Authors response to RC1

We thank the editor and the two reviewers for their feedback on the manuscript. Here we provide our answer to RC1. We numbered all comments and provided answers below in blue. Lines are numbered according to the new manuscript with track change disabled.

Reviewer 1

General comments:

The authors examine global patterns in key nitrogen cycling pathways in the ocean using genetic datasets. The tool used -- the statistical model CEPHALOPOD -- has been previously published and provides a thorough and statistically rigorous analysis of how the key genes vary with environmental factors. This is a step up from simple DIY correlations between datasets. Most of the results themselves are not novel -- we know that assimilation of N by cyanobacteria should dominate in oligotrophic gyres, and nitrification at depth, for example, but it is still useful to see how these known patterns emerge using this tool. It is very interesting to see the widespread but low abundance coverage of the denitrification genes. However, the conclusion of high relative abundance of DNRA across the ocean, especially at high latitudes in the surface, needs more work to be convincing. Not only is it very surprising (why would an anaerobic process show up in the most oxygenated waters at high latitudes?), but the genetic marker may not be appropriate here. I am not a geneticist or omics expert, but my understanding is that nir B and nir D can be involved in ANRA as well as DNRA (see for example this review: https://doi.org/10.1016/j.synbio.2025.12.017), and ANRA is actually what we expect in high latitude surface waters. I suggest the authors look for nrfA instead and reevaluate their conclusions about DNRA.

Thank you for this thoughtful overview of our work and constructive suggestions that have substantially improved the manuscript. All the specific points raised in your general comment are addressed in detail in the responses below, including the reevaluation of DNRA by considering the new marker genes you suggested.

Specific scientific comments:

1. Use of "aerobic" (ex: L. 17): I would not use the term aerobic for autotrophic processes/photosynthesis. I realize the ANRA may be associated with aerobic heterotrophy as well in the surface, which may explain why a small section of the pie in Fig. 6c for ANRA is not cyanobacteria. I and many others consider aerobic to mean O2-consuming, not O2-producing, so may be confusing.

Thank you for this comment. We agree that the term “aerobic” may be misleading in this context because it is often interpreted as referring to oxygen-consuming metabolisms. Our intention was instead to describe metabolisms associated with oxic environments, including processes that do not consume oxygen, such as photosynthesis. To avoid confusion, we have replaced “aerobic metabolism” with “metabolism associated with oxic environments” where appropriate.

L20-22: Indeed, we identify that *Cyanobacteria* primarily encodes for metabolic pathways associated with oxic environment and biosynthetic pathways, while *Gammaproteobacteria* and *Nitrososphaeria* encode for nitrogen transformations related to energy requirements.

L439-443: In the epipelagic layer, nitrogen cycling metabolic pathways associated with cellular biosynthesis and oxic conditions (i.e., ANRA, Nitrogen fixation) are primarily supported by *Cyanobacteria* (relative contribution between 0.6 and 0.9), in a lesser extent by non-cyanobacterial groups such as *Gammaproteobacteria* (relative contribution between 0.1 and 0.4; **Fig. 6c**).

2. Genes for nitrification: Again, I am not an expert, but why not look for amoA and nxr genes for ammonia and nitrite oxidation? This is what is used for this database on nitrifiers: Tang et al "Database of nitrification and nitrifiers in the global ocean," https://doi.org/10.5194/essd-15-5039-2023 -- which should be cited and with which results here could be compared.

Thank you for these suggestions regarding nitrification, following Tang et al. (2023). In our study, however, the amoA gene (K28504) was not available in the metagenomic annotation dataset used for our analyses (Paoli et al. 2022). In addition, the nxrA gene (K00371) is associated not only with nitrification but also with denitrification and DNRA (according to the KEGG database), making it unsuitable as a pathway-specific marker under our selection criteria. We therefore excluded nxrA to avoid ambiguity in pathway attribution. Consequently, nitrification was represented using the genes pmoB (K10945), pmoC (K10946), and hao (K10535; **Fig. 2**). We also removed pmoA (K10944) from the list of nitrification markers, as it is primarily associated with methane

oxidation and was incorrectly assigned to nitrification in the previous version of the manuscript. We have now cited the nitrifier database of Tang et al. (2023) in the Methods and Discussion to place our results within the context of previous work on the global distribution of marine nitrifiers.

