The MS by Dutkiewicz et al. presents a detailed account on the drivers of the marine phytoplankton diversity in a numerical model.

The effort is of great interest since the model considers all the processes that are considered relevant when using a trait-based framework. It is also of interest since the impact the main traits and processes are discussed separately, using a set of well defined sensitivity experiments. Finally, it focusses the discussion on the immediate implications for the interpretation of real data. In particular it points out that the selection of the environmental variables that are used as explanatory variables in statistical analyses has to be coherent with hypotheses drawn from the current theories. Dramatically, if one uses a trait-based framework the outcome is that most of the variables are hard to be constrained quantitatively (eg, the nutrient fluxes).

Indeed, the actual phytoplankton richness is possibly orders of magnitudes higher than the one emerging from this model exercise. This limitation is possibly due to the strong limitations of the "classical" trait-based approach. In addition, in this exercise there are no significant conceptual novelties on specific processes. Nevertheless, as stated by the authors, a general synthesis of the lessons learnt using this framework is going to be very useful for future studies and, with given its pedagogic clarity, for students and young researchers.

I thus recommend it for publication after addressing some very minor points.

We thank the reviewer for the positive comments. We address these points below in blue text. We note that the diversity is indeed significantly lower than in the real world. This now stated this several times in the revised version. For instance, in the model description (lines 165, revised version 175):

"We also emphasis that the level of richness that the model captures, though large for a model, is orders of magnitude lower than the real ocean. Thus this is not a fully comprehensive study of diversity, but does never-the-less provide a promising avenue for understanding some of the controls on diversity."

In the model limitation section 6, lines 417 (new version lines 449-450):

"Our model only captures a tiny (probably orders of magnitude less) amount of the diversity found in the real ocean. Including more resolution along these axes and including additional trait axes would allow for further diversity, but is beyond the scope of this present study."

A check is required for all the citations (missing parenthesis or points).

In the revision we have checked for missing parentheses and citations. (And yes, we found several).

Specific comments:

Introduction. The last sentence is generally correct for the whole diversity but in most cases studies focus on single groups. Is it still true?

Yes, we believe this is still true even for a distinct group. For instance, if the group is diatoms, then our study suggests that transport will still be important for hotspots of diatom diversity, while size/species specific losses and resource supply will dictate size diversity within diatoms. We did obliquely refer to this in lines 427-429:

"Our results suggest that observed patterns of "total" diversity (or for any grouping of phytoplankton types, such as for nano and micro-eukaryotes along the AMT) are a result of multiple controllers: supply rate of limiting resource, imbalance in supply of different resources relative to competitor's demands, top-down control, particularly in terms of size-dependent grazing, and transport processes."

Introduction: A statement on the different definitions of "diversity" is missing. A general issue with the literature on plankton is the lack of discussion about the importance and the technical and ecological implications of the choice of the metrics for diversity.

We do discuss what we mean by "diversity" in the context of the paper (lines 159-162). We also have a discussion about some of the techniques of measuring diversity in the Discussion (see lines 475-485). However, we agree with the reviewer that this also warrants a statement in the discussion (and also agree that there is a lack ok of such discussion in many studies on "diversity"). As such we have added the following at line 68:

"In this study we will almost exclusively consider diversity in terms of "richness", the number of locally co-existing species. This definition is often referred to as alpha-diversity. We focus on richness here as the ecological theories we use explain co-existence, rather than other common metrics of diversity such as Shannon Index or evenness. Given the model setup, we also do not consider the rare biosphere."

The study by Lima-Mendez is not on diversity but on interactions. Their conclusions that biotic interactions are more important than environmental factors in setting the community network derive from the analysis of a dataset that contains much more "species" than this model. I think it is just not possible to compare the two approaches with the current state of understanding. In addition, the model has no interactions except for grazing. Thus citing it is useful especially to discuss how these results represent a challenge for the current modelling approaches.

We agree that the Lima-Mendez paper is not about diversity, and as such we remove it from the Introduction. However, this is a valuable paper and we therefore do still cite it in the Discussion, but now make clearer that that paper was about community structuring rather than diversity. We did not mean to sound as though we were comparing the two approaches, and have changed the wording in the Discussion so that is no longer misleading (line 455)

"In a study focusing on the interactions (and hence community structure) showed little statistical links to nutrient concentrations (e.g. Lima-Mendez et al., 2015)."

L45 "there is evidence suggest"?

Now changed to "suggesting"

Section 2.

What is the definition of richness used for the AMT data? What is the reason for not using a rarefaction of the data prior to define richness? The issue should be discussed shortly, also considering the method used here (L94-95).

