nature portfolio

Peer Review File

Regional restructuring in planktic foraminifera communities through Pliocene-early Pleistocene climate variability

Corresponding Author: Dr Ekaterina Larina

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This ms is very important study on 4D reconstruction of global plankton diversity in the intermediate time scale. While it's known in longer (Cenozoic through, eg Woodhouse, Swain, Fenton) and shorter time scales (Yasuhara et al 2020 PNAS), it's not very well understood in this time scale. It's exciting results and I enthusiastically recommend publication in your journal after moderate revision as explained below.

Major comments
1. While the MS is generally well written, it's bit rough here and there.
You may polish introduction more (in general, I tend to feel you can add a few more words there and there. Eg, justification to your statement that you say foram trend is similar to shark etc).
Also, some parts of discussion is little confusing (Pls see my specific comments below).

2. The discussion is rather complicated (I am not criticizing about it per se. Good to know such details that we didn't know before). If possible, it's good to try to add a summary figure (cartoon, conceptual diagram, etc).

3. Some S-Figs (on ecogroup relative abundances) show more distinct/remarkable trends than eg fig 3. You may consider to bring some to the main text fig (perhaps by combining some).

Specific comments:
1. Line 50,
It's good to note that the closure of CAS has two meanings.
Paleoceanographers tend to think it as closure of deep water connection or major water exchange.
Paleontologists (coastal and terrestrial) thin it as complete closure that prevent coastal benthos to move between Caribbean and Pacific and allow terrestrial mammals to move between S and N America. The timing of the former is much earlier (may see eg Daniela Schmidt paper, discussion between Aaron O'dea and Carlos Jaramillo).
So, CAS closure that affect planktic foram may be the former (and so may be earlier than the coastal benthic extinction of the Caribbean Sea eg studies by Aaron O'Dea and Jeremy Jackson)?

2. Good to explain how new from Woodhouse & Swain bit more. One is as I mentioned above you focused on the intermediate time scale that remained elusive. You may like to cite Yasuhara et al 2020 PNAS for short time scale study.

3. Line 117-118. A bit confusing sentence. I read as the LDG was stable for the last 15 but changed from 8 Ma? It sounds these two conflict? You may rephrase.

4. Line 123-124, ESI lowest at mid latitudes.
This mean you have no high latitude data, right?
This bit sounds ESI is high in low and high latitudes (=misleading).
You may revise like "ESI lower in higher latitudes"?

5. Line 128-129, this shift to the (sub)tropics is not super clear in fig 2C andf G.
I see what you mean though, Any way to better visualize this trend?

6. Line 133, FIg. 2B-D. You mean 2B, D, and F?

7. How about the artefact from differences in latitudinal range covered by S and N hemispheres. It could affect S and N hemisphere averaged ESIs (fig 2B, D, F)?

For sampling bias (between high and low latitudes and also N and S hemispheres),
you may like to have a look at discussion&summary in Yasuhara et al 2024 (hope this is useful):
Yasuhara et al 2024 Tropical paleobiology discovers biodiversity in a warmer past. PNAS

8. Line 166. New paragraph starting from "this". Meaning of "this" is not very clear.

9. Line 168-169. N Atlantic ecogroup diversity decline after the closure of CAS.
I don't see it in fig 4 green curve. Instead it look decreasing in the mPWP?

10. Line 182-183, stability before 3.6 Ma. To show it, it's good to show longer data like past 10 or 15 myr? Otherwise, it's bit difficult to clearly understand or see for readers. You have the data.

11. Line 186-187. It's good to specify ecogroups 1-5 that you mean in fig 3.
(for high-latitude (cool-water) and subthermocline (deep-water) ecological communities)

12. Line 188. you refer fig 2C but it;'s on the Atlantic, and here you are talking about Pacific?
You mean E?
Also here you say "high-latitude (cool-water) and subthermocline (deep-water) ecological
187 communities (Fig. 3, Supplementary Figs. 2, 6B, 7B) expanded their niches from mid to low188 latitudes (Fig. 2C),"
But Fig 2 shows only ESI and doesn't show such community information.
It may be something wrong here.

13. Line 205–. the formation of cryosphere. Good to specify the timing you are discussing here (when it was happened?). Bit confusing what you are discussing here.

14. Line 230-235, discussion on ecogroups. This is very important discovery!! I love this!!

15. Line 234 "higher trophic levels are expected to migrate towards the poles"
You may like to cite Yasuhara and Deutsch 2022 that has very relevant conceptual fig:
Yasuhara, M., Deutsch, C. A., 2022. Paleobiology provides glimpses of future ocean: Fossil records from tropical oceans predict biodiversity loss in a warmer world. Science: 375 (6576), 25–26.

16. North Atlantic is something special, right? As far as I remember, future cooling is projected in mid-high latitudes?. I think I discussed a bit in Yasuhara et al 2020 PNAS.

17. Fig 3. shading for iNHG, nPWP, and cCAS is in light colors. Good to make sure they are clear enough to be distinguiished.

18. Fig 4. may consider to change NH (green) and SH (blue) colors to be more distinct each other?

Sincerely yours,
Moriaki Yasuhara (this is signed review)

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This is the 3rd in a series of studies (Woodhouse et al., 2023; Swain et al., 2024) utilizing the Triton data base of geological planktonic foraminifera records to explore spatial and temporal variation in responses of pelagic biodiversity to climate change during major events in the Cenozoic.

Swain, A., Woodhouse, A., Fagan, W.F., Fraass, A.J., Lowery, C.M., 2024. Biogeographic response of marine plankton to Cenozoic environmental changes. Nature 629, 616-623.
Woodhouse, A., Swain, A., Fagan, W.F., Fraass, A.J., Lowery, C.M., 2023. Late Cenozoic cooling restructured global marine plankton communities. Nature 614, 713-718.

Triton is an exceptional resource, that was always going to have great potential to illuminate pelagic ecosystem responses to during significant ocean-climate episodes, leading to high impact papers (Nature and Science). Although the previous 2 papers coming out of Triton analysis, and now this one, follow a similar formula, i.e. bipartite analysis, and presentation of diversity metrics, the focus on different major climatic states or transitions for each manuscript, is sufficiently novel and the results sufficiently interesting and relevant to concerns about marine ecosystem changes, to be paradigm shifting in each case, and thus, potentially, worthy of high profile publication.

This current ms probes the major climate transitions of the Pliocene-Pleistocene, including the Pliocene warm period and the Pliocene-early Pleistocene onset of Northern hemisphere glaciation, which, as stated, “provides critical information about global climate and biodiversity dynamics that are widely regarded as the best analog to climate change projected for the end of this century”.

The manuscript is well written and well-organized and the results are extremely interesting. However, I have a major issue that to me stalls this manuscript going forward.

The connection to actual planktonic foraminifera observations, a reality check linking the black-box of a large database to the original foraminifera observations, including subsequent transformations to ecogroups, is missing. That is, the practical aspects of what is actually in the assemblage referred to, which species and species groups respond, and the robustness of their interpretation/classification into ecogroups). We are merely referred to previous interpretations of species ecology from a paper by Aze et al. 2011. Even the opening figure (Fig. 1) leaves me with doubt that the authors are in touch with the actual organisms in question because the icons used to represent the various ecogroups are a series of species that went extinct in the early Cenozoic. Maybe this is an oversight but combined with the minimal reference to species throughout the paper or methods or SI section (N. pachyderma and Pulleniatina are mentioned once)it does not add confidence. This to me is really important, since many of the species that existed during the time period of study, in contrast to the studies of earlier time periods, exist today and have an ecology that is known not only from stable isotope inferences (which is the case of extinct taxa) but from real plankton net studies.

For this reason I suggest rejection of the manuscript in this stage. The authors could come back and say this is an oversight, easily resolved, or if they cannot present convincing background methodology to show that the ecogroups are valid then reanalysis would be needed. This bcomes especially relevant because the Pliocene -Pleistocene species involved have living counterparts whose ecology is known from direct plankton net and culture studies, some knowledge on which might have evolved since the Aze et al. 2011 study.

Thus any revised manuscript needs at least in the supplementary information tables showing the species that make up the eco groups, modern updates, and a digest of the results /discussion, that connect the broad ecogroup diversity evolution back to tangible variations at the species level in the text, that is useful not only for big picture conclusions/predictions about patterns of change, but also to the fossil/plankton-collector observer level, who have work remaining to fill gaps in documenting and understanding especially higher latitude changes.

Another angle to emphasize in any revised manuscript would be the latitudinal limits of the study due to the data base content: it is limited to mid latitudes/temperate latitudes (no data from polar or subpolar latitudes).

OTHER COMMENTS

Lines 58-60: “The intensification of the Northern Hemisphere Glaciation impacted ocean temperatures and circulation and began a climate system dominated by glacial and interglacial cycles that continues to the present day”. The important aspects for planktonic forams will be ocean temperature, stratification, seasonality and food. Please connect these factors more directly to the climate transition in question upfront in the introductions. What happened to ocean circulation, how much cooling was there across latitudes in the different basins, how did this impact planktonic foram habitat distributions? These are background palaeoceanographic factors that can be summarized.

Line 78-79: ‘Modern’ is mentioned. Here an important comparison should also be the study (Strack et al., 2022-ref 31) about community shifts during one interglacial cycle, i.e. end of the last ice age, which as far as I am aware, is the closest (large scale warming even) of the past and an important yard stick for this study, especially considering a similar suite or species (or their close relatives) existed. I’d like to see more about this also in the discussion.
And it is important to make comparisons to studies of the modern and the Strack et al., 2020 deglacial-transition study: How do the methodologies compare? Are the resulting data comparable? Please elaborate.

Line 106: Good to have this line explaining how one interprets ESI values, which is a tricky concept to get your head around – you could prove this further by relating this directly to the figure, i.e. the red – blue colour scaling system on Figure 2. Its tricky to appreciate the ESI concepts and this supporting guide to what e.g. a strong blue colour, or strong red colour means xxx, and the means inverse yy, would be really helpful. Maybe also add it to the Fig caption for clarity (will be appreciated). Also the axis limits and divisions on the time axis. Can you use something more logical, than 0.3 my tick marks, i.e. integers (major ticks) and 0.2 or 0.5 myr (minor ticks) seems more user friendly to me.

Line 108-11
Given my comment about figure 1 (see above), please bring us back to something tangible regarding the species that go into

these eco groups by including a table in the supplementary information (modified from Aze et al) presenting lists of the species belonging to each Eco group utilized here, with appropriate icons in Fig. 1. This will restore my faith in the authors having knowledge of the species whose diversity data they are analyzing and connect readers/uses back to micropalaeontological observations, which is really important. It feels like we have lost touch in the abyss of the database.

Results section
-This section would benefit greatly from some anchoring sentences to frame the method here in the ms body text - so that the reader can directly connect with the results. At the moment the body text and analysis feels far way from the Triton data base and what it actually comprises. For example, I suggest at least mention how many sites are included in the Triton data base for the time internal of interest, and the latitudinal range (and refer directly to the Supp info figure here), also mention directly the types of materials/lithology – e.g. fossil census data based on pelagic carbonate rich pelagite core samples.
Also, Hhw did the Pliocene-early Pleistocene pelagic ecosystem changes compare with the older system responses? The Pliocene warming compared to previous Cenozoic warmings?

Discussion:
o In terms of the oceanographic explanation for the asynchronous patterns of biodiversity change during the Pliocene and Pliocene-Pleistocene transition, i.e. heterogeneous between ocean basins. The regions pointed to likely transitioned through different ocean overturning states/strengths through this period, for which there is a body of literature. AMOC is mentioned. Changes in surface currnents is also mentioned but less about the overturning states and their possible impact on stratificaiton, which would be important drivers of the surface current strengthening and stratification. This could be brought out more strongly, even if the data in the end point to relative stability in diversity through the peak warmth.There is even modelling evidence for major a possible PMOC in the Pliocene warm peak:
Burls, N.J., Fedorov, A.V., Sigman, D.M., Jaccard, S.L., Tiedemann, R., Haug, G.H., 2017. Active Pacific meridional overturning circulation (PMOC) during the warm Pliocene. Science Advances 3, e1700156.
o Point out that the analysis is limited by the limited latitudinal extent of the records in the data base. You do not have anything at polar latitudes, or even subpolar latitudes, the database content is limited to mid latitudes/temperate latitudes, according to ocean climate zone definitions.

FIGURES

• Figure 1. (see my comments above) I note that this figure has just been copied more or less from Woodhouse and Swain et al., 2024, showing silhouettes of planktonic forams representing the different ecogroups that absolutely did not exist (had gone extinct) during the time-interval under consideration (e.g. Morozovelloides/ hantkeninids). It's a bit concerning -since it suggests a disconnection between actual knowledge of Plio-Pleistocene planktonic foram assemblages and a blind analysis of data. This figure needs to be properly customized to include representative morphologies for all the eco groups that are actually characteristic of the Plio-Pliestocene, underpinned by isotopic/and or plankton net ecology data.

• Figure 2: Environmental proxies for the Pliocene-early Pleistocene 671 Climate and Ecogroup Specialization Index.
-I have to say, compared to the previous papers using the bipartite analysis, the more resticted time interval focus (i.e. of the bitartie analysis space-time map) is here at the edge of comfortable visual resolution. i.e. in the previous longer time studies these figures were visually impressive and informative, but here, with the 150 kyr-long time binds are stretched out to their limits so that the figures are not pleasent on the eye and require some standing back and squinting to see a pattern. If there is the possibilty to change the axis limits to 1.6- 4 Myrs, this could help, laterally squeeze everything. Or can you resample at 100 kyr bins? This could help with the visuals.
-What are the white areas in the figure? Data gaps. Clarify.

