

# Response to the comments from Anonymous Referee #1

We thank Referee #1 for his/her very thorough and constructive comments that helped us greatly improve our manuscript. We have responded (in blue fonts) to the comments point by point and revised the manuscript accordingly.

5 Shao and Luo compile published abundance data of Gamma A (qPCR *nifH* gene counts), a  
putatively heterotrophic diazotroph widespread in the oceans. Using ancillary data, atlas, satellite  
products and models, they perform a thorough statistical analyses to infer relationships between  
Gamma A and environmental variables. Their results suggest that Gamma A benefits from primary  
production by-products and is mostly dominant in warm and iron-poor waters of the ocean. The  
10 data analyses are extensive and the results worth publishing.

However, the authors should improve the comparison between their study and Langlois et al. 2015,  
who also compiled Gamma A data and performed statistical analyses to define their niche. How  
does the present study build up on previous ones?

15 Response: Thanks for your comments. Although Langlois et al. (2015) has been referred in multiply  
places in the previous version of the manuscript, we agree with the reviewer that we should revise the  
manuscript so that our results can be better compared to Langlois et al. (2015).

20 In Langlois et al. (2015), the authors statistically analyzed the relationship between the Gamma A *nifH*  
abundance and a suit of environmental parameters including nutrients, salinity, temperature and oxygen.  
They found Gamma A was mostly distributed in the warm (tropical) and oligotrophic surface. With more  
data becoming available in the recent years, we used 80% more (1795 vs 992 data points) Gamma A *nifH*  
25 abundance data in our study than those in Langlois et al. (2015). We also used five additional variables  
including primary production, iron and DOC concentrations, solar radiation and mixed layer depth, as  
well as submesoscale eddies, to more thoroughly analyze potential controlling factors on Gamma A. Part  
of our conclusions is consistent to Langlois et al. (2015) that Gamma A prefers warm environment. But  
our study has revealed that Gamma A also prefer high primary production and cyclonic eddies, suggesting  
that sufficient supply of organic matter can be the more important determinant of Gamma A distribution.

30 We have revised the manuscript as follows:

(1) The last paragraph of 1. Introduction:  
“Langlois et al. (2015) analyzed the distribution of Gamma A phylotype in the Pacific and Atlantic Oceans  
35 and suggested that Gamma A preferred warm and oligotrophic surface oceans. 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, compiling a dataset with 1795 data points, 80% more 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

40 climatological monthly averages. To more thoroughly analyze potential controlling factors on Gamma A, we included 5 variables, including primary production, iron and DOC concentrations, solar radiation and mixed layer depth, in addition to temperature and concentrations of nitrate, phosphate and silicate that used in Langlois et al. (2015). We further explored the influence of mesoscale eddies on Gamma A abundance.”

45

(2) First paragraph of 4. Summary and outlook:

... in the global ocean. “The results of our study did not agree with the conclusion of a previous study that Gamma A preferred oligotrophic oceans (Langlois et al. 2015). Instead, most of our findings” ...

50 I also found several mis-citations, where the wrong citations are given to justify a statement or where the message of a given paper was not well understood.

In all, the exercise seems statistically correct but of relatively poor ecological interpretation significance unless several points are improved. Below I provide a list of comments.

55 Response: Thank you very much for your comments. We have responded to all specific comments point by point below.

Specific comments

60 L7: Delete “the” in “to the global marine”.

Response: Corrected.

L10: What is the carrying capacity? This term is used throughout the manuscript whereas it is never really explained.

65

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.

70

L15: “in addition” not “in additional”.

75 Response: Corrected.

L17: Eddies are not short-term features, they may last several months. Please choose another term.

Response: We have changed “short-term features” to “mesoscale features”.

80

L18: “organic matter” not “organic matters”

Response: Corrected.

85 L19: Weird wording, please rephrase.

Response: We have rephrased that “and therefore provide an insight into niche differentiation between the heterotrophic and autotrophic N<sub>2</sub> fixation.”