The definition of richness in the AMT is now included in this section (see text quoted below). Since the model has to impose a threshold of abundance (or biomass) for defining presence/absence of population types, and thus for defining richness for the AMT we consider that the cleanest comparison is with raw species richness data rather than by using rarefaction (i.e. such that we do not encompass the rare species). We also make this assumption clearer in the text.

Near line 87 (revised text line 89-91):

"Here diversity is determined as richness, which in this study is defined as the number of species detected in sample volumes in the range 10-100 ml."

And after line 100 (revised version 105-106):

"Given how these data are compared to model output (see below) we purposely neglect the rare biosphere, so do not attempt any techniques such as rarefraction to account for the rare species."

And altered text around 159-165 (revised version 166-174):

"As mentioned in the introduction, in this study we primarily discuss diversity in term of "richness" defined here as the number phytoplankton types that co-exist at any location above a threshold. We, in particular, look at the annual mean of the instantaneous surface richness (though see Supplemental for examples with depth). Technically we use a threshold value (10^{-5} mmolC/m^3) to determine if a type is in existence at any spot. This value would convert to about 10 Prochlorococcus cells/ml (typical oligotrophic waters are above 10^3 cells/ml), or only a tiny fraction (10^{-4}) of a larger diatom cell/ml. Thus this definition neglects the rare species that would be difficult to separate from numerical noise. This is why we do not account for the rare species in the AMT observations discussed above."

The model resolution is very low for the current standards for the ocean physics. Presumably, the computational requirements to run the biogeochemical model are such that using a higher resolution was too demanding. Nevertheless, in discussing the limits of the study the lack of mesoscale and submesoscale processes should be mentioned.

We do mention the coarse resolution of the model (lines 185-187 and lines 387-388). But agree that this is not sufficiently discussed as a limitation. We now include additional text and feel that this significantly improves this article.

Near line 113 (revised version 119-121):

“At this horizontal resolution, the model does not capture mesoscale features such as eddies and sharp fronts, a limitation of the model that must be kept in mind when considering the results.”

And near line 354 (revised version lines 383-385):

“Both Clayton et al (2013) and Levy et al (2014) showed the importance of eddies in enhancing this process of transport mediated diversity. Thus the hotspots in the default experiment would likely be even higher in a model that did resolve the mesoscale.”

And also section 6 (Limitation of this study), after line 420 (revised version lines 455-460):

“Given computational constraints with this complexity of ecosystem model, we have use a coarse resolution physical model that does not capture explicit meso (or sub) scale features. Previous studies (e.g. Clayton et al 2013; Levy et al 2014) have shown the importance of such features in enhancing diversity. Mesoscale features are important in temporal increases in nutrient supplies (see e.g. Clayton et al., 2017), and from this study this suggests temporal increase in size classes during such events. Sub- and mesoscale mixing in frontal regions will also enhance the richness in hotspots (Clayton et al 2013), but also in a general increase richness (Levy et al 2014).”

L146 Missing the verb?

Not a verb, but rather a qualifies. Thanks for catching this. Now reads:

“Following empirical evidence, mixotrophic dinoflagellates are assumed to have lower maximum photosynthetic growth rates than other phytoplankton of the same size (Tang, 1995; Fig 4a) and lower maximum grazing rates than heterotrophic dinoflagellates of the same size (Jeong et al., 2010, Supplemental Fig S2).”

L149 micron?

Yes, now changed

Results. L191 “Though note. . . “ could be in parenthesis.

Agreed, this has been done in the revised version

L196 “given distributions”?

Thanks, this was a typo, Text is now altered to say: “compared to”

L199 “likely”?

Yes – large classes are definitely under-estimated. In revised text “likely” is removed.

L214 “enhanced”?

Changed to “enhanced”

L218 manuscript or article?

We’ve change to “study” to be consistent to the rest of the article.

L234 Please add the total diversity to the figure on sensitivity. The pattern looks similar to the thermal Norm one and thus it seems to suggest that processes that impact the Thermal Norm diversity (notably, transport here) can be very important in setting the total diversity.

We are a little confused here, as Figure 8 (which is discussed in the paragraph starting at line 234) does have the total diversity. Though at a quick glance the total and thermal norm richness looks similar, the total is indeed made up of all the dimensions. To avoid this confusion, we now add at lines 241 (revised text 260-263):

“At first glance total diversity (Fig 8a) may look most like the thermal norm diversity (Fig 8d), but this is mostly because our eyes are drawn to the hotspots. In reality total diversity patterns are strongly impacted by all three dimensions of diversity as will be shown more clearly by the sensitivity experiments discussed later.”

