observed Gamma A abundance increased with NPP (i.e. red line in Fig. 2). Considering the coherence and clarity of the paper, we think it will be better to describe this here rather than in the methods part.

We defined $\Delta_{Gamma\text{-}A}$ as the observed Gamma A abundance minus corresponding NPP-supported maximal observed Gamma A in logarithmic space, which practically removed the impact of NPP in $\Delta_{Gamma\text{-}A}$. Then, we analyzed other controlling factors on $\Delta_{Gamma\text{-}A}$ (mentioned in section 3.3) by using GAM. Therefore, predicted Gamma A abundance can be received by predicted $\Delta_{Gamma\text{-}A}$ plus the modeled NPP-supported maximal Gamma A abundance.

We have revised this sentence as:

"As described above, $\Delta_{Gamma\text{-}A}$ was defined as the Gamma A abundance minus corresponding NPP-supported maximal Gamma A abundance. After $\Delta_{Gamma\text{-}A}$ was predicted by controlling factors other than NPP using GAM (Figs. 5a-b), it was added back to the NPP-supported maximal Gamma A abundance (i.e., the red line in Fig. 2) to form a prediction model for the Gamma A abundance (Fig. 5c).

Line 300 – I would begin this discussion with an emphasis that the model predicts high abundances where gamma A is not observed, like the Southern Ocean and coastal areas.

Response: Thanks for your comment. We would rephrase this sentence as: "The results suggested that the Gamma A was most abundant in the Southern Ocean and the upwelling region in the Eastern Tropical South Pacific (Fig. 6A) where, however, Gamma A was not sampled (Fig. 1)." (Note that by reassessing we have decided to remove "coastal areas" from the sentence.)

Line 305 – remove "where"

Response: Corrected.

Line 354 - I'm confused why the PCR bias is mentioned here - there is no end-point PCR data included in this study. I think a more relevant discussion could include unknown copy #s in gamma A's genome, or even accuracy of qPCR in general, due to the reliance on standard curves.

Response: Thanks for your comment. We have deleted PCR bias here, and revised this paragraph substantially as listed above (in our response to general comments).

Line 359 - N2 fixers have been shown to be very patchy in space and time, see Robidart et al., 2014.

Response: Thanks for your comment. We agree that N_2 fixers can be very patchy (and therefore many zeros are true zeros). We have added this argument in our manuscript (please see our response to general comments).

Line 367 – remove "confirming its heterotrophy" – over interpretation.

Response: Corrected.

Line 368 – replace "include" with "suggest" or the like

Response: Corrected.

Line 379 – there are many "universal" nifH primers with varying performance – do you mean a universal qPCR assay (which is unrealistic and would be difficult to interpret data from)?

Responses: Thanks for your comment. We agree that the statement is incorrect. What we wanted to express was that more NCD phylotypes were needed to be quantified, as Gamma A can only represent part of gammaproteobacterial diazotrophs. Therefore, we have revised this sentence as:
"Lastly, future study should also consider qPCR primer and probe sets targeting other NCDs such as Alphaproteobacteria and Cluster III phylotype, 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)."