90 L20: “sampling” not “samplings”, and delete “better” from the end of the sentence.

Response: Corrected.

L32: “oxygen deficient zones”

95

Response: We have corrected “oxygen deplete zones” to “oxygen deficient zones”.

L33: “heavy” sounds weird, please rephrase.

100 Response: We have switched “heavy” to “significant”.

L33: I would not say heterotrophic N<sub>2</sub> fixation is “not well quantified”, it’s just not quantified at all. There is -currently- no assay able to isolate heterotrophic N<sub>2</sub> fixation from autotrophic N<sub>2</sub> fixation.

105 Response: Thanks for your comment. We have revised the sentence as “Although the N<sub>2</sub> fixed by NCDs has not been quantified, ...”.

L41: This is not what these papers really say. In Benavides 2018b Gamma A was not detected, so it does not necessarily mean it was not stimulated by DOM, it just was not present in the samples at all. In Benavides 2015 N<sub>2</sub> fixation in dark waters was stimulated by amino acids.

110

Response: Thank you for your comment. In Benavides et al. 2018b, abundant Gamma A *nifH* DNA copies were detected (Table 1 of 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).

115

The referee was correct that amino acids did stimulate N<sub>2</sub> fixation in dark waters in Benavides 2015, while the effect of sugar was not obvious.

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. In addition, the response of aphotic N<sub>2</sub> fixation to different DOM composition could also vary (Benavides et al., 2015).”

120

125 L42-43: NCDs are thought to be attached to particles, but they haven't been found to be attached to particles.

Response: Thanks for this comment. We have changed “attach to particles” to “associate with particles”.

130 L45: Bonnet 2016 find that N released by *Trichodesmium* is taken up by diatoms. Not N released by NCDs.

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”.

135

L45: “to equip” sounds weird.

Response: We have changed “to equip with” to “contain”.

140 L52-53: How did those studies look at DIN inhibition of individual NCDs strains? It seems this is not what these studies really did.

Response: Shown in FIG 6 of Bentzon-Tilia et al. (2015),  $N_2$  fixation in NCD strains *P. stutzeri* BAL361 and *R. ornithinolytica* BAL286 decreased significantly when  $NH_4^+$  was added. However, in another NCD strain *R. palustris* BAL398,  $N_2$  fixation increased dramatically upon the addition of  $NH_4^+$ , which they speculated was because the nitrogenase complex had a function in addition to N acquisition, such as using  $N_2$  fixation as an electron sink for  $NH_4^+$  consumption.

145

The inhibition of  $NH_4^+$  to NCD strain *S. castanea* was also observed in Martínez-Pérez et al. (2018).

150 L58: This is quite unfair to say, please cite:

Bombar, Deniz, Ryan W. Paerl, and Lasse Riemann. 2016. “Marine Non-Cyanobacterial Diazotrophs: Moving beyond Molecular Detection.” *Trends in Microbiology* 24 (11): 916–27.

Cornejo-Castillo, Francisco M., and Jonathan P. Zehr. 2020. “Intriguing Size Distribution of the Uncultured and Globally Widespread Marine Non-Cyanobacterial Diazotroph Gamma-A.” *The ISME Journal*. <https://doi.org/10.1038/s41396-020-00765-1>.

155

Langlois, Rebecca, Tobias Großkopf, Matthew Mills, Shigenobu Takeda, and Julie LaRoche. 2015. “Widespread Distribution and Expression of Gamma A (UMB), an Uncultured, Diazotrophic,  $\gamma$ -Proteobacterial NifH Phylotype.” *PloS One* 10 (6): 1–17.

160 Moisander, Pia H., Mar Benavides, Sophie Bonnet, Ilana Berman-Frank, Angelicque E. White, and Lasse Riemann. 2017. “Chasing after Non-Cyanobacterial Nitrogen Fixation in Marine Pelagic Environments.” *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2017.01736>.