Perhaps the reviewer is suggesting adding the total diversity to Fig 11 (the sensitivity experiments)? We agree that this is a good idea. The revised version of the figure has the total (see below). This is a rather nice illustration of how the diversity decreases in all sensitivity experiment and that thermal norm diversity is not the same pattern as the total, so we have added additional text.

(revised version 317-318):

“However, the total diversity reduces dramatically (Fig 11, top row). Patterns of hot spots are however still apparent, but the increases in diversity with higher nutrient supply is no longer apparent.”

after line 303 (revised version lines 366):

“Total diversity is reduced everywhere, but mostly in the lower latitudes where the loss of diazotrophs and coccolithophores has a high impact.”

Line 354 (revised version line 382-385):

“Total diversity is reduced everywhere, but most dramatically in these hotspot regions.”

L387 Possibly? Several time in the text there are statements that are too strong. This is the case also for the comment on Lima-Mendez et al. The authors of this MS maybe right but they have no direct evidences to oppose. They can only suggest or hypothesize.

We have revised the text to emphasize where we can only hypothesize. We have removed the mention of Lima-Mendez in the introduction and have clarified our statement of this article in the discussion. For instance Line 446-456 (revised version 488-500):

"Though observational studies have hypothesized a multi-factorial control on diversity in the ocean (e.g. Rodriquez-Ramos et al 2015), they were unable to find significant correlations with any combination of factors such as latitude, temperature or biomass, or even nutrient concentrations. Correlating with factors such as temperature, latitude is a logical first step for trying to understand observed patterns of diversity, as these are often the only additional data that is available from a field study, and for instance "latitude" could potentially stand in for a range of biotic and abiotic processes. Our study, however, suggests that to some degree these factors are unlikely to help disentangle controllers of diversity. For instance, in our study it is mixing of different temperature water masses, potentially hinted at by local temperature variances rather than temperature itself, that is important. In a study focusing on the interactions (and hence community structure) showed little statistical links to nutrient concentrations (e.g. Lima-Mendez et al., 2015). On the other hand nutrient supply rates (a harder variable to measure) did show some measure of identifying communities (see e.g. Mouriño-Carballido et al. 2016)."

In other parts of the text we have added qualifiers or removed sentences that we, on hindsight, deem to be too strongly stated.

L392 and following: The limitation due to the low model resolution is never mentioned.

Yes, this was an oversight. As discussed above, we have now added several sentences on this issue in several parts of the articles, in particular in this section, after line 420 (revised version lines 455-460):

"Given computational constraints with this complexity of ecosystem model, we have use a coarse resolution physical model that does not capture explicit meso (or sub) scale features. Previous studies (e.g. Clayton et al 2013; Levy et al 2014) have shown the importance of such features in enhancing diversity. Mesoscale features are important in temporal increases in nutrient supplies (see e.g. Clayton et al., 2017), and from this study this suggests temporal increase in size classes during such events. Sub- and mesoscale mixing in frontal regions will also enhance the richness in hotspots (Clayton et al 2013), but also in a general increase richness (Levy et al 2014)."

More importantly, as only briefly discussed at the very end of the Discussion, the traitbased modeling approach, while being much improved here, is still far from reproducing the observed richness (especially if quantified using genetic or genomic approaches). There are issues with data, indeed. But it is unclear from this manuscript which should the future directions of research based upon this kind of modeling approach.

We have added the following sentences to address this issue, starting in the introduction, (lines 165, revised version 175):

"We also emphasis that the level of richness that the model captures, though large for a model, is orders of magnitude lower than the real ocean. Thus, this is not a fully comprehensive study of diversity or species richness, but does never-the-less provide a promising avenue for understanding some of the controls on diversity."

In the model limitation section 6, lines 417 (new version lines 449-450):

"Our model only captures a tiny (probably orders of magnitude less) amount of the diversity than is in the real ocean. Including more resolution along these axes and including additional trait axes would allow for further diversity, but is beyond the scope of this present study."

