Response to the comments from Anonymous Referee #2

Thanks to Shao and Luo for thoughtful responses to both reviewers' comments. Two of my original comments were not addressed fully, so I raise them here again.

Response: Thank you again for your helpful comment. Please kindly see our responses to these two comments below.

First and most importantly, in the Figure 2 caption (with implications elsewhere). Although more description is supplied, I am still unclear why the high data points at 10^2.0 were left out of the linear regression. Please provide justification.

Response: We apologize for this confusion. We now have included the data point at NPP of 10^2 in the linear regression. In addition, partly suggested by another reviewer, we have shortened the intervals of $Log_{10}NPP$ to 0.05 when identifying the maximal observed Gamma A abundance, thus having more data points for a more reliable linear regression between the highest observed Gamma A abundance and NPP (see Fig. 2).

Secondly, it still seems that there is data in the global plots (now only fig. 2) that is not cited in Table 1 from the Eastern South Pacific (at a minimum). Please triple check that all data is properly cited.

Response: Although all the Gamma A abundance data shown in Fig. 1 in the previous version had been cited in Table 1 (the data in the the Eastern South Pacific were from Shiozaki et al. (2018a)), as suggested by the reviewer, we rechecked all of our data sources and found two additional papers that also reported undetected (i.e. zero-value) Gamma A abundance in the Eastern South Pacific (Halm et al., 2012; Turk-Kubo et al., 2014). (It was missed in the previous version probably because Gamma A were reported using an uncommon acronym "AO") We have added these new zero-value data in Fig. 1 and cited the papers in Table 1.

Halm, H., Lam, P., Ferdelman, T. G., Lavik, G., Dittmar, T., LaRoche, J., . . . Kuypers, M. M. M. (2012). Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. The ISME Journal, 6(6), 1238-1249. doi:https://doi.org/10.1038/ismej.2011.182

Shiozaki, T., Bombar, D., Riemann, L., Sato, M., Hashihama, F., Kodama, T., . . . Furuya, K. (2018a). Linkage between dinitrogen fixation and primary production in the oligotrophic South Pacific Ocean. Global Biogeochemical Cycles, 32(7), 1028-1044. doi:https://doi.org/10.1029/2017gb005869

1

40 Turk-Kubo, K. A., Karamchandani, M., Capone, D. G., & Zehr, J. P. (2014). 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(10), 3095-3114. doi:https://doi.org/10.1111/1462-2920.12346

45

Response to the comments from Anonymous Referee #3

This is an interesting paper displaying the putative controlling factors of Gamma A, the most sampled non-cyanobacterial diazotroph (NCD) in the ocean. However, I noticed outstanding over-
50 speculations throughout the manuscript, in which some of the discussions were even baseless. Although the previous reviewers had also pointed out the same problem, it seems that the authors insisted on their speculations.

55 Response: We thank the reviewer for the very constructive comments that have greatly improved our manuscript. After carefully reading the reviewer's general and specific comments, we now have modified our analysis and revised the manuscript to fully avoid those speculations (in particular, the speculation that Gamma A is heterotrophy and uses organic matter from primary producers). In the previous version, we assumed that Gamma A was supported by NPP, and therefore the dependent variable of the GAM
60 analysis was the reduction (residual) of observed Gamma A abundance from the "NPP-supported maximal abundance". In the present version, we have discarded this method. In the GAM analysis, the variable to be predicted is now the Gamma A abundance itself, and NPP is added to the predictors with other environmental variables. The GAM analysis now applies to the entire dataset, not to two separated groups according to NPP as was done in the previous version. This new GAM analysis revealed some
65 different features from the previous analysis. We believe that the new GAM analysis is more objective and that the results are more robust. For example, after directly adding NPP into GAM as one of the predictors, the GAM reached a higher explanatory power while not identifying a substantial relationship between Gamma A abundance and temperature or DOC. We would thank the reviewer again for her/his very useful comments that make the analysis more solid.