• Fig. 3. In the caption, explain ecogroup 3 for completeness.
-the blue colours are not so easy to distinguish in as curves. Can you choose a set of more ´distinctive colours for the ecogroup curves?
-The axis tick mark choice is not look especially nice. Other options? The axis looks better in Fig 4, so perhaps use the same axis limits and formatting for Fig. 3.

SM Figure 1. Please explain why there are no records above 60 degrees N in the North and nothing below 50 degrees S in the south. These high latitudes regions would be important for a holistic understanding of latitudinal pelagic sensitivities to global change but are missing. I know that there are deep sea drilling cores (and thus foram data) at least for the Atlantic and Indian Oceans with move Antarctic and sub Arctic coverage. Is this excluded from Triton?

(Remarks on code availability)
The code page exists where is says it exists. I have not looked further.

Reviewer #3

(Remarks to the Author)
Please see attached review.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I appreciate your careful and thorough revision. Figures are now much better and I like them very much.

Only a few things:
Line 76 “linked to biodiversity hotspots of important marine groups, including sharks”
Better to say “pelagic group” and just “biodiversity (instead of biodiversity hotspot)”, as coastal benthic groups have very different biodiversity trend/hotspot and also pelagic groups don’t show very distinct “hotpots”.

From line 88: “Although comprehensive studies of planktic foraminiferal 88 diversity over various temporal
89 scales, ranging from thousandse.g.,34,37-38 to millionse.g.,3,39 of years, exist; to date, none have
90 investigated patterns on large spatial scales with hundreds of sites.”

“none have (should be “has”?) investigated patterns on large spatial scales with hundreds of sites”
This is bit misleading as Swain et al Woodhouse et al and Fenton et al actually did such using Triton.
You may add a few more words to better specify (such as intermediate time scale or regionally resolved? I suspect somehow you missed to specify what you like to say here enough).

The paragraph from Line 100 on Triton:
You may consider to combine it into the previous paragraph, as it looks a bit redundant.

Otherwise I think the manuscript is now well revised.

Sincerely yours,
Moriaki Yasuhara

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Helen Coxall review round 2.

This is a much improved manuscript. My major comments have been addressed more or less completely – i.e. the authors have given attention to the details of the planktonic foraminifera assemblages and eco-groupings at the heart of this study, connecting us back to the data and observations.

I am largely happy, however, the new list of species raises a couple of foram specific questions that I would like to see addressed before this should be accepted for publication.

• Figure 1. A representative diagram of the planktic foraminifera ecogroups used in this
695 study.
Great. A vast improvement (even if the Neogene species do not have such nice silhouettes!)

• Figure 5. Conceptual representation of eco-group distributions during warmer (mPWP = the mid-Pliocene Warm Period
This is also a great addition and complement to Fig 1 bringing the paper together. Really helpful.

• Supplementary Table 1. Species of planktic foraminifera with assigned eco-groups used in our study from Aze, compared to ecogroups assigned by modern studies.

Thank you for making this table. Even though the eco-group interpretations have not much changed since Aze et al., its important and useful for the modern planktonic foraminifera expert to keep these check lists of species alive, and keep the references up to date. It also highlights the level of taxonomic detail that is contained in the data base, which is important.
-Please in the SI Table caption include a list of the eco-groups (again) to remind the reader what the numbers mean. Or perhaps since this is all about the ecor groups its worth while sorting them by eco group so we see the composition of each group.

• Now I see the lists of species I have a question about the North Atlantic. Why is the species Neogloboquadrina atlantica not listed in the table or eco-groups? It is extremely abundant/dominant in the North Atlantic in the Pliocene going extinct in the early Pleistocene – it is also an important biostratigraphic marker in the North Atlantic/subpolar region. It seems to be the key high latitude North Atlantic species before N. pachyderma took over that niche. If it is not in the Triton database then it must have been overlooked and binned in e.g. N. dutertrei. But is this true? It is important comment on this.

The data base and paper focus on macroperforate planktic foraminifera (single-celled calcareous zooplankton), the reason given that “Planktic foraminifera with microperforate and medioperforate wall textures were excluded from the analysis since they are currently characterized by a less refined phylogeny compared to macroperforate forms.” That may be the case, although there are some species that become important at high latitudes eg. Globigerinita uvula/glutinata. Morad et al., 2019, hints of hidden diversity among some of these small tax such as these.
-Can the authors at last comment on the extent to which the omission of microperforates might result in significant information missing in terms of diversity signals?

Morard, R., Vollmar, N.M., Greco, M., Kucera, M., 2019. Unassigned diversity of planktonic foraminifera from environmental sequencing revealed as known but neglected species. PLOS ONE 14, e0213936.

This microperforate issue also becomes relevant for this section:

Line 192-: “The negative correlation emerged concurrently with the onset of the intensification of the Northern Hemisphere Glaciation at ~2.7 Ma, and ecogroup 5 played a significant role in driving the observed ecological changes during the Pliocene - early Pleistocene interval (Supplementary Fig. 8). Curiously, the population of ecogroup 5 is represented by only two species — Neogloboquadrina pachyderma and Turborotalita quinqueloba — in contrast to other ecogroups (See supplementary table 1). Yet, despite its lower diversity, ecogroup 5 plays a significant role in driving changes observed during the Pliocene - early Pleistocene interval.”

This is where inclusion of microperforate taxa would make a difference to the diversity (+ 1 or 2 species). Also the inclusion of N. atlantica.

Important N. atlantica references

Weitkamp, T.M., Razmjooei, M.J., Pearson, P.N., Coxall, H.K., 2025. Upper Oligocene to Pleistocene planktonic foraminifera stratigraphy at North Atlantic DSDP Site 407, Reykjanes Ridge: diversity trends and biozonation using modern Neogene taxonomic concepts. J. Micropalaeontol. 44, 1-78.

Bolton, C.T., Bailey, I., Friedrich, O., Tachikawa, K., de Garidel-Thoron, T., Vidal, L., Sonzogni, C., Marino, G., Rohling, E.J., Robinson, M.M., Ermini, M., Koch, M., Cooper, M.J., Wilson, P.A., 2018. North Atlantic Midlatitude Surface-Circulation Changes Through the Plio-Pleistocene Intensification of Northern Hemisphere Glaciation. Paleoceanography and Paleoclimatology 33, 1186-1205.

Dowsett, H., Robinson, M., Foley, K., Herbert, T., Hunter, S., Andersson, C., Spivey, W., 2023. The Relative Stability of Planktic Foraminifer Thermal Preferences over the Past 3 Million Years. Geosciences 13, 71.

Flower, B.P., 1999. Data report: Planktonic foraminifers from the subpolar North Atlantic and Nordic Seas: Sites 980-987 and 907. Proceedings of the Ocean Drilling Program: Scientific Results 162, 19-34.

Weaver, P.P.E., Clement, B.M., 1986. Synchroneity of Pliocene planktonic foraminiferal datums in the North Atlantic. Marine Micropaleontology 10, 295-307.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Please see attached file.

(Remarks on code availability)

Response to Reviewers

Point-by-point responses to the reviewer comments are provided below in italics and in green font. All line numbers in our responses refer to the updated submission.

Reviewer #1:

This ms is very important study on 4D reconstruction of global plankton diversity in the intermediate time scale. While it's known in longer (Cenozoic through, eg Woodhouse, Swain, Fenton) and shorter time scales (Yasuhara et al 2020 PNAS), it's not very well understood in this time scale. It's exciting results and I enthusiastically recommend publication in your journal after moderate revision as explained below.

Thank you for your positive feedback and for recognizing the significance of our study on the intermediate time scale of global plankton diversity, Moriaki Yasuhara. We greatly appreciate your recommendation for publication and will address the suggested revisions carefully to further improve the manuscript.

Major comments
1. While the MS is generally well written, it's bit rough here and there.
You may polish introduction more (in general, I tend to feel you can add a few more words there and there. Eg, justification to your statement that you say foram trend is similar to shark etc).

Also, some parts of discussion is little confusing (Pls see my specific comments below).

Thank you for your thoughtful comments. We have revised the introduction to improve its clarity and flow. While we did not explicitly add a statement about sharks to the paper, we would like to provide additional context and references to clarify our reasoning.

The study by Worm et al. (2003, PNAS) discusses diversity hotspots for oceanic predators, stating:
"In the open ocean, many large predators such as tunas, sharks, billfishes, and sea turtles are of current conservation concern because of their vulnerability to overfishing and ecosystem role. Here we use scientific-observer records from pelagic longline fisheries in the Atlantic and Pacific Oceans to show that oceanic predators concentrate in distinct diversity hotspots. Predator diversity consistently peaks at intermediate latitudes (20–30° N and S), where tropical and temperate species ranges overlap. Individual hotspots are found close to prominent habitat features such as reefs, shelf breaks, or seamounts and often coincide with zooplankton and coral

reef hotspots." This study includes planktic foraminifera as part of the zooplankton associated with these hotspots.

Similarly, Titterson et al. (2010, Nature) provide a visual representation in their Figure 1, showing overlapping patterns of species richness for sharks, squids, and foraminifera, highlighting the consistency in diversity trends across these groups.

We have also refined the discussion to address the specific points raised and to improve its overall clarity. We hope these changes better address your comments and clarify our reasoning.

2. The discussion is rather complicated (I am not criticizing about it per se. Good to know such details that we didn't know before). If possible, it's good to try to add a summary figure (cartoon, conceptual diagram, etc).

Thank you for your suggestion. We have created a conceptual figure to summarize the discussion; please see Figure 5. We hope this addition provides clarity and enhances the overall understanding of the key points.

3. Some S-Figs (on ecogroup relative abundances) show more distinct/remarkable trends than eg fig 3. You may consider to bring some to the main text fig (perhaps by combining some).

As suggested, we have moved SM Figure 2 into the main text, where it is now Figure 3, and moved the previous Figure 3 to the supplementary material. We hope this improves the clarity and impact of the presentation.

Specific comments:
1. Line 50,
It's good to note that the closure of CAS has two meanings.
Paleoceanographers tend to think it as closure of deep water connection or major water exchange.
Paleontologists (coastal and terrestrial) thin it as complete closure that prevent coastal benthos to move between Caribbean and Pacific and allow terrestrial mammals to move between S and N America. The timing of the former is much earlier (may see eg Daniela Schmidt paper, discussion between Aaron O'dea and Carlos Jaramillo).
So, CAS closure that affect planktic foram may be the former (and so may be earlier than the coastal benthic extinction of the Caribbean Sea eg studies by Aaron O'Dea and Jeremy Jackson)?

Thank you for your insightful comment. The planktic foraminiferal distribution should not be affected by the closure of deep water flow, and the isolation of the two ocean basins at mixed-layer depth is clearly demonstrated by the divergent values of planktonic isotopes starting around 4 Ma (see Figure 3D in O'Dea et al., 2016). We have revised the text as suggested (Lines 47ff; Line 232-236).

2. Good to explain how new from Woodhouse & Swain bit more. One is as I mentioned above

you focused on the intermediate time scale that remained elusive. You may like to cite Yasuhara et al 2020 PNAS for short time scale study.

Thank you for your comment. We have clarified how our study differs from Woodhouse & Swain and other. Additionally, we have cited Yasuhara et al. (2020, PNAS) to provide context on shorter timescale studies (see lines 88-99.)

3. Line 117-118. A bit confusing sentence. I read as the LDG was stable for the last 15 but changed from 8 Ma? It sounds these two conflict? You may rephrase.

Thank you for pointing out the potential confusion in this sentence. To improve clarity and avoid conflicting interpretations, we have removed the discussion of the last 15 Ma.

4. Line 123-124, ESI lowest at mid latitudes. This mean you have no high latitude data, right? This bit sounds ESI is high in low and high latitudes (=misleading).
You may revise like "ESI lower in higher latitudes"?

Thank you for your thoughtful comment. The Triton database does include records from high latitudes, and in our study, we analyzed data from the North Atlantic up to 65°N, covering the subpolar region. However, high-latitude data availability is often limited to specific intervals, such as the mid-Piacenzian Warm Period, due to patchy and inconsistent sampling across the Pliocene–early Pleistocene. Given the high-resolution nature of our study, which required consistent temporal and spatial coverage, these regions frequently failed to meet the criteria for sample completeness when analyzed in 150-kyr by 5° latitude bins. We have revised the text to clarify this and to emphasize that the lowest ESI zones, indicative of the highest ecogroup equitability, are predominantly located in mid-latitude regions compared to lower latitudes across all basins (see lines 170-171; 175-177).

5. Line 128-129, this shift to the (sub)tropics is not super clear in fig 2C andf G.
I see what you mean though, Any way to better visualize this trend?

To address your suggestion and enhance the visualization, we have added a black line to the figure to better highlight the pattern. See Figure 2b,2f.

6. Line 133, FIg. 2B-D. You mean 2B, D, and F?

We updated the figure numbering.