L200-204: For nitrification, although suggested by Tang et al. (2023), the amoA gene was not available in the metagenomic annotation dataset (Paoli et al. 2022), whereas nxrA was excluded because it is also associated with denitrification and DNRA (Kanehisa and Goto, 2000). Consequently, nitrification was represented by the pathway-specific marker genes pmoB (K10945), pmoC (K10946), and hao (K10535; **Fig. 2b**).

L488-493: The genomic potential, i.e., the relative abundance of those genes negatively correlates to PAR and positively correlates to nitrate and oxygen concentration in the epi- and mesopelagic layers (**Fig. 6b**), known drivers of nitrification rates (Tang et al., 2023; Ward, 2008; Zhang et al., 2020). Moreover, the estimated genomic potential distribution, showing elevated nitrification gene abundance in high-nutrient, low-chlorophyll (HNLC) regions such as the North Pacific and Atlantic, and Southern Ocean, aligns with higher measured nitrification rates (i.e., ammonia and nitrite oxidation; Tang et al., 2023).

L494-496:Therefore, the genomic potential of nitrification aligns with known nitrifiers distribution and nitrification rates (Pajares and Ramos, 2019; Tang et al., 2023).

3. Nitrification taxonomy: It is reassuring that the taxonomy associated with nitrification is Nitros*, but it is perhaps a bit circular to pitch it as a conclusion (considering the literature) since for nitrification, taxonomy and function are so related. It would make the paper stronger overall to pitch this more as a "validation" that the model can find known patterns such as this one.

Indeed, this result is not an original finding *per se*, but rather a validation that the taxonomy associated with our genomic potential matches current knowledge. We now state it as follow:

L486-488: Under oxic conditions, Nitrification encoding genes are predominantly associated to *Nitrososphaeria* archaea, which corresponds to established knowledge in the literature (Tang et al., 2023; Ward, 2008), thus validating taxonomic annotation of our selected marker genes.

4. Fig. 1: There are many wide black arrows, but then you say you did not consider them in this study. Perhaps there is another way to say this that is not so disserving? You are using methods that just identify correlations among datapoints, so probably OK to just not even mention that you cannot consider transport in the model. Same for anammox: you DID consider it, but it was just not sufficiently abundant to include in the statistical analysis.

Thank you for this comment. Our objective is to model the genomic potential associated with prokaryote genes and their environmental niches. Physical transport processes (Fig. 1, black arrows) are therefore outside the scope of the framework, as they are driven by ocean circulation and density gradients rather than by microbial genes. To clarify this point, we added the following statement to the Figure 1 caption:

L76-77: Note that physical transport processes are represented for completeness of the nitrogen cycle but were not modeled, as our study focuses on the environmental niches and the genomic potential of microbial communities.

Regarding anammox, we did evaluate this pathway but were unable to generate robust global projections because the available marker genes lacked pathway specificity and sufficient coverage in our dataset. To clarify this point, we added the following statement at the end of the Introduction, with a reference to the Methods section for further details:

L162-164: Although anammox was initially considered, it was not retained for global projections because the available marker genes lacked sufficient pathway specificity and coverage to support robust habitat modeling (*see Sect. 2.1.2.*).

We also clarified this point in the Methods section:

L204-208: Among the three enzymes associated with the Anammox pathway (**Fig. 1**, pathway 6) and detected in our metagenomic dataset (Paoli et al., 2022), the nitrite reductase NO-forming enzymes K00368 and K15864 did not meet our specificity criteria because they are also associated with denitrification, whereas hydrazine synthase

(K20934) was detected only once in the eastern pacific tropical upwelling system (-84.61°E, -1.89°N). Consequently, the Anammox pathway (**Fig. 1**, pathway 6) was not retained as candidate for global-scale modeling.

5. L. 120: "Habitat modeling" seems like it should represent a much broader category of modeling, rather than just simply this statistical approach. I would assume that any ecosystem model that aims to identify the ecological niche of a population or functional type is a habitat model.