70

Also, I have some concerns regarding their approaches. I doubt if the authors should use monthly climatological factors as the predictors of Gamma A, given that the ocean is highly dynamic and the diazotrophs are usually patchily distributed. More details please see the specific comment.

75

Response: We completely agree with the reviewer that the ocean is highly dynamic and that the real environment can be different from climatological conditions. That was also one of the reasons why we

also analyzed the relationship between mesoscale eddies and Gamma A abundance as an example to show that the dynamic ocean can also influence Gamma A. This is discussed in the first paragraph of 3.4: "The root-mean-square error (RMSE) of 0.84 and an R^2 of 43% in the prediction model (Fig. 4c) 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 ..."

Many climatological data of biogeochemical properties are available only in monthly intervals, particularly the nutrients. Therefore, we cannot conduct our analysis using data with shorter temporal resolutions. Nevertheless, the monthly climatological factors can also provide a large-scale background for understanding the general habitats of organisms. The practice of using monthly climatological data in meta-analysis appears common in marine ecology, including those for diazotrophs (e.g., Tang and Cassar, 2019). In a previous analysis of ocean N_2 fixation rates, our group even used yearly average climatology data as preditors (but with less data points) (Luo et al. 2014).

Tang, W., and N. Cassar (2019), Data-driven modeling of the distribution of diazotrophs in the global ocean, *Geophys. Res. Lett.*, 46, 12258-12269, doi:10.1029/2019gl084376.

Luo, Y.-W., I. D. Lima, D. M. Karl, C. A. Deutsch, and S. C. Doney (2014), Data-based assessment of environmental controls on global marine nitrogen fixation, *Biogeosciences,* 11(3), 691-708, doi:10.5194/bg-11-691-2014.

Regarding the patchness of diazotrophs, it was also raised by a reviewer in the first round and we have fully addressed this issue (fourth paragraph in Section 2.1), largely by only analyzing non-zero Gamma A abundance data. The reviewer who asked this question appeared to agree with our modification.

L8: Why did the authors presume that the NCDs are heterotrophs? In particular, the physiology of Gamma A is basically unknown. I suggest deleting "are presumably heterotrophic bacteria".

Response: Thank the reviewer for this comment. We stated that NCDs "are presumably heterotrophic bacteria" from the suggestions of other reviewers and also based on statements from other papers (Bombar et al., 2016; Zehr and Capone, 2020). However, we now agree with the reviewer that NCD should not be assumed to be heterotrophy (see response to general comments) and have deleted "are presumably heterotrophic bacteria" from the abstract.

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.

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

120 L13-15: I am skeptical about the relationship between Gamma A abundances and the NPP estimated by remote sensing… Would it be caused by the relationships between GammaA and other factors like temperature or chlorophyll? Also, "NPP-supported maximal abundance" is highly speculative, which sounds like that you have already proven the direct relationship between Gamma A and NPP. It should be noted that correlation may not imply causation. Also, let's say the

125 positive relationship between Gamma A and NPP is true, you cannot tell if Gamma A contributed to NPP or were supported by NPP…

Response: We understand that NPP from satellite remote sensing is derived from temperature and chlorophyll using certain algorithms, which however is beyond the scope of this paper. As a reference, in

130 our modified multivariate GAM analysis, NPP and temperature, together with other environmental parameters, are added into potential predictors and the results suggest that Gamma A abundance has a generally positive relationship with NPP, but no clear pattern between Gamma A and temperature is generated particularly for temperature > 15 degreeC in which most Gamma A data were reported (Fig. 4g). In other words, Gamma A may be more related to NPP than temperature.

135

Regarding the "NPP-supported maximal abundance", we completely agree with the reviewer (see our more response of this issue to the general comments) and the term "NPP-supported maximal abundance" has been removed from the abstract and the entire paper.

140 L16-17: These interpretations are highly speculative, which should be avoid in the abstract.

Response: Thank you for this comment. We have deleted these interpretations.