7. How about the artefact from differences in latitudinal range covered by S and N hemispheres. It could affect S and N hemisphere averaged ESIs (fig 2B, D, F)?
For sampling bias (between high and low latitudes and also N and S hemispheres),
you may like to have a look at discussion&summary in Yasuhara et al 2024 (hope this is useful):
Yasuhara et al 2024 Tropical paleobiology discovers biodiversity in a warmer past. PNAS

Thank you for your suggestion. We have included a discussion of the statistical tests performed to address the limitations of unequal sample sizes between hemispheres and basins. Specifically, we *incorporated resampling analyses and sampling coverage metrics to ensure the robustness of our findings (see lines 133-139; Supplementary Figs. 13-17). These analyses confirm that observed patterns are not artifacts of sampling bias.*

8. Line 166. New paragraph starting from "this". Meaning of "this" is not very clear.

We edited the sentence for further clarity.

9. Line 168-169. N Atlantic ecogroup diversity decline after the closure of CAS.
I don't see it in fig 4 green curve. Instead it look decreasing in the mPWP?

We have reviewed the text and corrected it to align with the pattern observed in the Northern Hemisphere of the Atlantic in Fig. 4. We appreciate your careful observation.

10. Line 182-183, stability before 3.6 Ma. To show it, it's good to show longer data like past 10 or 15 myr? Otherwise, it's bit difficult to clearly understand or see for readers. You have the data.

Thank you for your suggestion. Although we have data extending beyond 3.6 Ma, the analysis of the last 15 Ma falls outside the scope of this study. To address your concern, we have revised the text (see lines 234-239). For broader context, the study by Woodhouse and Swain et al. (2023) demonstrates similar patterns over longer timescales, albeit at a coarser resolution and at an averaged global scale, supporting the overall picture presented in our analysis.

11. Line 186-187. It's good to specify ecogroups 1-5 that you mean in fig 3. (for high-latitude (cool-water) and subthermocline (deep-water) ecological communities)

We have revised the text as suggested.

12. Line 188. you refer fig 2C but it;'s on the Atlantic, and here you are talking about Pacific? You mean E?
Also here you say "high-latitude (cool-water) and subthermocline (deep-water) ecological
187 communities (Fig. 3, Supplementary Figs. 2, 6B, 7B) expanded their niches from mid to
low188 latitudes (Fig. 2C),"
But Fig 2 shows only ESI and doesn't show such community information.
It may be something wrong here.

Thank you for your detailed feedback. You are correct that Fig. 2 shows only ESI and does not display community information, and that the reference should be to Fig. 2E, not Fig. 2C. We have revised the text to clarify that we are referring to changes in the proportional abundances of the ecogroups rather than niche expansion.

13. Line 205–. the formation of cryosphere. Good to specify the timing you are discussing here (when it was happened?). Bit confusing what you are discussing here.

We have revised the text as suggested by the Reviewer 1. See line 271.

14. Line 230-235, discussion on ecogroups. This is very important discovery!! I love this!!

Thank you for your enthusiastic feedback! We're thrilled that you find the discovery significant and appreciate your support for our work.

15. Line 234 “higher trophic levels are expected to migrate towards the poles”
You may like to cite Yasuhara and Deutsch 2022 that has very relevant conceptual fig:
Yasuhara, M., Deutsch, C. A., 2022. Paleobiology provides glimpses of future ocean: Fossil records from tropical oceans predict biodiversity loss in a warmer world. Science: 375 (6576), 25–26.

We have added the citation for Yasuhara and Deutsch (2022) to support the statement regarding the migration of higher trophic levels towards the poles. We appreciate you pointing us to this relevant reference and conceptual figure.

16. North Atlantic is something special, right? As far as I remember, future cooling is projected in mid-high latitudes? I think I discussed a bit in Yasuhara et al 2020 PNAS.

We have added the suggested citation (Yasuhara et al., 2020 PNAS) to emphasize the unique characteristics of the North Atlantic, particularly its sensitivity to climate-driven changes and the strong influence of sea surface temperature on planktic species distribution.

17. Fig 3. shading for iNHG, nPWP, and cCAS is in light colors. Good to make sure they are clear enough to be distinguiished.

As suggested, we improved the figure quality.

18. Fig 4. may consider to change NH (green) and SH (blue) colors to be more distinct each other?

As suggested, we changed the colors. See Figure 4.

Sincerely yours,
Moriaki Yasuhara (this is signed review)

Reviewer #2:

This is the 3rd in a series of studies (Woodhouse et al., 2023; Swain et al., 2024) utilizing the Triton data base of geological planktonic foraminifera records to explore spatial and temporal variation in responses of pelagic biodiversity to climate change during major events in the Cenozoic.

Swain, A., Woodhouse, A., Fagan, W.F., Fraass, A.J., Lowery, C.M., 2024. Biogeographic response of marine plankton to Cenozoic environmental changes. Nature 629, 616-623.
Woodhouse, A., Swain, A., Fagan, W.F., Fraass, A.J., Lowery, C.M., 2023. Late Cenozoic cooling restructured global marine plankton communities. Nature 614, 713-718.

Triton is an exceptional resource, that was always going to have great potential to illuminate pelagic ecosystem responses to during significant ocean-climate episodes, leading to high impact papers (Nature and Science). Although the previous 2 papers coming out of Triton analysis, and now this one, follow a similar formula, i.e. bipartite analysis, and presentation of diversity metrics, the focus on different major climatic states or transitions for each manuscript, is sufficiently novel and the results sufficiently interesting and relevant to concerns about marine ecosystem changes, to be paradigm shifting in each case, and thus, potentially, worthy of high profile publication.

This current ms probes the major climate transitions of the Pliocene-Pleistocene, including the Pliocene warm period and the Pliocene-early Pleistocene onset of Northern hemisphere glaciation, which, as stated, "provides critical information about global climate and biodiversity dynamics that are widely regarded as the best analog to climate change projected for the end of this century".

Thank you for recognizing the importance of our study and for your detailed and thoughtful feedback.

The manuscript is well written and well-organized and the results are extremely interesting. However, I have a major issue that to me stalls this manuscript going forward.

The connection to actual planktonic foraminifera observations, a reality check linking the black-box of a large database to the original foraminifera observations, including subsequent transformations to ecogroups, is missing. That is, the practical aspects of what is actually in the assemblage referred to, which species and species groups respond, and the robustness of their interpretation/classification into ecogroups. We are merely referred to previous interpretations of species ecology from a paper by Aze et al. 2011. Even the opening figure (Fig. 1) leaves me with doubt that the authors are in touch with the actual organisms in question because the icons used to represent the various ecogroups are a series of species that went extinct in the early Cenozoic. Maybe this is an oversight but combined with the minimal reference to species throughout the paper or methods or SI section (N. pachyderma and Pulleniatina are mentioned once) it does not add confidence. This to me is really important, since many of the species that existed during the time period of study, in contrast to the studies of earlier time periods, exist today and have an ecology that is known not only from stable isotope inferences (which is the case of extinct taxa) but from real plankton net studies.

For this reason I suggest rejection of the manuscript in this stage. The authors could come back and say this is an oversight, easily resolved, or if they cannot present convincing background methodology to show that the ecogroups are valid then reanalysis would be needed. This becomes especially relevant because the Pliocene -Pleistocene species involved have living counterparts whose ecology is known from direct plankton net and culture studies, some knowledge on which might have evolved since the Aze et al. 2011 study.

Thus any revised manuscript needs at least in the supplementary information tables showing the species that make up the eco groups, modern updates, and a digest of the results /discussion, that connect the broad ecogroup diversity evolution back to tangible variations at the species level in the text, that is useful not only for big picture conclusions/predictions about patterns of change, but also to the fossil/plankton-collector observer level, who have work remaining to fill gaps in documenting and understanding especially higher latitude changes.

Another angle to emphasize in any revised manuscript would be the latitudinal limits of the study due to the data base content: it is limited to mid latitudes/temperate latitudes (no data from polar or subpolar latitudes).

We have addressed your concerns as follows:

Connection to Observations and Ecogroup Validity:

We acknowledge the importance of validating the ecogroup classifications against modern studies and observations. To address this, we have included a supplementary table listing the species that comprise each ecogroup, updated with modern ecological studies of extant planktic foraminifera (see supplementary Table 1). Specifically, we compared the ecogroups used in our study, based on Aze et al. (2011), with modern studies employing geochemical and plankton net data, and updated them where appropriate. The only discrepancy identified in our study was the classification of Turborotalita humilis. While we initially placed it within ecogroup 2, modern studies confirm that it belongs to ecogroup 1. We have updated this information and included T. humilis in our analysis as part of ecogroup 1. This adjustment did not impact any results, and the current manuscript now contains the updated figures.

Additionally, we have included spatio-temporal raw occurrences of species across various ecogroups in the supplementary materials. See extended data files. We hope this addition improves transparency and connects our ecogroups back to species-level observations.

Correction of Figure 1: *Thank you for noting the issue with the icons used in Figure 1. We have corrected this oversight and replaced the faulty forms with representations of extant species corresponding to the studied time interval. We believe this revision aligns the figure with the scope of the study and addresses your concerns.*

Latitudinal Limits and Subpolar Regions: *The Triton database does include records from high latitudes, and in our study, we analyzed data from the North Atlantic up to 65°N, covering the subpolar region. However, in high latitudes across different basins, data availability is often limited to specific intervals, such as the mid-Piacenzian Warm Period or other short timeframes.*

Given the high-resolution nature of our study, which required consistent temporal and spatial coverage, high-latitude regions frequently failed to meet the criteria for sample completeness when analyzed in 150 kyr by 5° latitude bins. These regions often showed patchy, inconsistent data across the Pliocene–early Pleistocene. Therefore, to maintain the robustness and resolution of our analyses, we excluded these regions.

We hope these revisions address your concerns and further strengthen the manuscript. Thank you again for your time and effort in reviewing our work.

OTHER COMMENTS

Lines 58-60: "The intensification of the Northern Hemisphere Glaciation impacted ocean temperatures and circulation and began a climate system dominated by glacial and interglacial cycles that continues to the present day". The important aspects for planktonic forams will be ocean temperature, stratification, seasonality and food. Please connect these factors more directly to the climate transition in question upfront in the introductions. What happened to ocean circulation, how much cooling was there across latitudes in the different basins, how did this impact planktonic foram habitat distributions? These are background palaeoceanographic factors that can be summarized.

Thank you for your thoughtful feedback. Considering that this is the introduction and needs to remain succinct, we have incorporated key aspects of ocean circulation, cooling across latitudes, and their impacts on marine organisms into this section to provide essential context (see lines 63-70). Additional details addressing your questions are included in the discussion to elaborate on these palaeoceanographic factors in the appropriate depth (lines 270ff). We hope this approach balances clarity and conciseness.

Line 78-79: 'Modern' is mentioned. Here an important comparison should also be the study (Strack et al., 2022-ref 31) about community shifts during one interglacial cycle, i.e. end of the last ice age, which as far as I am aware, is the closest (large scale warming even) of the past and an important yard stick for this study, especially considering a similar suite or species (or their close relatives) existed. I'd like to see more about this also in the discussion.
And it is important to make comparisons to studies of the modern and the Strack et al., 2020 deglacial-transition study: How do the methodologies compare? Are the resulting data comparable? Please elaborate.

Thank you for your insightful comment and for highlighting the importance of Strack et al. (2022) in understanding community shifts during the last deglacial transition. We recognize the significance of their work and have cited it where appropriate in the manuscript to acknowledge its relevance. However, we believe that detailed comparisons to the last ice age and its transitions are beyond the scope of this study. Our focus is on broader temporal trends and major climate transitions during the Pliocene–Pleistocene, which provide critical insights into biodiversity dynamics over longer timescales.

While comparisons to the deglacial-transition study would be valuable, there are many other important time intervals or other studies that could be used for comparison and exploring all of these would dilute the focus of our current analysis.

Line 106: Good to have this line explaining how one interprets ESI values, which is a tricky concept to get your head around – you could prove this further by relating this directly to the figure, i.e. the red – blue colour scaling system on Figure 2. Its tricky to appreciate the ESI concepts and this supporting guide to what e.g. a strong blue colour, or strong red colour means xxx, and the means inverse yy, would be really helpful. Maybe also add it to the Fig caption for clarity (will be appreciated). Also the axis limits and divisions on the time axis. Can you use something more logical, than 0.3 my tick marks, i.e. integers (major ticks) and 0.2 or 0.5 myr (minor ticks) seems more user friendly to me.

Thank you for your helpful feedback. We have updated Figure 2 and its caption to clarify how ESI values relate to the yellow-blue color scaling and what the colors represent. Additionally, we revised the time axis to use more logical scaling. See Figure 2.

Line 108-11
Given my comment about figure 1 (see above), please bring us back to something tangible regarding the species that go into these eco groups by including a table in the supplementary information (modified from Aze et al) presenting lists of the species belonging to each Eco group utilized here, with appropriate icons in Fig. 1. This will restore my faith in the authors having knowledge of the species whose diversity data they are analyzing and connect readers/uses back to micropalaeontological observations, which is really important. It feels like we have lost touch in the abyss of the database.

Thank you for your detailed comment. In response, we have updated Figure 1 to include appropriate icons representing the species relevant to the ecogroups studied. Additionally, as suggested, we have included a supplementary table listing the species belonging to each ecogroup from Aze et al. (2011) and from modern studies (see Supplementary Table 1). We hope these additions provide the tangible connection to species-level data and micropalaeontological observations you requested and restore confidence in the robustness of our analysis.

Results section
-This section would benefit greatly from some anchoring sentences to frame the method here in the ms body text - so that the reader can directly connect with the results. At the moment the body text and analysis feels far way from the Triton data base and what it actually comprises. For example, I suggest at least mention how many sites are included in the Triton data base for the time internal of interest, and the latitudinal range (and refer directly to the Supp info figure here), also mention directly the types of materials/lithology – e.g. fossil census data based on pelagic carbonate rich pelagite core samples.
Also, Hhw did the Pliocene-early Pleistocene pelagic ecosystem changes compare with the older system responses? The Pliocene warming compared to previous Cenozoic warmings?