Thank you for this comment. We agree that habitat modeling encompasses a broad range of applications and is not limited to a specific statistical methodology. This is already stated in the manuscript as follows:

L135–136: Rooted in ecological niche theory, it is widely used to predict the spatial distribution of organisms based on the environmental conditions at observed locations (Peterson and Soberón, 2012).

Our approach falls within this broader framework because it uses environmental conditions to characterize the niches of microbial taxa and the genomic potential associated with the relative abundance of nitrogen-cycle marker genes. To clarify this point, we have added the following statement at the end of the Introduction in the revised manuscript:

L156-157: To address these questions, we apply the core principles of habitat modelling to model the environmental niches of nitrogen-cycle markers genes in prokaryotes, and to project the global distribution of their genomic potential.

6. Cutoff (50 observations; line 169): You say that a gene must be present in at least 50 observations "to ensure sufficient spatial coverage" -- this does seem incompatible with many of the anaerobic N cycling transformations in anoxic zones unless you have 50 samples IN anoxic zones. This meant that you couldn't show anammox patterns. Perhaps for these types that didn't make that cut (anammox, and perhaps nrf as a replacement for nirBD for DNRA), it's useful to explain that they do "show up where they should," i.e., the small volumes of anoxic zones, rather than saying "we didn't consider them." You did consider them, but they are just not candidates for global-scale representation which is reasonable!

Thank you for this comment, which is closely related to your comment #4. We agree that the previous wording may have suggested that anammox was not considered, whereas it was in fact evaluated during marker-gene selection. We have revised the text to clarify that the pathway was excluded from global-scale modeling because the only anammox-specific marker gene lacked sufficient spatial coverage, despite being detected in a location where anammox activity is expected.

L204-208: Among the three enzymes associated with the Anammox pathway (**Fig. 1**, pathway 6) and detected in our metagenomic dataset (Paoli et al., 2022), the nitrite reductase NO-forming enzymes K00368 and K15864 did not meet our specificity criteria because they are also associated with denitrification, whereas hydrazine synthase (K20934) was detected only once in the eastern pacific tropical upwelling system (-84.61°E, -1.89°N). Consequently, the Anammox pathway (**Fig. 1**, pathway 6) was not retained as candidate for global-scale modeling.

7. L. 225: "outliers in the biological inputs were removed." Is this reasonable, given that we do think there is much heterogeneity in microbial communities and BGC fluxes in the environment? What are the implications of this? Is it more like you are finding "average" patterns and ignoring the known deviations? (This seems reasonable to do.)

Thank you for this comment. We agree that microbial communities and biogeochemical fluxes exhibit substantial heterogeneity in space and time. However, the definition of outliers depends on the scale of analysis. In our case, the CEPHALOPOD framework operates at a 1° x 1° monthly climatological scale, with the objective of identifying large-scale climatological relationships rather than resolving sub-mesoscale or short-term variability. We have clarified this point in the manuscript as follows:

L253-255: This step also aims to improve the robustness of model fitting to large-scale patterns and to reduce the influence of extreme values that may reflect local or transient conditions not representative of the climatological scale at which CEPHALOPOD is applied.

8. DNRA results: As I state above, I'm deeply skeptical that DNRA is actually this abundant in oxygenated surface waters. It would be really amazing if these results WERE true, and so, you need to convince me and the other readers that I can believe this. See above major comment about nrf vs. nirBD.

Thank you for this major comment. Although nirB and nirD are annotated as DNRA-associated genes in KEGG, this is contested in the literature. In this context, and in agreement with your suggestion and reviewer #2 (RC2), we revised our definition of the DNRA pathway by replacing the previously used nirB and nirD markers with nrfA and nrfH. These are more widely recognized in the literature as specific markers of DNRA. All analyses, figures, and results were updated accordingly. With this revised and more specific definition, DNRA-related marker genes are predominantly detected in oxygen-depleted regions, particularly in the eastern tropical Pacific upwelling system, consistent with the current literature describing this pathway. The relevant changes are detailed below:

L328-331: We observe a clear dominance of nitrogen transformation pathways associated with bioavailable over bio-unavailable forms of nitrogen – namely ANRA, and nitrification – in both epi- and mesopelagic layers (Fig. 3). In the epipelagic layer, ANRA accounts for the largest share of the genomic potential (0.79), followed by nitrification (0.17). However, DNRA that is also associated with bioavailable forms of nitrogen only accounts for a minor share of the genomic potential (<0.01).