L19: I think your result only tells that GammaA abundances were higher in the waters with higher
145 SLA, while it is uncertain whether these higher SLA values really mean eddies…

Response: We identified the cores of mesoscale eddies by the outermost closed contour lines of the SLA field which has been mentioned in the Methods section (below Eq. 2). We also slightly revised the text to make the statement clearer: "The cores of mesoscale eddies were identified by the outermost closed
150 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.".

L20: How were Gamma A affected by the organic matters? It needs to be more specific.

155 Response: As we no longer assume Gamma A is heterotrophy, this sentence has been removed.

L21: It is not true. Gamma A are positively correlated with temperature, just like other cyanobacterial diazotrophs.

160 Response: We tried to state that some (not all) predictors for Gamma A were different from those for autotrophic diazotrophs. We have revised the abstract with our new results and made the statement more

4

precise: " 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. "

L55: Gamma A were not detected in aphotic waters.

Response: Thank you for the comment. This sentence has been deleted in the revised manuscript.

L63: Yes, some NCDs could be autotrophic/mixotrophic, and that's why it is inappropriate to presume that NCDs are heterotrophs. Also, as Gamma A are generally more abundant in surface water, they could be photoheterotrophic or even phototrophic.

Response: We thank the reviewer for mentioning this. As stated in the response to general comments, we have avoided the speculation that Gamma A is heterotrophy. Additionally, our new results revealed that Gamma A tended to habitat well-lit waters, which was discussed in the revised manuscript.

L66: Please specific what "active" means here. Were the NCDs active in fixing nitrogen in these habitats?

Response: We changed to "substantial presence of NCDs are found in DIN-replete environment …"

L126-132: We don't know the real suitable conditions for Gamma A yet… This problem could be simply because the currently available/commonly used environmental data does not cover the real controlling factors of Gamma A.

Response: Thanks for this comment. We agreed and revised the text to incorporate this comment: "It can also indicate our limited understandings of environmental conditions: The currently available environmental data do not include all the controlling factors of Gamma A."

L154: The resolution of monthly averaged environmental data seems too low to predict the abundances of Gamma A.

Response: Please see our response to the general comments for the same question.

L230-L235: What will the correlation look like if you calculate the correlation coefficients using all the data points? Also, the amount of data points in "high NPP" and "low NPP regions" are largely different, which should be considered.

5

Response: This correlation stated here (0.21) was calculated using all the data points. The number of data points in "low NPP" is two times greater than that in "high NPP". As stated in our response to the general comments, the dataset was no longer separated into low-NPP and high-NPP groups in the GAM analysis in the revised manuscript.

L245: We don't know if Gamma A need organic matters from primary producers. It is simply your speculation from your observation, and you cannot use your speculation to explain your observation…

Response: We agree with the comment. We have revised this paragraph entirely to more objectively and logically evaluate the implication of the positive correlation between Gamma A and NPP:

" 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."

L255: The statement about the "NPP-supported maximal Gamma A abundance" is baseless. As we don't even know about the trophic status of Gamma A.

Response: It has been removed.

L286: GAM can delineate partial effects of different variables, including NPP. I am wondering why the authors artificially separate the dataset based on the NPP? Analysing the full dataset with GAM may result in more universal conclusions about the putative determinants of Gamma A. Also, I doubt that the linear regression based on 6 data points (in Fig. 2) means any valid relationship between Gamma A and NPP.

Response: Thank you for this comment. We have modified our analysis following this comment by including NPP in GAM and no longer separating the dataset based on NPP in GAM (see our response to the general comments).
In the revised manuscript, the linear regression line is an estimate just showing an approximate upper bound of the Gamma A abundance and is not used in future analysis (see our response to the general comments). Nevertheless, we shorten the intervals to pick more (12) data points for the linear regression.

245

L307: The discussion about DOC is high-speculative, which is fully depending on what the authors believe in… Based on the result of GAM, the correlations between DOC and Gamma A are contradictory with the authors' speculations about the NPP…

250   Response: DOC is no longer a significant predictor after we modified our analysis by including NPP as one of the GAM predictors and not separating high- and low-NPP groups (see our response to general comments). Therefore, the discussion for DOC has been removed entirely.

255

260