Thank you for your thoughtful suggestions. We have included the number of sites analyzed (370 cores) in Supplementary Figure 1 as well as in the Methods section (see lines 331-333).

Additionally, the latitudinal range for each basin can be observed in Figure 2 and in Supplementary figures 1, 10-12, 17. Regarding lithology, due to the large number of cores analyzed across 3 different basins (370 in total), the materials vary, making it impractical to discuss them in detail within the manuscript and beyond the scope of this study. However, anyone can access the core IDs in our Triton_df_Oceans.RData file (available at https://github.com/ekatlarina/forams-ecogroups) and explore additional lithological details through the corresponding core repositories.

Discussion:

o In terms of the oceanographic explanation for the asynchronous patterns of biodiversity change during the Pliocene and Pliocene-Pleistocene transition, i.e. heterogeneous between ocean basins. The regions pointed to likely transitioned through different ocean overturning states/strengths through this period, for which there is a body of literature. AMOC is mentioned. Changes in surface currnents is also mentioned but less about the overturning states and their possible impact on stratificaiton, which would be important drivers of the surface current strengthening and stratification. This could be brought out more strongly, even if the data in the end point to relative stability in diversity through the peak warmth. There is even modelling evidence for major a possible PMOC in the Pliocene warm peak:
Burls, N.J., Fedorov, A.V., Sigman, D.M., Jaccard, S.L., Tiedemann, R., Haug, G.H., 2017. Active Pacific meridional overturning circulation (PMOC) during the warm Pliocene. Science Advances 3, e1700156.

Thank you for your constructive feedback. We have incorporated a discussion on overturning states and their impact on stratification and surface currents, referencing Burls et al. (2017) to highlight the potential role of the Pacific Meridional Overturning Circulation (PMOC) during the mPWP. These revisions aim to provide a stronger oceanographic context to our findings (see lines 252ff).

o Point out that the analysis is limited by the limited latitudinal extent of the records in the data base. You do not have anything at polar latitudes, or even subpolar latitudes, the database content is limited to mid latitudes/temperate latitudes, according to ocean climate zone definitions.

Thank you for your thoughtful comment. The Triton database does include records from high latitudes, and in our study, we analyzed data from the North Atlantic up to 65°N, covering the subpolar region. However, in high latitudes across different basins, data availability is often limited to specific intervals, such as the mid-Piacenzian Warm Period or other short timeframes. Given the high-resolution nature of our study, which required consistent temporal and spatial coverage, high-latitude regions frequently failed to meet the criteria for sample completeness when analyzed in 150 kyr by 5° latitude bins. These regions often showed patchy, inconsistent data across the Pliocene–early Pleistocene. Therefore, to maintain the robustness and resolution of our analyses, we excluded these regions. See lines 178-180.

FIGURES

Figure 1. (see my comments above) I note that this figure has just been copied more or less from Woodhouse and Swain et al., 2024, showing silhouettes of planktonic forams representing the

different ecogroups that absolutely did not exist (had gone extinct) during the time-interval under consideration (e.g. Morozovelloides/ hantkeninids). It's a bit concerning -since it suggests a disconnection between actual knowledge of Plio-Pleistocene planktonic foram assemblages and a blind analysis of data. This figure needs to be properly customized to include representative morphologies for all the eco groups that are actually characteristic of the Plio-Pliestocene, underpinned by isotopic/and or plankton net ecology data.

Thank you for noting the issue with the icons used in Figure 1. We have corrected this oversight and replaced the faulty forms with representatives of extant species corresponding to the studied time interval. We believe this revision aligns the figure with the scope of the study and addresses your concerns.

Figure 2: Environmental proxies for the Pliocene-early Pleistocene 671 Climate and Ecogroup Specialization Index.
-I have to say, compared to the previous papers using the bipartite analysis, the more resticted time interval focus (i.e. of the bitartie analysis space-time map) is here at the edge of comfortable visual resolution. i.e. in the previous longer time studies these figures were visually impressive and informative, but here, with the 150 kyr-long time binds are stretched out to their limits so that the figures are not pleasent on the eye and require some standing back and squinting to see a pattern. If there is the possibilty to change the axis limits to 1.6- 4 Myrs, this could help, laterally squeeze everything. Or can you resample at 100 kyr bins? This could help with the visuals.
-What are the white areas in the figure? Data gaps. Clarify.

Thank you for your detailed feedback. The white areas in the figure indeed represent data gaps or bins that did not meet our data completeness criteria and were therefore excluded from the analysis. We clarified in the figure caption as suggested (see Figure 2). Regarding your suggestion about time bins, we tested the analysis using different time bin sizes, including 100 kyr, but found that 150 kyr time bins by 5° latitude provided the most complete and comprehensive coverage for our time frame. To address your concerns about visual resolution, we have improved the figure by adding space between panels, modifying the color scheme for better distinction, and including solid black lines to emphasize key patterns. We hope these updates enhance the figure's clarity and visual appeal.

Fig. 3. In the caption, explain ecogroup 3 for completeness.
-the blue colours are not so easy to distinguish in as curves. Can you choose a set of more ´distinctive colours for the ecogroup curves?
-The axis tick mark choice is not look especially nice. Other options? The axis looks better in Fig 4, so perhaps use the same axis limits and formatting for Fig. 3.

Thank you for your suggestions. We have updated the figure as recommended: the colors for the ecogroup curves have been adjusted to make them more distinctive, and the axis limits and formatting have been aligned with those used in Fig. 4 for consistency. Additionally, we have moved this figure to the supplementary material as Fig. 2 and updated the caption to include an explanation of all ecogroups, including ecogroup 3 for completeness.

SM Figure 1. Please explain why there are no records above 60 degrees N in the North and nothing below 50 degrees S in the south. These high latitudes regions would be important for a holistic understanding of latitudinal pelagic sensitivities to global change but are missing. I know that there are deep sea drilling cores (and thus foram data) at least for the Atlantic and Indian Oceans with move Antarctic and sub Arctic coverage. Is this excluded from Triton?

Thank you for your thoughtful comment. The Triton database does include records from high latitudes, and in our study, we analyzed data from the North Atlantic up to 65°N, covering the subpolar region. However, in high latitudes across different basins, data availability is often limited to specific intervals, such as the mid-Piacenzian Warm Period or other short timeframes. Given the high-resolution nature of our study, which required consistent temporal and spatial coverage, high-latitude regions frequently failed to meet the criteria for sample completeness when analyzed in 150 kyr by 5° latitude bins. These regions often showed patchy, inconsistent data across the Pliocene–early Pleistocene.

Therefore, to maintain the robustness and resolution of our analyses, we excluded these regions. However, we note that Woodhouse and Swain et al. (2023) provide complementary insights by analyzing polar and subpolar regions of ESI for all basins combined at 250 kyr time bins. Their study highlights the expected pattern of high ESI, or lower evenness, in high latitudes, aligning with ecological expectations. While our focus remains on high-resolution basin-specific analyses, these findings further reinforce the patterns observed at higher latitudes.

Reviewer #2 (Remarks on code availability):

The code page exists where is says it exists. I have not looked further.

Reviewer #3:

The authors have analyzed taxonomic and functional diversity of a well-resolved plankton group across an interval of Earth history relevant to present-day climate warming. I was interested to read this article from its premise alone, and I would be happy to see an article on this topic appear in Nature Communications. My comments are mostly minor (although see two larger methodological comments/questions), and most relate to improving clarity, especially of the methodological details, or softening the interpretive assertions to be less definitive. I apologize to the authors for my long delay in returning this review and hope the detail of constructive comments below helps offset the inconvenience.

Thank you for your thoughtful review and constructive comments. We are grateful for your positive feedback on the relevance of our study and your detailed suggestions, which will help us improve the clarity and robustness of the manuscript. We appreciate your time and effort in providing this valuable input.

Abstract and Introduction

1. Line 21 and 42: I suggest rewording "recent observations" to something like "observations of living species". "Recent observations" could mean samples of living species taken in the last few years or, geologically speaking, samples from the Recent.

We have edited the text to clarify that we are referring to recent studies of both past and modern planktonic communities (lines 22, 37-38).

2. Line 47: Do you mean "life system responses"? I wouldn't consider plankton part of the Earth system. You could also be wordier but explicit, "life system responses to Earth system change" or similar.

We have revised the sentence to explicitly state "how biological systems respond to changes in the Earth system," as this more accurately conveys the intended meaning (line 44).

3. Line 48: Define Ma at first use for non-geology/paleontology readers

We defined Ma as 'millions of years old' for non-geology readers (line 45).

4. Line 50: This is referring to the rise of the Panama isthmus, correct? I would include reference to that by name, since I think it's a more well-known geographic feature than the Central American Seaway. Also consider additional references – the single citation for the given age date was published several years ago now and there's been heated debate about the exact timing of the closure.

Thank you for your comment. We have updated the text and acknowledged the ongoing debate regarding the timing of the closure of the Central American Seaway (lines 48ff).

5. Line 80: I understand the typical narrative structure of introductions is to include a sentence like this ("despite some prior study, there is still a gap in understanding X"), but I suggest

omitting the sentence here beginning "Despite ongoing efforts…". There have arguably been more comprehensive and high-resolution comparative studies of zooplankton responses to paleoclimate change than for any other group except bivalves. That doesn't make the topic less deserving of study! Simply merge some the language from the sentence with the next two sentences for flow in the paragraph. For example, something like,

"The recently collated Triton dataset, which provides an extensive record of global fossil occurrences of planktic foraminifera, allows for comprehensive, high-resolution studies of zooplankton community responses to paleoclimate change across regional scales."

Thank you for your suggestion. We have revised the text to merge the language for better flow and to avoid overstating gaps in prior research (See lines 88ff).

Results

6. The results section needs to include a several-sentence introduction to what bipartite network analysis is and how it calculates ESI, including a definition of equitability that doesn't include the word equitable. I know there is prior literature applying this method to similar questions, but this paper needs to be fully understandable on its own for a reader who may be unfamiliar with that previous work and with network theory. How would you explain this to an advanced undergraduate learning about it for the first time?

Thank you for your feedback. We have revised the text to clarify the concept of the Ecogroup Specialization Index (ESI). Further clarification on how the Ecogroup Specialization Index (ESI) or Specificity Specialization Index is calculated can be found in Julliard[67] and Poissot[68]. Instead of defining "equitability" in general, we provide an explanation of what equitable means within the context of our study, which we hope increases clarity about this metric as suggested by the Reviewer (Lines 142-149)

7. The information currently in the Methods about how to interpret ESI (ranging from 0 to 1, with 1 indicating specialization) needs to be repeated here as ESI results are presented. Readers need this context to interpret Fig 2, for example.

We have revised the text as suggested by the Reviewer 3 (see lines 142-149).

8. Line 126: I would say "might be" the result of thermocline differences rather than "likely" due to it. There are alternative hypotheses, and reality is usually a mix of processes at work. Without data on the thermocline or an estimate of its effect relative to other processes, it feels too assertive to me to assume the pattern is due to thermocline differences alone.

We have revised the text as suggested by the Reviewer (lines 173).

9. Line 128: Reword to avoid the words "subtropical" and "tropical". The location of the tropics has changed through time, so it is better to discuss the location of biogeographic patterns relative to static latitudinal markers.

We have reworded the sentence as suggested (lines 176).

10. Line 128: Say "shift" rather than "noticeable shift". I agree it's present, but it's a slight and qualitative signal rather than jumping off the page. Also remove the word "significant" from the following sentence unless it is associated with a statistical test result.

We have reworded the sentence as suggested (lines 175).

11. Line 136: This study calculates relative occurrence frequency, not species' niche breadth or position. Changes in relative occurrences could be due to niche occupancy changes but could also result from changes in the spatial extent of species' habitats or other factors. Please reword to avoid the phrase "ecological niche expansion" here. Similarly, please remove the phrase "reconfiguration of ecological niches" in the caption to fig 3.

We have reworded the sentence as suggested by the Reviewer (lines 186-189) and revised the caption for Figure 3 to eliminate the phrase "reconfiguration of ecological niches," addressing your concern (see supplementary figure 2).

12. Line 150: Environment is presumably stable along both vertical and latitudinal axes, right?

Yes, that's correct.

13. Line 167: The in-text results in the sentence beginning "The highest rate of change…" don't appear match what I see in Fig 4. It's unclear that ecogroup diversity did drop after cCAS in the North Atlantic, and I certainly wouldn't call it a "sharp decline". Rather, the green line for the N Atlantic barely drops (not significant) in the first three time bins proceeding from the cCAS, and it then drops at the onset of the mPWP (Fig 4B). The "highest rate of change in ecogroup diversity" instead looks like it might be in the North Pacific around 3.4 Ma.

14. Line 171: As in comment #13, the text description here appears to differ from what I see in the figures. Ecogroup diversity does not consistently increase in the Northern Pacific (second panel of Fig 4B) as the sentence beginning on this line asserts ("a consistent increase in ecogroup diversity is evidence across all basins in the Northern Hemisphere"). I am also unconvinced that the changes happened exactly at the iNHG, in either fig 3 or 4, as stated in the captions there and in line 172. The onset of a directional trend could be argued to originate anytime between the mPWP and iNHG from figs 3 and 4, and it is ultimately a qualitative call rather than a quantitative break-point estimate.