L356-360: DNRA, typically associated with heterotrophic activity in suboxic to anoxic environments, shows peak genomic potential in the tropical Pacific upwelling system, with values ranging from 0.015 in the epipelagic to 0.03 in the mesopelagic layers (**Fig 4c**). Additional hotspots are predicted in ODZ of the northern Indian Ocean and in the Benguela upwelling system, with values ranging from 0.01 in the epipalgic to 0.02 in the mesopelagic layers (**Fig. 4c**). These are associated with a high uncertainty however (NSD > 0.5 ; see **Fig. S3** for detailed uncertainty pattern).

L364-373:

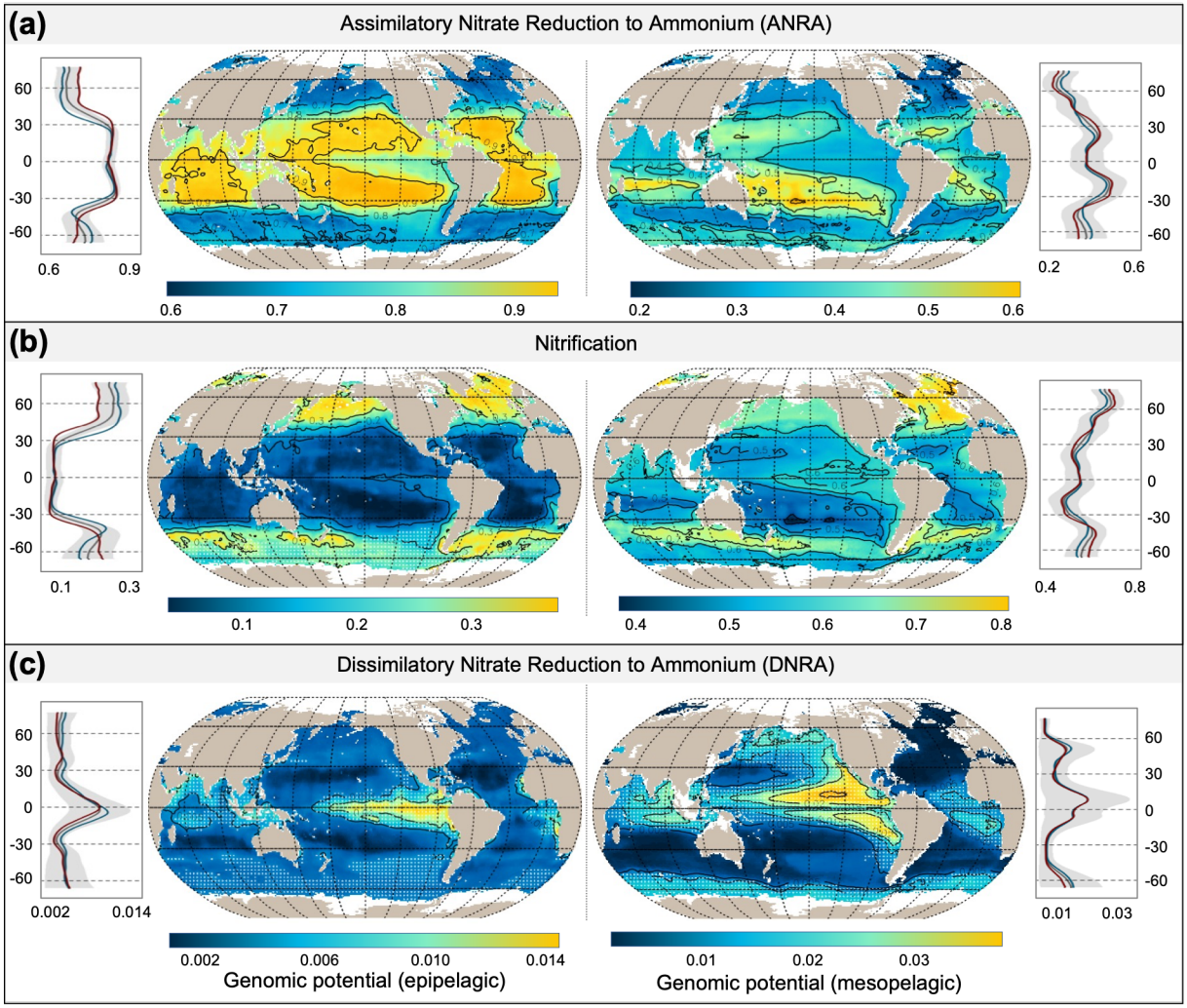

Figure 4: Genomic potential for bioavailable nitrogen transformation pathways represented by their relative abundance of metagenomic reads. Nitrogen transformation pathways are displayed in rows with: (a) Assimilatory Nitrate Reduction to Ammonium (ANRA), (b) Nitrification and (c) Dissimilatory Nitrate Reduction to

Ammonium (DNRA). Left panels correspond to the epipelagic projections and right panel correspond to the mesopelagic projections. Each panel represents the annual average spatial projection, with white stippling representing geographical areas associated with a high degree of uncertainty (i.e., standard deviation above 0.5 the global average). Latitudinal profiles are displayed in form of annual average (black), summer (red; northern hemisphere June to August) and winter (blue; northern hemisphere December to February). The projection uncertainty across bootstraps is presented as grey shading in the latitudinal profiles.

L446-448: Taxa that are obligatory or facultative anaerobes also show contributions to epipelagic DNRA (0.1 to 0.2 for each taxon) while mesopelagic DNRA is mainly associated with *Marinisomatia* in our dataset (0.8).

L597-502: The genomic potential for anaerobic nitrogen transformation pathways, such as denitrification and DNRA, positively correlates with low oxygen saturation and high apparent oxygen utilization (Fig. 6b). This is consistent with the dependence of Gammaproteobacteria on sub- or anoxic conditions to perform anaerobic respiration (Lam and Kuypers, 2011). The eastern Pacific upwelling, Arabian Sea, and Bay of Bengal ODZs show the highest genomic potential for DNRA and denitrification, reflecting known hotspots of anaerobic nitrogen transformation (Zhang et al., 2023).

