“Change in diazotrophic community structure associated with Kuroshio succession in the northern South China Sea”

Zhang et al., 2023

Review

Zhang et al., 2023 show how the diazotrophic community in the northern South China Sea (nSCS) responds to intrusion by Kuroshio current waters by measuring *nifH* abundances via qPCR and high throughput sequencing, nitrogen fixation rates, and the degree of Kuroshio intrusion (KI) over two cruises in 2017 and 2018. The big picture question the authors address is an important one, namely how do physical changes in a marine ecosystem affect microbial diversity? The authors show that *Trichodesmium* is more abundant at stations with more KI and that UCYN-B is more abundant at stations with low KI. In addition, the authors show that UCYN-C and gamma-proteobacteria are more prevalent at stations moderately affected by KI. The authors also perform statistical tests to assess the degree environmental factors affect the diazotrophic community structure. Their tests suggests that environmental factors have a bigger affect on diazotrophic community structure in 2017, when KI is strong, than in 2018 when KI is weak.

I have several minor comments but no major revisions and recommend the paper to be accepted with minor revisions. All line numbers refer to the preprint visible on EGUsphere.

General comment

I found it difficult to understand what the Mantel tests vs. the NCM were testing. The abstract and Figure 8 say that the neutral community model is being used to test the effect of environmental variables on diazotrophic community composition.

However, in the methods and elsewhere, the Mantel tests are described as the way that relationship is assessed.

I think the easiest way to solve this would be to add an explanation in section 2.9 that describes how the NCM was used to assess the variation due to environmental and spatial factors.

Line by line comments

Title - I agree with the other reviewers that it should be Changes not Change

Line 35 - This phrasing is awkward. I would rephrase “The current state of marine N2 fixation study . . .” to be “The field of marine N2 fixation is changing dramatically.”

Lines 39 - 40 - I would also change the phrasing here. I would replace “Such previously established concept” with “This previously established concept . . . “

Lines 35 - 52 - Since UCYN-C is discussed in the abstract, I would explicitly mention it when the new diatom-diazotroph symbioses discovered by Schvarcz, 2022 are discussed. I would also mention another way that the field of marine N2 fixation is changing, which is the discovery of widespread, high coastal or continental shelf N2 fixation (Tang et al., 2019).

Line 94 - I agree with the previous reviewer comment that a 100 micron pre-filter could reduce *Trichodesmium* abundances. I would add in a comment mentioning this if you have not already.

Line 173 - Please include a reference for your merging criteria.

Line 177 - Please also include a reference for the RDP classifier algorithm.

Line 190 - It is not clear how you define spatial vs. environmental factors. Please state all spatial and environmental factors you test explicitly.

Line 196 - Please also include a citation to justify the choice of a WIF threshold of < 20.

Line 202 - I do not understand the two phrases used to describe what the NCM is assessing the relationship between. I would rephrase to more clearly explain the difference between “the occurrence frequency of diazotrophic taxa” and “their relative abundance across the wider metacommunity”

Lines 218 - 219 - This sentence should be in the methods not in the results.

Lines 226 - 230 - I didn’t see any tables where the relative abundances were clearly stated. I would make a new SI table with the same format as Table S5 that has the relative abundances for the 10 species investigated via qPCR. It is also hard to evaluate the authors’ claims about day vs. nighttime abundances since many of the times (19:00, 19:30, 20:00 are on the border of day and night). I would explicitly state with a D or N in Table S5 and the new table which times are day and night. Please also say if times are local time or another timezone.

Line 232 - This statement is misleading because there is quite a bit of variability across stations, for example stn 1 vs. 12 vs. 4 in 2017 (Table S6) . I would instead describe that there is considerable station to station variability but that across all stations cyanobacteria and gamma proteobacteria are at approximately equal abundances.

Line 233 - Here, I think there also needs to be an acknowledgement that at some stations *Trichodesmium* is much different than 58% of the cyanobacterial nifH abundances. For example at stn 1 in 2017 - Tricho is ~75%. Like above, I would say that there is considerable variability but that overall Tricho is 58% of cyanobacterial abundances.

Overall comment here - describing the overall pattern while acknowledging station to station variability will allow you to transition to the next section - because the station to station variability

is consistent with what your correlation analyses show. For example, Stn 1 which is mostly Kuroshio waters has predominantly Tricho.

Line 242 and/or Figure 5 - In Figure 5, I NFR not S NFR is in the same cluster as SST, SSS, DCM etc. Either there is a typo at line 242 or a typo in Figure 5.

Line 246 and/or Figure 5 - In Figure 5, SPP looks like it is in a separate cluster from I DIN, I DIP, and Nit. If you meant the previous fork in the tree, I would list all six variables that are not in the strong KI cluster.

Line 257 - Based on figure 8, the variation due to only environmental factors is 0.13 in 2017. The variation due to only environmental factors as well as the variation due to **both** spatial and environmental factors is 0.33

Lines 275 - 276 - I would rephrase this sentence, because the 2006 paper does not show what this sentence describes. I would rephrase to "Additionally the relatively lower abundance of UCYN-B in Kuroshio waters may be because Trichodesmium has greater genetic resources which allow it to outcompete UCYN-B for P."

Line 285 - I think this point should be nuanced more. It looks like UCYN-C was not detected at all at station 10 and in station 9 was detected at lower levels than in 2017.

Lines 288 - 293 - This paragraph is very confusing. It is difficult to figure out what argument the authors are trying to make.

Line 331 - I think predominated is a more accurate word than was distributed.

Line 335 - This difference between selective and neutral is not discussed before. It isn't clear to me what makes spatial factors neutral. Please add a section to the methods that defines this.

Line 338 - Another interpretation could be that the rare taxa are not constrained by environmental variables at all (i.e. rare taxa are always rare in this environment)- a really interesting result. I would further discuss this.

Line 339 - I would replace featuring with “correlated with” or “connected to”.

Line 362 - Please explain more - why especially UCYN-A?

Lines 379 - 380 - There is a positive correlation between NFR and KI at just one of the strong KI stations so I don’t think this claim is justified.

Lines 383 - 384 - There is already N2 fixation in temperate waters that is high - see Tang et al 2019 paper. I think the bigger picture importance of the paper is better said along the lines of the last sentence in the abstract.

Figure 2 - Please make the numbers on the contours and axes bigger. They are very very small and hard to read right now. Please also clarify, is the data in panel A the real data used to model the points in panels b-e? I would state this in the legend.

References

Tang et al., 2019
https://www.nature.com/articles/s41467-019-08640-0