Reply to both point 13 and 14: *Thank you for your detailed feedback regarding Fig. 4B. We have carefully reviewed the patterns and revised the text accordingly. Our analysis indicates that the largest changes in ecogroup diversity occurred in both the North Atlantic and North Pacific, though during different periods. In the North Atlantic, ecogroup diversity increased immediately after the mPWP (around 3 Ma), while in the North Pacific, the increase occurred earlier, between 3.7 and 3.2 Ma.*

The magnitude of change in both basins was similar, around 0.8 corrected sampled-in-bin diversity (SQS diversity). We have clarified these points in the text and adjusted the interpretation to reflect the timing and magnitude of these changes more accurately (lines 220ff).

15. Consider discussing limitations of unequal sample size between hemispheres and basins

Thank you for your suggestion. We have included a discussion of the statistical tests performed to address the limitations of unequal sample sizes between hemispheres and basins. Specifically, we incorporated resampling analyses and sampling coverage to ensure the robustness of our findings. These analyses confirm that observed patterns are not artifacts of sampling bias (lines 133-139).

Discussion

16. Line 177: I would suggest an additional reference here as in the Introduction, as this timing has been a contentious topic.

We revised the text to as suggested by the Reviewer.

17. Sentence beginning 182: I'm unable to follow the reasoning of this sentence. The evidence cited in the preceding sentence seems to be from around the closure at 2.8 Ma, but also implies this study found matching changes beginning at 3.6. Those timings don't match up perfectly. In addition, the "global stability…prior to 3.6 Ma" mentioned here isn't previously mentioned in the Results and is based on only 2 data points (between 3.6 and 3.9 Ma). I agree that the closing of the CAS was a major biogeographic event, and it is worth mentioning in the Discussion. That said, looking at the figures, e.g., Fig 3 and 4, I struggle to see any features of the data around then that couldn't be attributable to white-noise variation (no rapid directional changes between 3.5 and 3.9 Ma). I suggest walking back the strength of the claims in this paragraph. If there were major shifts from before vs. after the cCAS (entirely possible and even expected!), the time series of the data analyzed here unfortunately don't extend back long enough to document it. There is still plenty of discussion material without this point.

Thank you for your feedback. We have revised the text to address your concerns. Specifically, we have softened the statement about "global stability prior to 3.6 Ma" to avoid overinterpretation (lines 234-239).

18. Paragraph 205: The events postulated here may all have occurred. We cannot know for certain that this is exactly how history played out, however. Please add in some qualifiers to indicate that this *might* have been what happened. For example: "ocean cooling *could have* enhanced water column stratification….cooling also *might have* contributed to the decrease….as a result, more organic material *would have* reached a mesopelagic zone".

We have modified the text to include qualifiers, as recommended, to reflect the uncertainty surrounding the proposed events (Lines 275-278).

19. Line 214: As commented in point #11, this study doesn't estimate species' niche breadth or position. From the data presented, it isn't possible to distinguish the effect of changes in niche occupancy from changes in the spatial extent (vertical and/or latitudinal) of habitats representing

the original (potentially unchanged) niche. A mechanism-agnostic reframe of this sentence would be, “An elevated nutrient availability might have enabled subthermocline populations to grow”.

We have modified the text as suggested (lines 278-281).

20. Line 215: As commented in point #14, the exact onset of the change in ecogroup and/or species diversity is hard to pin down as perfectly corresponding to the iNHG – in Fig 4, it looks like it could have started even has far back as 2.8 or 2.9 Ma in the Northern Atlantic. Rather than declaring that “the intensification of the Northern Hemisphere Glaciation catalyzed a reconfiguration of latitudinal diversity and equitability gradients,” perhaps the authors would accept a softening to something like, “there is a possible mechanistic pathway for the iNHC to have catalyzed a reconfiguration…”.

We modified the text as suggested (lines 282ff).

21. Line 248: There is only one prior point in the manuscript with any mention of results for the Indian Ocean (line 129). I know there are few data for the northern hemisphere component, because it has so little geographic area, but it would be nice to give the (southern) Indian Ocean more comparable page space as the Pacific throughout the article.

We have made an effort to incorporate more information about the Indian Ocean where appropriate in the manuscript. However, the Indian Ocean basin, particularly its Northern Hemisphere component, has limited data coverage and does not exhibit significant patterns comparable to those observed in the Atlantic or Pacific basins. As such, the scope for discussion is naturally more constrained. Nonetheless, we have aimed to ensure that the Indian Ocean is represented as fully as possible given the available data (see lines 175-177; 264-269; 274-277; 318-321).

Methods

22. Line 489: Define kyr at first use for non-geology/paleontology readers

We have revised the text as suggested by the Reviewer (line 93).

23. State how many species were studied

We now include the number of studied species (line 340) and include a table with all the studied species in the supplementary material (Supplementary Table 1-2).

24. State how many time bins there were

We have revised the text as suggested by the Reviewer (line 345).

25. Line 509: Is sampling coverage quantified as Good’s u? That seems to be what is described, but it needs to be specified if so – or if not, how else coverage is calculated

Sampling coverage in the iNext package is indeed based on Good's U estimator. We have clarified this in the text and provided additional details as needed (lines 366-368).

26. Is the ESI weighted by ecogroup diversity? It sounds like it is calculated by the relative, raw occurrence counts of the different ecogroups, but since there are many more species in some groups than others my naïve guess would be that changes in depauperate ecogroups (mainly the polar one) would under-influence ESI. Or maybe the point of null models is to mitigate this for results interpretation? This gets back to needing more text (perhaps even a whole paragraph) of thorough explanation about what bipartite analysis is and how exactly ESI is computed.

The ESI is calculated using the relative occurrence counts of ecogroups within each paleolatitudinal band, ensuring that the contributions of depauperate ecogroups, such as the polar group, are proportionate to their representation. While larger ecogroups inherently have more influence due to higher occurrence counts, the use of relative frequencies mitigates this bias. Additionally, null models are employed to account for expected distributions under random conditions accounting for statistical variations and sampling errors, allowing us to interpret deviations with more statistical rigor. We have expanded the manuscript to include a detailed explanation of the bipartite analysis, the calculation of ESI, and the role of null models in addressing any potential biases (see lines 371ff).

27. Line 525-528: Can you reword the explanation of the null model? I am especially confused by the wording, "taking into account the number of occurrences within specific latitudinal bands across all groups and within group temporal bin across all latitudes." I know the journal may impose strict word limits, but adding at least another sentence or even a whole paragraph would be worth it to make this very clear to visualize and follow. For instance, how is the model iterated? Nothing in the one-sentence overview of the model is described as probabilistic, but it must be for the model to differ between runs. The figures of null model results (supplemental figs 10-12) aren't mentioned in the main text, so I'm also unclear how they feed into or are compared against the main analysis.

We have now reworded the explanation of the null model to increase clarity and explicitly describe the probabilistic nature of the model, including how iterations are performed. The null models explicitly show whether the observed occurrences may happen by chance alone or not, given the occurrence data across the spatiotemporal bins. These models also inform us about the extent and the direction of deviation from random chance, and this ensures a secondary validation of the ESI values we calculate from the observed occurrence values (see lines 390-401). We have also now mentioned the supplemental figures 10-12, which present a summary form of the results from our null model analyses, in the main text.

28. Line 536: Explain what "corrected sampled-in-bin" metric is and give a citation
The 'corrected sampled-in-bin' metric is used to account for sampling biases when estimating the presence of taxa within a specific time bin. It adjusts the observed data to better reflect true ecological or evolutionary signals by correcting for variations in sampling effort across different intervals. This approach is particularly useful in paleontological and ecological studies to avoid over- or under-representing taxa due to uneven sampling. For further details, please refer to Alroy (2010), who provides a comprehensive methodology for implementing these corrections:

Alroy, J. (2010). Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. Paleobiology, 36(2), 227-243.
We added the explanation of the "corrected sampled-in-bin" metric to the methods and provided a citation as suggested (see lines 406-410).

29. Line 538: What is the smallest sample coverage estimated from the empirical samples and how does it compare to the value of 0.7 chosen as quorum value? The quorum shouldn't exceed twice the minimum coverage, and personally I prefer not to extrapolate at all (i.e., to set the quorum at or below the minimum attained by empirical samples). Please report the coverage and revise it if appropriate.

The sample coverage for species diversity in our dataset ranges from 0.89 to 1.00, while the coverage for ecogroup diversity is consistently 1.00. Based on your feedback, we reviewed our chosen quorum value. Initially, we had set a quorum of 0.7; however, we actually increased this to 0.8 during the analysis to better align with the observed sample coverage. For further details on our sample coverage, please see the supplementary figures (Supplementary Fig.13-14 and 17).

Major methods comments

30. Line 538: I am confused by the explanation of why to use 100 kyr time bins here when the study is conducted at 150 kyr resolution. What do you mean it's because of "the possibility" of running one analysis at a finer resolution? In practice, from what I understand, both taxonomic and functional diversity analyses were run at the same resolution of 150 kyr. Unless there is a simple explanation I am missing, I would ask the authors to repeat this at 150 kyr to match the analyses. If the analyses are run at different resolutions (e.g., one at 100 kyr and one at 150 kyr), then I would ask to repeat the one at 100 kyr again at 150 kyr, to be able to evaluate results on the same temporal scale.

To avoid confusion and ensure consistency, we now conduct all analyses, including taxonomic and functional diversity, at the same resolution of 150 kyr. We appreciate your suggestion to improve the clarity and rigor of our methodology.

31. Line 542: SQS has assumptions and limitations like any method, which need to be addressed here. I know that SQS is the go-to diversity estimation method in many quantitative paleobiology studies, most of which poorly justify this choice, and I also know that there's a widespread lack of understanding/lack of teaching about exactly how SQS works under the hood. My goal isn't to talk the authors out of SQS, but to provide some more information to help them make this methods choice more robust. To begin, SQS is a type of rarefaction (it goes by Coverage-Based Rarefaction or CBR in ecology literature), so it's incorrect to say it "offers a significant advancement over traditional subsampling approaches like rarefaction." Yes, CBR/SQS can lead to more accurate estimates than classical count-based rarefaction under certain conditions (cite both Chao and Jost 2012 and Alroy 2010, Science). CBR/SQS can also produce wildly *less* accurate estimates if the data fail to meet its assumptions. One of the more challenging assumptions is that evenness is constant across time/samples. Unfortunately, there's no single gold standard way to estimate evenness. Here, then, are my suggestions:

o Reword the Species and Ecogroup Diversity subsection to explain more clearly how SQS works. The supplement to Close et al. 2018 in MEE, “How should we estimate diversity in the fossil record,” is an accessible place to read about this besides Alroy’s original paper.

Thank you for the suggestion. We have revised the text as suggested (see lines 403ff).

o Add error bars to the diversity estimates in Figure 4. Or, if this makes the figure too cluttered, make it a separate supplemental figure that gets its own page to be a larger size. It’s easy to ask for, say, a 90% confidence interval from iNEXT output

Thank you for the suggestion. We have added 90% confidence intervals to the diversity estimates, which are now presented in Supplementary Figure 18-19 for clarity.

o Calculate one or more evenness metrics and plot them as a time series in a supplemental figure, with panels for each ocean basin. If they seem variable, or it’s hard to feel confident in them, repeat the diversity analysis using either raw diversity or classical count-based rarefaction. Make a new version of figure 4 with this different calculation of diversity, and put it in the supplement.

Thank you for your suggestion. We ran Pielou's Evenness (J) as the evenness metric, as it is particularly well-suited for capturing species balance in our dataset, which spans different time bins and hemispheres with variable richness and the presence of rare species. The results are plotted as time series, with panels for each ocean basin, and can be seen in Supplementary Figure 15-16 . Our sampling coverage is quite high, as well as the calculated evenness, so we have opted not to repeat the diversity analysis using count-based rarefaction.

Recurring comment throughout the text:
32. Please check and revise the way in-line citations are formatted to match journal guidance. I noticed through the article that for multi-author papers, the format is “Jones et al. found…” except when one of the authors of the present study is the second author. Then, the format changes to “Jones and Johnson et al. found…”. Only the first author should be named (or however journal formatting rules dictate).

The citation format for Woodhouse and Swain et al., 2023 is intentional, as both authors are listed as co-first authors. This format follows the journal's guidelines for acknowledging co-first authorship. We have ensured that all other in-line citations are formatted according to the journal's rules. If further clarification is needed, we are happy to make additional adjustments.

33. Cite version numbers for all individual software packages, e.g., for bipartite, divDyn, iNEXT, etc.

Version numbers of all individual software packages were included in the manuscript.

Figures
34. Fig 2: Please separate the panels with white space

We separated the panels in Figure 2 with white space, as suggested.

35. SM fig 1: Please change the color coding. The blue and green dots are difficult for me to discern and I expect will be inaccessible to people with color vision deficit. Coblis is a good tool to check for accessibility https://www.color-blindness.com/coblis-color-blindness-simulator/

Thank you for your suggestion. All figures in the manuscript use color schemes that are color-blind friendly, as defined by ColorBrewer, and were tested in a color blindness simulator. For SM Fig. 1, We adjusted the colors based on your feedback and hope the new scheme improves accessibility. We appreciate your input to ensure the figures are inclusive.

36. SM fig 2: Consider changing the definition of axis ticks from every 0.15 Ma to 0.1 or 0.2. Consider (optional) reversing the plot order of ecogroups 1 to 5, so group 1 is on bottom.