9. L. 497: Levine et al 2025 is a review, not a model. Cite the models you are referring to directly.

Thank you for this comment. We clarified this by referring to the Régimbeau et al. (2022), Giordano et al. (2024) and Reynolds et al. (2024) studies directly. The statement is now as follows:

L526-530: Moreover, integrating metagenomics with genome-scale metabolic models (Giordano et al., 2024; Régimbeau et al., 2022; Reynolds et al., 2024), offers a promising avenue to link gene content with microbial metabolic fluxes (Levine et al., 2025). These metabolic models leverage genomic data to predict microbial nutrient transformations under varying environmental conditions, thereby enhancing the predictive capacity of biogeochemical cycles (Giordano et al., 2024; Régimbeau et al., 2022; Reynolds et al., 2024)

Smaller presentation comments:
10. L. 85: "ammonium that is being released" -- note that this may also be considered a key N transformation process! Perhaps also mention urea somewhere, as urea cycling is becoming more recognized currently.

Thank you for this comment. We agree that urea cycling is an important component of the marine nitrogen cycle and contributes to regenerated nitrogen, notably through the production of ammonium during urea degradation. This process may be included in the statement in L85 ("ammonium that is being released"). In the present study, however, we focus on redox-associated inorganic nitrogen transformation pathways that can be consistently resolved using available marker genes. In contrast, urea cycling belongs to the broader organic nitrogen cycle and is interconnected with multiple metabolic processes, including amino acid metabolism. For this reason, it was not explicitly included in the scope of our analysis. We clarified this statement and mention the urea cycle as follows:

L86-91: But nitrification thrives in the layers just below that, in the depth range between 100 and 500 m in the mesopelagic zone, where it oxidizes the ammonium released during the remineralization of the organic nitrogen that is sinking or transported downwards from the near-surface layers. These ammonium-releasing processes, that may include urea degradation among other pathways of organic nitrogen cycling (e.g., Wan et al., 2024), are not explicitly represented in this work, which focuses on redox-associated inorganic nitrogen transformations only.