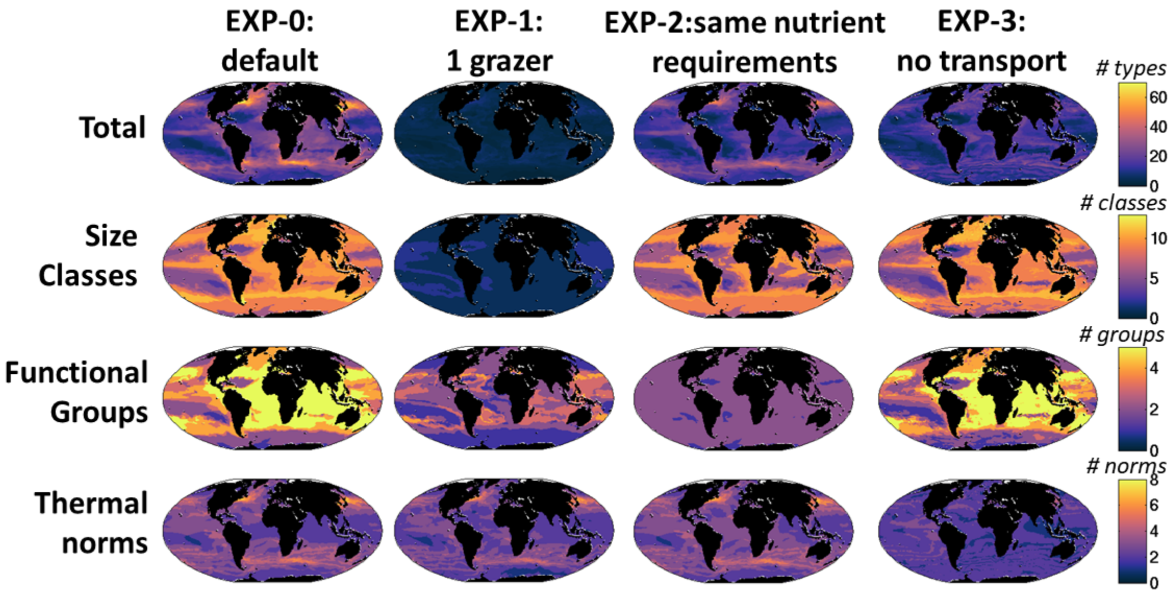
Supplementary: please provide the main parameters values. Is the term in parenthesis in eq. S1.4 (1/T-1/TN) or actually (1/TN-1/T)?

The phytoplankton allometric parameters are already given in Table 1 of the supplement (now Supplemental Table 2). We now add a new Supplemental Table 1(see below) which includes the values for all the other parameters mention in the supplemental text. We now direct the reader to these tables in the revised Supplemental Material). We feel that it would be confusing to include all the other model parameters not mentioned in the text since we could not adequately explain these. We however direct the reader to Dutkiewicz et al (2015) which has all the equations and all the parameters values listed. Almost all parameter values used here are identical to those used in that study. We have included a new section in the Supplemental (revised version Section S1.4: Model Parameters) where we explain more clearly where to find the appropriate parameters (e.g. Dutkiewicz et al (2015) Table 1 and 2 and those in our previous study), and detail the very few parameters that have been changed from Dutkiewicz et al 2015.

Equation S1.4 is correct.

New section in supplement:

"S1.4. Model Parameters: *We provide the values for the non-allometric parameters mentioned in the text above in Supplemental Table 1 and for the allometric parameters in Table 2. We refer the reader to Dutkiewicz et al (2015a) Tables 1 and 2 for the values of all other ecological and biogeochemical parameters used in this model. We note here only the few changes in parameter values: In Dutkiewicz et al (2015a) we had preferential remineralization of dissolved organic phosphorus (DOP) relative to other elements, here we do not. In this study, DOP remineralizes with same values (0.0333 d^{-1}) as the other elements. We found that CDOM was too high in this version of the model and increased the CDOM bleaching rate to 0.2592 d^{-1} from 0.167 d^{-1}."*



(Revised) Figure 11: Sensitivity simulations, model annual mean richness. EXP-1 has no size-dependent loss rates (i.e. only one grazer); EXP-2 has no nutrient requirement differences between functional groups; EXP-3 has no transport of the plankton (all nutrients and non-living organic pools are still transported). Top row: total richness; Second tow: size class richness determined by number of co-existing size classes; Thirdrow: functional richness determined by number of co-existing biogeochemical functional groups; Bottom row: thermal richness determined by number of co-existing temperature norms. The left most column are the same output as shown in Fig 9a,b,c,d for the original ("default") experiment, but with absolute values, not normalized.

	Symbol	Value	Units
normalization factor for temperature function	τ_T	0.8	unitless
	A_T	-4000	K
reference temperature	T_N	293.15	K
factor determining width of norms	B_T	3x10^{-4}	1/K
norm optimum temperature	T_{oj}	271.15 to 304.15 in 4K intervals	K
decay coefficient for norms	B	4	unitless
palatibility matrix	σ_{jk}	1 if grazer *k* is 10 times larger the prey *j*.	unitless

		0.3 if grazer *k* is 5 or 15 times larger than prey *j*	
grazing half saturation rate	k_p	1.5	mmolC/m^3

Supplemental Table S1: Non-allometric ecological parameters mentioned in this Supplemental