We have updated the axis ticks to intervals of 0.2 Ma as recommended and reversed the plot order of Ecogroups 1 to 5, placing Group 1 at the bottom. We appreciate your input in improving the figure's readability.

37. SM fig 3-7: Define iNHG, mPWP, and cCAS in captions. Add confidence interval shading around the linear fit lines so the viewer can tell whether they are significantly different from a zero-slope/horizontal line. I realize that might add visual clutter, in which case you could remove the lines connecting points and just leave the points (not necessary, just a suggestion to try in case it's helpful).

The terms iNHG, mPWP, and cCAS have been defined in SM Fig. 2 caption. We also added 90% confidence interval shading around the linear fit lines to help viewers assess whether they are significantly different from a zero-slope/horizontal line. See supplementary figures 18-19. We appreciate your feedback in improving the clarity of these figures.

38. All plots: Convention differs by scientific field, but I often recommend authors consider flipping age axes so time moves from left (oldest) to right (youngest). It's up to the authors, just make sure to update the slope definitions on line 563 if changing this.

We understand the convention of flipping age axes in some scientific fields; however, we prefer to retain the current orientation of our axes, as it aligns with the conventions commonly used in Neogene vs Phanerozoic plots. We believe this approach maintains consistency and clarity for our intended audience. We appreciate your understanding.

Response to Reviewers

Point-by-point responses to the reviewer comments are provided below in italics and in green font. All line numbers in our responses refer to the updated submission.

Reviewer #1:

I appreciate your careful and thorough revision. Figures are now much better and I like them very much.
Thank you for your positive feedback! We are glad the revised figures are clearer and appreciate your suggestions that helped improve them.

Only a few things:
Line 76 “linked to biodiversity hotspots of important marine groups, including sharks”
Better to say “pelagic group” and just “biodiversity (instead of biodiversity hotspot)”, as coastal benthic groups have very different biodiversity trend/hotspot and also pelagic groups don’t show very distinct “hotpots”.

We edited the sentence for further clarity as suggested (see line 76).

From line 88: “Although comprehensive studies of planktic foraminiferal diversity over various temporal scales, ranging from thousands e.g.,34,37-38 to millions e.g.,3,39 of years, exist; to date, none have investigated patterns on large spatial scales with hundreds of sites.”
“none have (should be “has”?) investigated patterns on large spatial scales with hundreds of sites”
This is bit misleading as Swain et al Woodhouse et al and Fenton et al actually did such using Triton.
You may add a few more words to better specify (such as intermediate time scale or regionally resolved? I suspect somehow you missed to specify what you like to say here enough).

We edited the sentence for further clarity as suggested (see lines 88-90).

The paragraph from Line 100 on Triton:
You may consider to combine it into the previous paragraph, as it looks a bit redundant.
We combined the text to avoid redundancy, as suggested (see lines 97ff).

Otherwise I think the manuscript is now well revised.

Sincerely yours,
Moriaki Yasuhara

Reviewer #2:

Helen Coxall review round 2.

This is a much improved manuscript. My major comments have been addressed more or less completely – i.e. the authors have given attention to the details of the planktonic foraminifera assemblages and eco-groupings at the heart of this study, connecting us back to the data and observations.
Thank you for your positive feedback. We're pleased to hear that you found the manuscript much improved and that the revisions helped strengthen the connection between the data, assemblages, and eco-groupings central to the study.

I am largely happy, however, the new list of species raises a couple of foram specific questions that I would like to see addressed before this should be accepted for publication.

• Figure 1. A representative diagram of the planktic foraminifera ecogroups used in this
695 study.
Great. A vast improvement (even if the Neogene species do not have such nice silhouettes!)
Thank you!

• Figure 5. Conceptual representation of eco-group distributions during warmer (mPWP = the mid-Pliocene Warm Period
This is also a great addition and complement to Fig 1 bringing the paper together. Really helpful.
Thank you!

• Supplementary Table 1. Species of planktic foraminifera with assigned eco-groups used in our study from Aze, compared to ecogroups assigned by modern studies.
Thank you for making this table. Even though the eco-group interpretations have not much changed since Aze et al., its important and useful for the modern planktonic foraminifera expert to keep these check lists of species alive, and keep the references up to date. It also highlights the level of taxonomic detail that is contained in the data base, which is important.
-Please in the SI Table caption include a list of the eco-groups (again) to remind the reader what the numbers mean. Or perhaps since this is all about the ecogroups its worth while **sorting them by eco group** so we see the composition of each group.

We've updated the captions for Supplementary Tables 1 and 2 to include the list of ecogroups and clarify what the numeric codes represent. Additionally, we have sorted the species by ecogroup in Supplementary Table 1 and 2 to highlight the composition of each ecogroup, as suggested.

• Now I see the lists of species I have a question about the North Atlantic. **Why is the species *Neogloboquadrina atlantica* not listed in the table or eco-groups?** It is extremely abundant/dominant in the North Atlantic in the Pliocene going extinct in the early Pleistocene – it is also an important biostratigraphic marker in the North Atlantic/subpolar region. It seems to be the key high latitude North Atlantic species before *N. pachyderma* took over that niche. If it is not in the Triton database then it must have been overlooked and binned in e.g. *N. dutertrei*. But is this true? It is important comment on this.

This is correct, and the general lack of N. atlantica in Triton is due to several compounded effects:

1. The Aze phylogeny doesn't recognize N. atlantica as a species, which is rectified in the upcoming Lamyman phylogeny (of which Woodhouse is a co-author) and we deemed it to be important enough to include in this work - referring back to the drilling reports where N. atlantica has been recognized,

2. Due to the nature of Triton being based almost exclusively on scientific ocean drilling efforts, as you say, it has often been overlooked or misidentified amongst the other Neogloboquadrinids,

3. Although N. atlantica is included in the Triton database, we evaluated all temporal bins for data completeness and applied a sample coverage threshold of >0.7 using the iNEXT package as well as records count of over 15 in each temporal bin. It happens that N. atlantica was primarily present in bins that did not meet this completeness criterion and were therefore excluded from the final analysis. As a result, N. atlantica did not appear in the diversity and ecogroup calculations, in order to maintain consistency across taxa and temporal bins.

Co-author Woodhouse is actively updating Triton and seeking to populate it with more detailed studies beyond the scientific ocean drilling reports to include more records of species such as N. atlantica.

The data base and paper focus on macroperforate planktic foraminifera (single-celled calcareous zooplankton), the reason given that "Planktic foraminifera with microperforate and medioperforate wall textures were excluded from the analysis since they are currently characterized by a less refined phylogeny compared to macroperforate forms." That may be the case, **although there are some species that become important at high latitudes eg. *Globigerinita uvula/glutinata*.** Morad et al., 2019, hints of hidden diversity among some of these small tax such as these. **-Can the authors at last comment on the extent to which the omission of microperforates might result in significant information missing in terms of diversity signals?**

Great point, and we agree that there is vital information contained within the microperforate record, which needs to be explored as well. Other than the less well resolved phylogeny, the other key factor that required their exclusion from the analyses is the comparably less complete ecogroup data amongst the microperforate species. In the new Lamyman phylogeny, 43/81 *microperforate species have an assigned ecogroup, whilst 377/379 macroperforate species do (See lines 337-341). Clearly, we have a lot of work to do!*

Morard, R., Vollmar, N.M., Greco, M., Kucera, M., 2019. Unassigned diversity of planktonic foraminifera from environmental sequencing revealed as known but neglected species. PLOS ONE 14, e0213936.

This microperforate issue also becomes relevant for this section:

Line 192-: "The negative correlation emerged concurrently with the onset of the intensification of the Northern Hemisphere Glaciation at ~2.7 Ma, and ecogroup 5 played a significant role in driving the observed ecological changes during the Pliocene - early Pleistocene interval (Supplementary Fig. 8). Curiously, the population of ecogroup 5 is represented by only two species — *Neogloboquadrina pachyderma* and *Turborotalita quinqueloba* — in contrast to other ecogroups (See supplementary table 1). Yet, despite its lower diversity, ecogroup 5 plays a significant role in driving changes observed during the Pliocene - early Pleistocene interval."

This is where inclusion of microperforate taxa would make a difference to the diversity (+ 1 or 2 species). Also the inclusion of *N. atlantica*.

Another issue to follow up those from the comment above, is that of the 265 occurrences of N. atlantica in our study interval, only 9 actually occur higher than 66.6 degrees of latitude, so it would not fit into the criteria for a high latitude species. Interestingly, this draws attention to the actual criteria of defining these forms based solely on biogeography, where the high latitudes have a disadvantage in being notoriously undersampled, especially in deeper geological time.

Important N. atlantica references

Weitkamp, T.M., Razmjooei, M.J., Pearson, P.N., Coxall, H.K., 2025. Upper Oligocene to Pleistocene planktonic foraminifera stratigraphy at North Atlantic DSDP Site 407, Reykjanes Ridge: diversity trends and biozonation using modern Neogene taxonomic concepts. J. Micropalaeontol. 44, 1-78.

Bolton, C.T., Bailey, I., Friedrich, O., Tachikawa, K., de Garidel-Thoron, T., Vidal, L., Sonzogni, C., Marino, G., Rohling, E.J., Robinson, M.M., Ermini, M., Koch, M., Cooper, M.J., Wilson, P.A., 2018. North Atlantic Midlatitude Surface-Circulation Changes Through the Plio-Pleistocene Intensification of Northern Hemisphere Glaciation. Paleoceanography and Paleoclimatology 33, 1186-1205.

Dowsett, H., Robinson, M., Foley, K., Herbert, T., Hunter, S., Andersson, C., Spivey, W., 2023.

The Relative Stability of Planktic Foraminifer Thermal Preferences over the Past 3 Million Years. Geosciences 13, 71.

Flower, B.P., 1999. Data report: Planktonic foraminifers from the subpolar North Atlantic and Nordic Seas: Sites 980-987 and 907. Proceedings of the Ocean Drilling Program: Scientific Results 162, 19-34.

Weaver, P.P.E., Clement, B.M., 1986. Synchroneity of Pliocene planktonic foraminiferal datums in the North Atlantic. Marine Micropaleontology 10, 295-307.

Reviewer #3:

I commend the author team on an improved draft of a manuscript that already had many merits. Particularly big and helpful additions in this draft are the redesign of figures; supplemental tables with species information; and sampling-standardization methods and their explanations. I also found the text overall to be clearer and stronger.
Thank you for your positive feedback!

The revision addressed most of my original concerns, although there are a few outstanding comments and a few new issues I have flagged below. There are 10 suggestions in total, most of which are minor wording changes, but a few of which are more involved and/or more influential. For instance, in this draft I found an error in a figure and an error in describing a figure. The clearer methods text and data visualizations in this draft also meant I was also able to evaluate some pieces of the work I found too unclear to comment on before (e.g. the null models), so I have follow-on comments about the details of those methods/results where last time I could only ask for clarification on them.

1. Line 25 and 73: Zebrafish and hydras are marine model organisms; planktic foraminifera are not. We lack knowledge about many parts of their basic natural history, like their lifespans. Contrast this to a true model organism like *C. elegans*, where we've literally counted and studied every individual cell in the animal! Please just call forams marine organisms or such, not model organisms.
We have removed the references to planktic foraminifera as model organisms in Lines 25 and 73.

2. Line 89: I don't understand the claim of studies of planktic foram diversity, "to date, none have investigated patterns on large spatial scales with hundreds of sites". Four of the five papers cited in the sentence study planktic forams on global scales, some with the same dataset – how is that not an equally large scale? I see from the reviewer response document that the authors came up with this language after I asked for changes in this paragraph, but unfortunately, the new language doesn't solve the original issue for me and introduces new ones.
Please find a way to introduce the justification/positioning of this study that avoids claims to be the first to address an issue or to "uniquely address" an issue. Many journals explicitly dissuade

this language, as claims to primacy are notoriously difficult to prove, and unnecessary. I repeat my original suggestion below. (The authors don't have to take this suggestion, but they do need to find a wording that follows the recommendations above.)
I understand the typical narrative structure of introductions is to include a sentence like this ("despite some prior study, there is still a gap in understanding X"), but I suggest omitting the sentence here….Simply merge some the language from the sentence with the next two sentences for flow in the paragraph. For example, something like,
"The recently collated Triton dataset, which provides an extensive record of global fossil occurrences of planktic foraminifera, allows for comprehensive, high-resolution studies of zooplankton community responses to paleoclimate change across regional scales."
We understand the concern about overstating the novelty of our study and have revised the sentence to remove any claims of being the first or uniquely addressing a specific issue (See lines 88ff).

3. Line 133: I am excited to see the results of the resampled data in SM fig 9. Please state what the smallest occurrence count was that you matched in all time bins. Is this a count of cores? That would make the most sense to me as a way to address the issue of spatially inconsistent sampling that I think multiple reviewers may have pointed out, but I suspect it may be individual foram records that's being counted instead. What would results look like to resample distribution data using cores/sites as the unit of distribution?

Thank you for this insightful comment. You are correct that in the original version, the smallest occurrence count referred to the number of individual foraminiferal records. To test for the influence of sampling strategy, we repeated the resampling analysis using the number of unique cores (rather than individual foraminiferal occurrences) as the matched unit within each time bin. The resulting ecogroup proportions were identical to those from the original approach, indicating that the observed patterns are robust to the resampling method. We have, therefore, slightly adjusted the wording in the manuscript to clarify that the number of cores was used for resampling and have also added the smallest number of cores included in the analysis to the figure heading (See Supplementary figure 9). See lines 133, 434.