11. L. 107: do you mean "such genes" instead of "such enzymes"?

We apologize for the confusion. While genes code for enzymes, we are referring to the genes. The statement has been clarified, also following reviewer #2 comments 4, 9 and 10 in our answer to RC2.

L115-118: The relative abundance of genes encoding a given metabolic function in prokaryotic metagenomes can be estimated from genome-wide metagenomic read coverage. Following Schickele et al. (2024), we refer to this relative abundance (e.g. of $nif_{H,D,K}$) as the genomic potential of a prokaryotic community for that function (e.g., nitrogen fixation).

12. L. 113: "the one[s] associated" (not "the one")

Thank you for pointing this out. We corrected the text accordingly.

13. L. 145: "Doing so, [we] explore"

Thank you for pointing this out. We added the missing word [we]

14. L. 272: "The three features" -- could you name them here? I am curious at this stage!

Thank you for this suggestion. We added the mention of the features with reference to the corresponding result section:

L305-307: The three features explaining most of the variance in the observation presented a cumulative feature importance (> 50 %) of 75 % for the epipelagic model (POC, nitrate and oxygen concentrations) and 81 % for the mesopelagic model (POC, apparent oxygen utilization and PAR; *see Sect. 3.5*.).

15. L. 302: Could you clarify/specify that "bio-unavailable forms" refers to N fixation and denitrification? This makes sense to look at these two separately from the others, but just took me a second to realize what was going on.

Thank you for your comment. We clarified this point as follows:

L337-340: This also means that the first order spatial gradient does not reflect whether nitrogen transformations are associated with cellular biosynthesis or energy metabolism, but rather whether they involve the bio-available nitrogen pool (ANRA, DNRA, and nitrification) or the exchange between the bio-unavailable nitrogen pool and atmospheric (N_2) nitrogen (nitrogen fixation and denitrification).

16. Figs. 4 and 5: Could you just label each of the maps with the process for much easier reference?

Thank you for this suggestion. We added labels referencing the process above each global projection for clarity.

17. L. 408: Again, I would not call cyanobacteria "aerobic."

Following our response to your comment #1, we do not refer to cyanobacteria as aerobic in the revised manuscript.

18. L. 410/11: Nitros* are named the way they are because they are associated with nitrification, so seems naive to state a result this way (see above comment).

Thank you for this comment. Following our response to your comment #3, we agree that this result is not an original finding *per se*, but rather a validation that the taxonomy associated with our genomic potential matches current knowledge. We state it as follow:

L486-488: Under oxic conditions, Nitrification encoding genes are predominantly associated to Nitrososphaeria archaea, which corresponds to established knowledge in the literature (Tang et al., 2023; Ward, 2008), thus validating taxonomic annotation of our selected marker genes.

19. L. 415: "In the lat[t]er"?

Thank you for pointing this out, however the corresponding statement has been removed in the present in the revised manuscript.

20. L. 416: Is it known that these are diazotrophs? Perhaps more fair to just say non-cyanobacterial groups?

Thank you for this comment. Non-cyanobacterial diazotrophs are reported in the literature, but we did not specifically test if the taxa reported in Fig. 6c are actively expressing diazotrophy. While the corresponding statement has been removed in the present in the revised manuscript, we applied your suggestion to a similar statement above:

L439-442: In the epipelagic layer, nitrogen cycling metabolic pathways associated with cellular biosynthesis and oxic conditions (i.e., ANRA, Nitrogen fixation) are primarily supported by *Cyanobacteria* (relative contribution between 0.6 and 0.9), in a lesser extent by non-cyanobacterial groups such as *Gammaproteobacteria* (relative contribution between 0.1 and 0.4; **Fig. 6c**).