Riemann, Lasse, Hanna Farnelid, and Grieg F. Steward. 2010. “Nitrogenase Genes in Non-Cyanobacterial Plankton: Prevalence, Diversity and Regulation in Marine Waters.” *Aquatic Microbial Ecology: International Journal* 61 (3): 235–47.

165 Response: Thanks for your suggestion. In Bombar et al. (2016), the discussed controlling factors were presence of oxygen, presence of reactive inorganic nitrogen and the availability of energy.. Iron was mentioned as an important component of nitrogenase but not its effect on NCDs. Langlois et al. (2015) mentioned Gamma A may rely on DOC accumulated in the upper water column due to vertical stratification. In Cornejo-Castillo et al. (2020), the controlling factor discussed is the abundance of  
170 Gamma A in different size fractions. Moisander et al. (2017) reviewed studies related to NCDs and the major controlling factor discussed is DIN and DOM. Riemann et al., 2010 also discussed the impact of carbon, nitrogen and oxygen on NCDs. These studies did not directly analyzed the relationship between NCDs and iron/stratification.

Therefore, we have revised the texts as:

175 “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<sub>2</sub> (Bombar et al. 2016), although, as discussed above, the N<sub>2</sub> 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, there have been, to our knowledge, no studies analyzing the effects of Fe or stratification on NCDs.

180

L59: change “can” to “may”.

Response: Corrected.

185

L63: You may cite Benavides, M., and J. Robidart. 2020. “Bridging the Spatiotemporal Gap in Diazotroph Activity and Diversity With High-Resolution Measurements.” *Frontiers in Marine Science* 7. <https://doi.org/10.3389/fmars.2020.568876>.

190 Response: We have added this citation.

L69: “suggesting” would be fairer than “revealing” here. Note that nif genes can be used for other purposes.

195 Response: Thanks for your suggestion. We have made this correction.

L84: “the upper 100 m of the water column”.

Response: Corrected.

200

L87: there is no common qPCR detection limit, it depends on the assay, the machine, the lab, the volume of water filtered.

Response:

205

The referee was correct that there was no common qPCR detection limit. Usually, the detection limit ranges from  $10^1$  to  $10^2$  copies  $L^{-1}$ . The number of presence data points (72) under detection can be considered overestimated when the larger detection limit ( $10^2$  copies  $L^{-1}$ ) was taken into calculation if the data can be assumed normally distributed and all the zero-value data present low abundance below the detection limit.

210

However, we have removed the sentence in the revised manuscript upon the comments from Referee #2. We now agree with Referee #2 that some reported zero-value data were true zeros and the distribution of Gamma A can be patchy. We have added a new subsection in Results to discuss the zero-value data.

215

Table 1: I am a bit puzzled at 0 m depths, this is unlikely. Please check.

Response: Thanks for your reminding. We rechecked the original data reported in published papers, depths of 0 m were included. We supposed this may represent that the data were sampled at sea surface.

220

L103-104: I wonder why an artificial neural network was considered for DOC concentrations when there is a now a global database available <https://odv.awi.de/data/ocean/dom-compilation-hansell-et-al-2021/> Please reconsider using it instead.

225

Response: Thanks for your suggestion. Most our Gamma A samplings do not have DOC data available in the Hansell’s global database sampled in the same spatial and month grids. Indeed, the DOC data we used is produced from an artificial neural network model based on the same DOC observation database mentioned by the referee.

230

L116: It is unclear how SLA data was used, data was extracted from the same days as Gamma A samples were taken? Please explain.

Response: Thanks for your comment. Yes, SLA data was extracted from the same days as Gamma A samples were taken. 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:

235 “To identify if the Gamma A abundance was sampled in cyclonic or anticyclonic eddies, we extracted  
from AVISO program ([www.aviso.altimetry.fr](http://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  
240 anticyclonic (positive SLA) eddies cores were recorded. Otherwise, data points were recorded as ‘outside  
the eddy’.”

L149: *nifH* abundance also decreases with depth in the North Pacific (see work from Church at station ALOHA).

