

Biogeosciences Discuss., referee comment RC2
<https://doi.org/10.5194/bg-2021-349-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2021-349

Anonymous Referee #2

Referee comment on "Controlling factors on the global distribution of a representative marine non-cyanobacterial diazotroph phylotype (Gamma A)" by Zhibo Shao and Ya-Wei Luo, Biogeosciences Discuss., <https://doi.org/10.5194/bg-2021-349-RC2>, 2022

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?

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.

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

Abstract ----

Line 7 – First sentence is awkward. Perhaps "non-cyanobacterial diazotrophs (NCDs) may

be contributors to global marine....”

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?

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.

Line 17 – “mesoscale” in place of “short-term”

Line 18 – “matter” in place of “matters

Line 19 – “provide insight into” in place of “insight a”

Introduction ----

Line 25 – remove heterotrophic here

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

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

Line 30 – remove “dominant” or find a way to rephrase

Line 31 – “have also been” in place of “was also”

Line 33 – rephrase “heavy”

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.

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.

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

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

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

Line 46 – “equip” is awkward

Line 55 – is “supposably” needed here?

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

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

Line 85 – zero nifH copies can also be true zeros

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.

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.

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

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?

Results and Discussion -----

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

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.

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.

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.

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

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

Line 219 – “is presumed to be” in place of “was supposedly”

Line 235 – too speculative

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.

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

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

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.

Line 305 – remove “where”

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.

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

Robidart, Julie C., et al. "Ecogenomic sensor reveals controls on N₂-fixing microorganisms in the North Pacific Ocean." *The ISME journal* 8.6 (2014): 1175-1185.

Line 367 – remove “confirming its heterotrophy” – over interpretation.

Line 368 – replace “include” with “suggest” or the like

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)?