4. Line 139: Start a new paragraph with "After verifying…"
We started a new paragraph with "After verifying…". See line 138.

5. Please can you find a substitute for the word equitable in defining ESI in line 142? I commented on this word being unclear to me in the first draft and the revision still has this issue. I worry it will be opaque to ESL readers. Wouldn't "evenness" or "balance" serve just as well here? Outside of this, the revised explanations in the methods help a lot.
We have substituted the word "equitable" with "evenness" where applicable. (Lines 138,141-143, 168, 176).

6. Line 195: Why is it curious that ecogroup 5 is represented by only two species (N. pachyderma and T. quinqueloba)? Ecogroup 5 is high-latitude species, i.e. polar species, and it's typical of nearly every organismal group that there are many fewer polar species than temperate or equatorial ones. I would leave out the adverb "curiously" – just state that ecogroup 5 is

composed of these two species, which seems straightforward. Also omit the next sentence, beginning line 197, which is verbatim duplication from the end of preceding one.
We removed the word 'curiously" as suggested (see line 196).

7. Line 337: This line gives in-text citations as Aze et al.(21) and Fenton & Woodhouse et al.(22). The review response letter said that second authors were named in places where first authorship was joint. However, I looked up Fenton et al. (2021), and nowhere does the article mention joint first authorship. Indeed, Fenton is the single corresponding author as well as first author. I strongly recommend making the back-to-back citations here consistent, or it gives an arrogant impression to the author whose inclusion is a formatting exception.
Thank you for bringing this issue to our attention. Co-author Woodhouse had been under the impression since Triton was published that he was co-first author as was agreed with the rest of the writing team, however this is clearly not the case. Woodhouse has contacted Fenton to get this amended as soon as possible. We will therefore maintain this formatting style.

8. With regards to the paragraph beginning line 220:
A). Line 225: "The South Atlantic exhibited an earlier increase in ecogroup diversity, preceding the North Atlantic by ~150 kyrs." This isn't apparent to me from Figure 4: The South Atlantic ecogroup diversity curve looks close to flat, and it's a subjective call as to when exactly it may have begun to increase. I would leave this sentence out and let viewers draw their own interpretation from the figure.
The sentence was removed as suggested.

B). Line 226: "In the Indian Ocean, ecogroup diversity remained relatively stable in the Southern Hemisphere, while the Northern Hemisphere displayed a gradual increase that started during the closure of the Central American Seaway." This sentence is backwards to what I see in the plot – the *Northern* hemisphere Indian ocean ecogroup diversity curve looks flat, while the *Southern* diversity increased around the same early time as the North Pacific.
Thank you for catching this mismatch between the text and the figure. We've corrected the sentence to accurately reflect the patterns shown in the plot. See lines 228-230.

C. Line 228: "shifts…occur over a broad interval spanning the mPWP and the intensification of the Northern Hemisphere Glaciation." By the authors' own description in this paragraph, the shifts also "started during the closure of the Central American Seaway," so this sentence should say shifts spanned the iNHG, mPWP, *and* cCAS.
See also the comment below about this paragraph and the figures it describes.
Thank you for your suggestion. We've added the closure of the Central American Seaway to the sentence as suggested. See line 231-232.

9. Thank you for adding confidence intervals to the diversity curves (new supplemental figures 18 and 19)! I was really curious to look at this. There are two new concerns I have upon being able to see the fuller data picture now, however. First, there is an error in one of the figures: panel 18a is a duplicate of panel 19a. Please update this and check what happened in the code to generate it.

Second, all the ecogroup diversity curves look indistinguishable when plotted with their confidence intervals. I think this makes sense when we consider that the plot/SQS estimation is comparing diversity values of 1 vs 2 vs 3. There's just not much difference between 2 and 3 – they're both small numbers, if that makes sense. In that light, it feels a little silly to do SQS on a dataset with a maximum diversity of 74 (species) and assess differences between that and SQS on a dataset with a maximum diversity of only 5 (ecogroups). Here is my suggestion: fine to leave the figures as-is (with a fix to panel 18a), but add a sentence to the main text that acknowledges there is broad overlap in confidence intervals for ecogroup diversity – that is, it's hard to claim statistical support for differences within time series or between time series (SM fig 19).

Thank you for pointing this out. We apologize for the duplication error in panel 18a; the final figures were prepared in Adobe Illustrator, so the issue was not related to the R code. We have corrected panel 18a accordingly. Regarding the confidence intervals, we agree that the overlap across ecogroup diversity curves makes it difficult to draw statistically supported conclusions from time-series comparisons. We have added a sentence to the main text acknowledging this, as suggested. See lines 233-236.

Just to note, in general, the ecogroup diversity curves for the Northern and Southern Hemispheres show substantial overlap in the Atlantic and Pacific oceans. In contrast, the hemispheric diversity trajectories in the Indian Ocean are more distinct, with limited overlap in both the curves and their associated confidence intervals.

10. Thank you for the additional methods text and figures to explain the null models. I preface this comment by saying it's possible I am completely misunderstanding the current explanations of the null models. If that is the case, then the authors should revise to prevent such misunderstanding.

I am looking at SM figures 10-12, which show the proportion of ecogroups per time bin that have stronger vs. no difference vs. weaker preference than random assortment, by latitudinal bin. The result, strongly and consistently across all three figures (corresponding to the three ocean basins studied), is that ~ALL ecogroups in all time bins and all latitudes show *no preference* compared to their respective null models. Given the stated purpose of the null models to "determine how observed patterns deviate from random expectation," the fact the output seems to indicate there is *not* any deviation from random seems like it should undermine confidence in the main results. Either that or the null models are irrelevant to the main analyses, in which case it feels like their inclusion is a red herring meant to give an illusion of methodological rigor. Please help me understand what I'm missing, or put clear caveats in the main text about the limits of detectability of the stated patterns, given most follow a random distribution exactly.

(Maybe my misunderstanding is rooted in the fact it's undefined what the preference is *for*. Preference for what? It can't be species' preference for ecogroups, because the phrasing is *ecogroup preference for X*. Ecogroup preference for ocean/longitude?)

We apologize for any confusion regarding the null models and we certainly did not include them to give an illusion of methodological rigor. The results from the null models do not undermine confidence in the main results as they represent something different from the main analyses and give us further confidence in our main results.

The reasoning behind using the null models is to see if there are certain ecogroups which have differential affinity to certain latitudinal bins/biogeographic areas (i.e., higher or lower chance of a finding certain ecogroup being present in a certain latitudinal bin than expected by random chance). As most of them do not (as shown by the results), these results systematically allow us to show that the changes in the relative composition of ecological communities are a consequence, direct or indirect, of temporal/environmental changes, rather than being strict biogeographical/regional contraints/patterns (and therefore, we call them null models). This is highly important as it emphasizes the 'adaptability' of foraminiferal species and how their response to external changes can be studied. See lines 396ff.

There are more results that can be derived from the null model analyses which are not directly related to the results of the current manuscript and we hope to explore them in a future project.

Review of “Impact of Climate Variability on Planktic Foraminifera”

The authors have analyzed taxonomic and functional diversity of a well-resolved plankton group across an interval of Earth history relevant to present-day climate warming. I was interested to read this article from its premise alone, and I would be happy to see an article on this topic appear in Nature Communications. My comments are mostly minor (although see two larger methodological comments/questions), and most relate to improving clarity, especially of the methodological details, or softening the interpretive assertions to be less definitive. I apologize to the authors for my long delay in returning this review and hope the detail of constructive comments below helps offset the inconvenience.

Abstract and Introduction

1. Line 21 and 42: I suggest rewording “recent observations” to something like “observations of living species”. “Recent observations” could mean samples of living species taken in the last few years or, geologically speaking, samples from the Recent.
2. Line 47: Do you mean “life system responses”? I wouldn’t consider plankton part of the Earth system. You could also be wordier but explicit, “life system responses to Earth system change” or similar.
3. Line 48: Define Ma at first use for non-geology/paleontology readers
4. Line 50: This is referring to the rise of the Panama isthmus, correct? I would include reference to that by name, since I think it’s a more well-known geographic feature than the Central American Seaway. Also consider additional references – the single citation for the given age date was published several years ago now and there’s been heated debate about the exact timing of the closure.
5. Line 80: I understand the typical narrative structure of introductions is to include a sentence like this (“despite some prior study, there is still a gap in understanding X”), but I suggest omitting the sentence here beginning “Despite ongoing efforts…”. There have arguably been more comprehensive and high-resolution comparative studies of zooplankton responses to paleoclimate change than for any other group except bivalves. That doesn’t make the topic less deserving of study! Simply merge some the language from the sentence with the next two sentences for flow in the paragraph. For example, something like,

   > “The recently collated Triton dataset, which provides an extensive record of global fossil occurrences of planktic foraminifera, allows for comprehensive, high-resolution studies of zooplankton community responses to paleoclimate change across regional scales.”

Results

6. The results section needs to include a several-sentence introduction to what bipartite network analysis is and how it calculates ESI, including a definition of equitability that doesn’t include the word equitable. I know there is prior literature applying this method to similar questions, but this paper needs to be fully understandable on its own for a reader who may be unfamiliar with that previous work and with network theory. How would you explain this to an advanced undergraduate learning about it for the first time?

7. The information currently in the Methods about how to interpret ESI (ranging from 0 to 1, with 1 indicating specialization) needs to be repeated here as ESI results are presented. Readers need this context to interpret Fig 2, for example.
8. Line 126: I would say "might be" the result of thermocline differences rather than "likely" due to it. There are alternative hypotheses, and reality is usually a mix of processes at work. Without data on the thermocline or an estimate of its effect relative to other processes, it feels too assertive to me to assume the pattern is due to thermocline differences alone.
9. Line 128: Reword to avoid the words "subtropical" and "tropical". The location of the tropics has changed through time, so it is better to discuss the location of biogeographic patterns relative to static latitudinal markers.
10. Line 128: Say "shift" rather than "noticeable shift". I agree it's present, but it's a slight and qualitative signal rather than jumping off the page. Also remove the word "significant" from the following sentence unless it is associated with a statistical test result.
11. Line 136: This study calculates relative occurrence frequency, not species' niche breadth or position. Changes in relative occurrences could be due to niche occupancy changes but could also result from changes in the spatial extent of species' habitats or other factors. Please reword to avoid the phrase "ecological niche expansion" here. Similarly, please remove the phrase "reconfiguration of ecological niches" in the caption to fig 3.
12. Line 150: Environment is presumably stable along both vertical and latitudinal axes, right?
13. Line 167: The in-text results in the sentence beginning "The highest rate of change…" don't appear match what I see in Fig 4. It's unclear that ecogroup diversity did drop after cCAS in the North Atlantic, and I certainly wouldn't call it a "sharp decline". Rather, the green line for the N Atlantic barely drops (not significant) in the first three time bins proceeding from the cCAS, and it then drops at the onset of the mPWP (Fig 4B). The "highest rate of change in ecogroup diversity" instead looks like it might be in the North Pacific around 3.4 Ma.
14. Line 171: As in comment #13, the text description here appears to differ from what I see in the figures. Ecogroup diversity does not consistently increase in the Northern Pacific (second panel of Fig 4B) as the sentence beginning on this line asserts ("a consistent increase in ecogroup diversity is evidence across all basins in the Northern Hemisphere"). I am also unconvinced that the changes happened exactly at the iNHG, in either fig 3 or 4, as stated in the captions there and in line 172. The onset of a directional trend could be argued to originate anytime between the mPWP and iNHG from figs 3 and 4, and it is ultimately a qualitative call rather than a quantitative break-point estimate.
15. Consider discussing limitations of unequal sample size between hemispheres and basins

Discussion

16. Line 177: I would suggest an additional reference here as in the Introduction, as this timing has been a contentious topic.
17. Sentence beginning 182: I'm unable to follow the reasoning of this sentence. The evidence cited in the preceding sentence seems to be from around the closure at 2.8 Ma, but also implies this study found matching changes beginning at 3.6. Those timings don't match up perfectly. In addition, the "global stability…prior to 3.6 Ma" mentioned here

isn't previously mentioned in the Results and is based on only 2 data points (between 3.6 and 3.9 Ma). I agree that the closing of the CAS was a major biogeographic event, and it is worth mentioning in the Discussion. That said, looking at the figures, e.g., Fig 3 and 4, I struggle to see any features of the data around then that couldn't be attributable to white-noise variation (no rapid directional changes between 3.5 and 3.9 Ma). I suggest walking back the strength of the claims in this paragraph. If there were major shifts from before vs. after the cCAS (entirely possible and even expected!), the time series of the data analyzed here unfortunately don't extend back long enough to document it. There is still plenty of discussion material without this point.