245 Response: Thanks for your comment. Yes, Church et al’s work found *nifH* abundance decreases with depth in the North Pacific. However, their study mainly focused on cyanobacterial diazotrophs and did not report *nifH* qPCR copies of Gamma A.

250 L170: please explain what the carrying capacity is.

Response: As mentioned above, we have changed “carrying capacity” to ‘NPP-supported maximal Gamma A abundance’ in our manuscript.

255 L172: Please cite Bombar 2016.

Response: We have added this citation.

260 L175: How are biogeographic patterns biased by the sparse and uneven sampling in different ocean regions? Can this be assessed statistically?

Response: Thanks for your comments. Yes, spatial biases in samples existed in our data set. To partly eliminate this bias caused by concentrated samplings in specific regions, in the previous version we have binned our data points into  $2^\circ \times 2^\circ$  grids. In addition, the standard errors estimated by GAM can also help  
265 to assess this kind of bias. Regions with undersampled biogeographic features would contain large uncertainties shown in Figure 6b. From our result, the largest uncertainties for the predictions exist in the Southern Ocean (Fig. 6b) because there were no Gamma A samples in this high-nitrate area. Other than this region, the uncertainty in the predicted Gamma A *nifH* abundance was at similar level in the global ocean (Fig. 6b), partly indicating that the spatial biases in samples may not impact our analyses greatly.  
270 We have added a discussion of this issue at the end of the revised manuscript: “The samples of Gamma A *nifH* copies are still limited particularly in the Southern Hemisphere, possibly causing spatial biases in our analyses. More sampling and studies are needed in the future to improve our understandings” ...

275 L190-191: This negative correlation is because low temperature anticorrelates with NPP, right?

Response: The response variable  $\Delta_{\text{Gamma-A}}$  used in the analyses here was the “residual” of observed Gamma A abundance to the NPP-supported maximal Gamma A abundance, which therefore practically removed the effect of NPP.

280

Nevertheless, we agree with both referees’ comments on the necessity of the univariate linear correlation analysis, and have decided to delete this section in the revised manuscript.

285 Section 3.4. The first sentence belongs in the methods. Why even show linear regressions at all if the model is deemed better? I would suggest just mentioning the correlations, maybe move them to the supplementary, and dive into the GAM directly in the main text. Why are, in any case, the effects found using GAM so different to the ones obtained with linear correlations? (e.g. L219).

290 Response: Thanks for your comments. Univariate analysis was used in linear correlation while multivariate analysis was used in GAM. Therefore, effect of every controlling factor in GAM is partial effect when other controlling factors were controlled. That is why effects found using GAM were different to those in linear correlation. Again, upon both referee’s comments, we have decided to delete the linear correlations section (and associated discussions in other places) in our manuscript.

295 L223: fuel with what?

Response: Particulate organic matter (POM) can fuel Gamma A with organics. We have revised the sentence that “Lastly, particulate organic matter (POM) can also supply necessary organic carbon and nutrients to Gamma A ...”

300

L224: I suggest replacing Farnelid 2019 for Riemann 2010.

Response: Thanks for your suggestion, we have made this replacement.

305 L235: This seems quite a speculative conclusion to make. DOC concentrations alone do not inform about lability, and, to date, we don’t know anything about the metabolism of Gamma A or which kind of DOM molecules they may use.

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

L243-244: note that NCDs also need P.

Response: Thanks for your comment. P\* in our study represents excess inorganic phosphate in



315 seawaters. NCD's source of P remains unknown. We have revised our manuscript to: "while our results tentatively indicate that the competition may not occur strongly between NCDs and phytoplankton, although it is still unclear whether NCDs use inorganic or organic P sources."

L301: this seems quite different from observations.