18. Paragraph 205: The events postulated here may all have occurred. We cannot know for certain that this is exactly how history played out, however. Please add in some qualifiers to indicate that this *might* have been what happened. For example: "ocean cooling *could have* enhanced water column stratification….cooling also *might have* contributed to the decrease….as a result, more organic material *would have* reached a mesopelagic zone".
19. Line 214: As commented in point #11, this study doesn't estimate species' niche breadth or position. From the data presented, it isn't possible to distinguish the effect of changes in niche occupancy from changes in the spatial extent (vertical and/or latitudinal) of habitats representing the original (potentially unchanged) niche. A mechanism-agnostic reframe of this sentence would be, "An elevated nutrient availability might have enabled subthermocline populations to grow".
20. Line 215: As commented in point #14, the exact onset of the change in ecogroup and/or species diversity is hard to pin down as perfectly corresponding to the iNHG – in Fig 4, it looks like it could have started even has far back as 2.8 or 2.9 Ma in the Northern Atlantic. Rather than declaring that "the intensification of the Northern Hemisphere Glaciation catalyzed a reconfiguration of latitudinal diversity and equitability gradients," perhaps the authors would accept a softening to something like, "there is a possible mechanistic pathway for the iNHC to have catalyzed a reconfiguration…".
21. Line 248: There is only one prior point in the manuscript with any mention of results for the Indian Ocean (line 129). I know there are few data for the northern hemisphere component, because it has so little geographic area, but it would be nice to give the (southern) Indian Ocean more comparable page space as the Pacific throughout the article.

Methods

22. Line 489: Define kyr at first use for non-geology/paleontology readers
23. State how many species were studied
24. State how many time bins there were
25. Line 509: Is sampling coverage quantified as Good's u? That seems to be what is described, but it needs to be specified if so – or if not, how else coverage is calculated
26. Is the ESI weighted by ecogroup diversity? It sounds like it is calculated by the relative, raw occurrence counts of the different ecogroups, but since there are many more species in some groups than others my naïve guess would be that changes in depauperate ecogroups (mainly the polar one) would under-influence ESI. Or maybe the point of null models is to mitigate this for results interpretation? This gets back to needing more text (perhaps even a whole paragraph) of thorough explanation about what bipartite analysis is and how exactly ESI is computed.

27. Line 525-528: Can you reword the explanation of the null model? I am especially confused by the wording, “taking into account the number of occurrences within specific latitudinal bands across all groups and within group temporal bin across all latitudes.” I know the journal may impose strict word limits, but adding at least another sentence or even a whole paragraph would be worth it to make this very clear to visualize and follow. For instance, how is the model iterated? Nothing in the one-sentence overview of the model is described as probabilistic, but it must be for the model to differ between runs. The figures of null model results (supplemental figs 10-12) aren’t mentioned in the main text, so I’m also unclear how they feed into or are compared against the main analysis.
28. Line 536: Explain what “corrected sample-in-bin” metric is and give a citation
29. Line 538: What is the smallest sample coverage estimated from the empirical samples and how does it compare to the value of 0.7 chosen as quorum value? The quorum shouldn’t exceed twice the minimum coverage, and personally I prefer not to extrapolate at all (i.e., to set the quorum at or below the minimum attained by empirical samples). Please report the coverage and revise it if appropriate.

Major methods comments

30. Line 538: I am confused by the explanation of why to use 100 kyr time bins here when the study is conducted at 150 kyr resolution. What do you mean it’s because of “the possibility” of running one analysis at a finer resolution? In practice, from what I understand, both taxonomic and functional diversity analyses were run at the same resolution of 150 kyr. Unless there is a simple explanation I am missing, I would ask the authors to repeat this at 150 kyr to match the analyses. If the analyses are run at different resolutions (e.g., one at 100 kyr and one at 150 kyr), then I would ask to repeat the one at 100 kyr again at 150 kyr, to be able to evaluate results on the same temporal scale.
31. Line 542: SQS has assumptions and limitations like any method, which need to be addressed here. I know that SQS is the go-to diversity estimation method in many quantitative paleobiology studies, most of which poorly justify this choice, and I also know that there’s a widespread lack of understanding/lack of teaching about exactly how SQS works under the hood. My goal isn’t to talk the authors out of SQS, but to provide some more information to help them make this methods choice more robust. To begin, SQS is a type of rarefaction (it goes by Coverage-Based Rarefaction or CBR in ecology literature), so it's incorrect to say it “offers a significant advancement over traditional subsampling approaches like rarefaction.” Yes, CBR/SQS can lead to more accurate estimates than classical count-based rarefaction under certain conditions (cite both Chao and Jost 2012 and Alroy 2010, Science). CBR/SQS can also produce wildly *less* accurate estimates if the data fail to meet its assumptions. One of the more challenging assumptions is that evenness is constant across time/samples. Unfortunately, there’s no single gold standard way to estimate evenness. Here, then, are my suggestions:
    - Reword the Species and Ecogroup Diversity subsection to explain more clearly how SQS works. The supplement to Close et al. 2018 in MEE, “How should we estimate diversity in the fossil record,” is an accessible place to read about this besides Alroy’s original paper.
    - Add error bars to the diversity estimates in Figure 4. Or, if this makes the figure too cluttered, make it a separate supplemental figure that gets its own page to be a larger size. It’s easy to ask for, say, a 90% confidence interval from iNEXT output

- Calculate one or more evenness metrics and plot them as a time series in a supplemental figure, with panels for each ocean basin. If they seem variable, or it's hard to feel confident in them, repeat the diversity analysis using either raw diversity or classical count-based rarefaction. Make a new version of figure 4 with this different calculation of diversity, and put it in the supplement.

Recurring comment throughout the text:

32. Please check and revise the way in-line citations are formatted to match journal guidance. I noticed through the article that for multi-author papers, the format is "Jones et al. found…" except when one of the authors of the present study is the second author. Then, the format changes to "Jones and Johnson et al. found…". Only the first author should be named (or however journal formatting rules dictate).
33. Cite version numbers for all individual software packages, e.g., for bipartite, divDyn, iNEXT, etc.

Figures

34. Fig 2: Please separate the panels with white space
35. SM fig 1: Please change the color coding. The blue and green dots are difficult for me to discern and I expect will be inaccessible to people with color vision deficit. Coblis is a good tool to check for accessibility https://www.color-blindness.com/coblis-color-blindness-simulator/
36. SM fig 2: Consider changing the definition of axis ticks from every 0.15 Ma to 0.1 or 0.2. Consider (optional) reversing the plot order of ecogroups 1 to 5, so group 1 is on bottom.
37. SM fig 3-7: Define iNHG, mPWP, and cCAS in captions. Add confidence interval shading around the linear fit lines so the viewer can tell whether they are significantly different from a zero-slope/horizontal line. I realize that might add visual clutter, in which case you could remove the lines connecting points and just leave the points (not necessary, just a suggestion to try in case it's helpful).
38. All plots: Convention differs by scientific field, but I often recommend authors consider flipping age axes so time moves from left (oldest) to right (youngest). It's up to the authors, just make sure to update the slope definitions on line 563 if changing this.

I commend the author team on an improved draft of a manuscript that already had many merits. Particularly big and helpful additions in this draft are the redesign of figures; supplemental tables with species information; and sampling-standardization methods and their explanations. I also found the text overall to be clearer and stronger.

The revision addressed most of my original concerns, although there are a few outstanding comments and a few new issues I have flagged below. There are 10 suggestions in total, most of which are minor wording changes, but a few of which are more involved and/or more influential. For instance, in this draft I found an error in a figure and an error in describing a figure. The clearer methods text and data visualizations in this draft also meant I was also able to evaluate some pieces of the work I found too unclear to comment on before (e.g. the null models), so I have follow-on comments about the details of those methods/results where last time I could only ask for clarification on them.

1. Line 25 and 73: Zebrafish and hydras are marine model organisms; planktic foraminifera are not. We lack knowledge about many parts of their basic natural history, like their lifespans. Contrast this to a true model organism like C. elegans, where we've literally counted and studied every individual cell in the animal! Please just call forams marine organisms or such, not model organisms.

2. Line 89: I don't understand the claim of studies of planktic foram diversity, "to date, none have investigated patterns on large spatial scales with hundreds of sites". Four of the five papers cited in the sentence study planktic forams on global scales, some with the same dataset – how is that not an equally large scale? I see from the reviewer response document that the authors came up with this language after I asked for changes in this paragraph, but unfortunately, the new language doesn't solve the original issue for me and introduces new ones.

Please find a way to introduce the justification/positioning of this study that avoids claims to be the first to address an issue or to "uniquely address" an issue. Many journals explicitly dissuade this language, as claims to primacy are notoriously difficult to prove, and unnecessary. I repeat my original suggestion below. (The authors don't have to take this suggestion, but they do need to find a wording that follows the recommendations above.)

> I understand the typical narrative structure of introductions is to include a sentence like this ("despite some prior study, there is still a gap in understanding X"), but I suggest omitting the sentence here….Simply merge some the language from the sentence with the next two sentences for flow in the paragraph. For example, something like,
>
> "The recently collated Triton dataset, which provides an extensive record of global fossil occurrences of planktic foraminifera, allows for comprehensive, high-resolution studies of zooplankton community responses to paleoclimate change across regional scales."

3. Line 133: I am excited to see the results of the resampled data in SM fig 9. Please state what the smallest occurrence count was that you matched in all time bins. Is this a count of cores? That would make the most sense to me as a way to address the issue of spatially inconsistent sampling that I think multiple reviewers may have pointed out, but I suspect it may be individual

foram records that's being counted instead. What would results look like to resample distribution data using cores/sites as the unit of distribution?

4. Line 139: Start a new paragraph with "After verifying…"

5. Please can you find a substitute for the word equitable in defining ESI in line 142? I commented on this word being unclear to me in the first draft and the revision still has this issue. I worry it will be opaque to ESL readers. Wouldn't "evenness" or "balance" serve just as well here? Outside of this, the revised explanations in the methods help a lot.

6. Line 195: Why is it curious that ecogroup 5 is represented by only two species (N. pachyderma and T. quinqueloba)? Ecogroup 5 is high-latitude species, i.e. polar species, and it's typical of nearly every organismal group that there are many fewer polar species than temperate or equatorial ones. I would leave out the adverb "curiously" – just state that ecogroup 5 is composed of these two species, which seems straightforward. Also omit the next sentence, beginning line 197, which is verbatim duplication from the end of preceding one.

7. Line 337: This line gives in-text citations as Aze et al.(21) and Fenton & Woodhouse et al.(22). The review response letter said that second authors were named in places where first authorship was joint. However, I looked up Fenton et al. (2021), and nowhere does the article mention joint first authorship. Indeed, Fenton is the single corresponding author as well as first author. I strongly recommend making the back-to-back citations here consistent, or it gives an arrogant impression to the author whose inclusion is a formatting exception.

8. With regards to the paragraph beginning line 220:

A). Line 225: "The South Atlantic exhibited an earlier increase in ecogroup diversity, preceding the North Atlantic by ~150 kyrs." This isn't apparent to me from Figure 4: The South Atlantic ecogroup diversity curve looks close to flat, and it's a subjective call as to when exactly it may have begun to increase. I would leave this sentence out and let viewers draw their own interpretation from the figure.

B). Line 226: "In the Indian Ocean, ecogroup diversity remained relatively stable in the Southern Hemisphere, while the Northern Hemisphere displayed a gradual increase that started during the closure of the Central American Seaway." This sentence is backwards to what I see in the plot – the *Northern* hemisphere Indian ocean ecogroup diversity curve looks flat, while the *Southern* diversity increased around the same early time as the North Pacific.

C. Line 228: "shifts…occur over a broad interval spanning the mPWP and the intensification of the Northern Hemisphere Glaciation." By the authors' own description in this paragraph, the shifts also "started during the closure of the Central American Seaway," so this sentence should say shifts spanned the iNHG, mPWP, *and* cCAS.

See also the comment below about this paragraph and the figures it describes.

9. Thank you for adding confidence intervals to the diversity curves (new supplemental figures 18 and 19)! I was really curious to look at this. There are two new concerns I have upon being

able to see the fuller data picture now, however. First, there is an error in one of the figures: panel 18a is a duplicate of panel 19a. Please update this and check what happened in the code to generate it.

Second, all the ecogroup diversity curves look indistinguishable when plotted with their confidence intervals. I think this makes sense when we consider that the plot/SQS estimation is comparing diversity values of 1 vs 2 vs 3. There's just not much difference between 2 and 3 – they're both small numbers, if that makes sense. In that light, it feels a little silly to do SQS on a dataset with a maximum diversity of 74 (species) and assess differences between that and SQS on a dataset with a maximum diversity of only 5 (ecogroups). Here is my suggestion: fine to leave the figures as-is (with a fix to panel 18a), but add a sentence to the main text that acknowledges there is broad overlap in confidence intervals for ecogroup diversity – that is, it's hard to claim statistical support for differences within time series or between time series (SM fig 19).

10. Thank you for the additional methods text and figures to explain the null models. I preface this comment by saying it's possible I am completely misunderstanding the current explanations of the null models. If that is the case, then the authors should revise to prevent such misunderstanding.

I am looking at SM figures 10-12, which show the proportion of ecogroups per time bin that have stronger vs. no difference vs. weaker preference than random assortment, by latitudinal bin. The result, strongly and consistently across all three figures (corresponding to the three ocean basins studied), is that ~ALL ecogroups in all time bins and all latitudes show *no preference* compared to their respective null models. Given the stated purpose of the null models to "determine how observed patterns deviate from random expectation," the fact the output seems to indicate there is *not* any deviation from random seems like it should undermine confidence in the main results. Either that or the null models are irrelevant to the main analyses, in which case it feels like their inclusion is a red herring meant to give an illusion of methodological rigor. Please help me understand what I'm missing, or put clear caveats in the main text about the limits of detectability of the stated patterns, given most follow a random distribution exactly.

(Maybe my misunderstanding is rooted in the fact it's undefined what the preference is *for*. Preference for what? It can't be species' preference for ecogroups, because the phrasing is *ecogroup preference for X*. Ecogroup preference for ocean/longitude?)