320 Response: Thanks for your comment. First, we agree that the statement of high Gamma A abundance in coastal regions can be misleading and should be removed. Second, Gamma A nifH copies were under sampled in the Southern Ocean and the upwelling region in the Eastern Tropical South Pacific (fig. 1), therefore we actually do not have direct supporting measurements. In the rest of the paragraph, we then discussed the largest uncertainties associated the high predicted Gamma A abundance in the Southern  
325 Ocean. The high Gamma A abundance predicted in the upwelling region in the Eastern Tropical South Pacific was mostly generated by its high NPP and temperature (Fig. S3 b and f).

We then have revised the manuscript as follows:

330 "The results suggested that the Gamma A was most abundant in the the Southern Ocean and the upwelling region in the Eastern Tropical South Pacific (ETSP) (Fig. 6A) where, however, Gamma A was not sampled (Fig. 1). The predicted high abundance in the Southern Ocean was mostly caused by its high nitrate concentration (Figs. S3g–h). However, the largest uncertainties for the predictions also exist in the Southern Ocean (Fig. 6b) as there were no Gamma A samples in this high-nitrate area (Fig. 1). The predicted high abundance in ETSP resulted from its simultaneously high NPP and temperature  
335 (Fig S3 a, b, e and f). Although the direct measurements of Gamma A nifH copies also lacked in ETSP, the sufficient samples in other regions with high NPP and temperature lowers the uncertainties of the predicted high abundance of Gamma A in ETSP (Fig. 6b). Nevertheless, future sampling in these two regions can then test our predictions and reduce the uncertainties."

340 Figure 6: why is the abundance "annual"? it's not a rate.

Response: Thanks for your comment. Annual mean abundance represents the mean value of Gamma A abundance from January to December. This term has been used often in other studies related to global distribution of species (e.g., Flombaum et al., 2013; Li, 1998). We then decided to keep this term

345

Reference:

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., ... & Martiny, A. C. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences*, 110(24), 9824-9829. (Fig. 2)

350

Li, W. K. (1998). Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnology and oceanography*, 43(7), 1746-1753.

Section 3.5. the connection or justification of why the effect of SLA is tested here is hard to follow.

355

Response: Thanks for your comment. For the environmental factors analyzed above, we used their climatological monthly values, which certainly may depart from in situ conditions (note that most of our Gamma A samples did not have sufficient in situ environmental parameters reported). Mesoscale eddies are one kind of phenomena that causes the in situ conditions different systematically from the climatological conditions. Also, as we mentioned in Introduction, mesoscale eddies can influence nitrogen fixation in the ocean. Therefore, we wanted to discover whether the Gamma A abundance in these eddies were systematically different from those predicted by our GAM model using climatological environmental conditions.

We have revised the first paragraph of Section 3.5 as  
“The root-mean-square error (RMSE) of 0.86 and an  $R^2$  of 41% in the prediction model (Fig. 5c) indicated that there was still substantial unexplained variance in Gamma A abundance. One possible reason was that we used the climatological monthly means in the environmental factors, while the in situ conditions can differ greatly from the climatological values. For example, oceanic mesoscale eddies can influence biogeochemical process 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.

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.

L316: eddies are not short term phenomena.

Response: Thanks for your comment. We have changed short term phenomenon to mesoscale phenomenon.

320: but also the number of data points in the NH is much higher than in the SH, potential bias, how can it be assessed?

Response: Thanks for your comment. Yes, the biases existed. However, this kind of biases is hard to be assessed. Similar to the biases caused by uneven sampling discussed above, we believed more samplings in the South Hemisphere are needed to reduce the bias (which has been emphasized in the last paragraph of the revised manuscript).

L367: “confirming heterotrophy” seems quite risky. We need genomic data and tracer experiments to confirm that.

Responses: Thanks for your comment. We have deleted “confirming its heterotrophy”.

L379: Unclear here, nifH primers are universal. There is no primer for cyanobacterial diazotrophs only. These primers target all diazotrophs with Mo nitrogenases.

400 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<sub>2</sub> fixation (Wu et al., 2019; Langlois et al., 2008; Martínez-Pérez et al., 2018; Chen et al., 2019b).”