nature portfolio

Peer Review File

Microeukaryote metabolism across the western North Atlantic Ocean revealed through autonomous underwater profiling

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Cohen et al.'s manuscript "Microeukaryote metabolism across the western North Atlantic Ocean revealed through autonomous underwater profiling" is an impressive work highlighting not only the exciting opportunities associated with the Clio sampling platform, but the power of using a multi-omics approach to examine metabolism. Overall, this manuscript is in great shape, but there are a few issues that are critical to address before publication.

The noteworthy results are the insights related to both surface and deeper ocean microeukaryote metabolism that are linked to macro- and micro-nutrient concentration data. This manuscript is a giant leap forward in terms of the vast quantity of multi-omic data presented. It's also one of very few manuscripts that has coordinated multi-omic and micronutrient concentration data and none that I am aware of have done this at this resolution and scale. For the most part, the work supports the conclusions and claims (see comments of main issue below), the methodology is sound, and the work meets the expected standards in the field. Also, the methods have the appropriate level of detail.

My main issue with the manuscript is in the presentation of the prokaryotic transcriptome data, which is not the main crux of the manuscript, but because it is right up front, it detracts from the rest of the manuscript. Here's my issue with this section, these metatranscriptomes were generated using poly-A selection. The fact that prokaryotic transcripts were identified is not surprising, as prokaryotes have been found to polyadenylate, however the role of poly-A in prokaryotes is not 100% known. What is known is that the role of polyadenylation in prokaryotes is not the same as what occurs in eukaryotes. The tails are shorter, for one. It seems to be more about marking RNA for degradation, not stabilization. That said, the research in that area is not fully developed, it's not clear that all prokaryotes polyadenylate and it's not clear that it's always done for marking degradation. Since this isn't really the point of the article, I would make it a supplemental discussion (if it's discussed at all). That supplemental discussion would have to dive further into the literature about polyadenylation in prokaryotes. Given that this doesn't

add much to the story and would involve an in-depth literature review, it is maybe less worthwhile to include. I would move figure 2 into the supplemental (or at least 2A) – certainly it shouldn't be the first figure showing multi-omic data in the manuscript.

My other comments on the manuscript are generally positive. The authors aren't overstating the significance of their dataset, which as I said earlier, is very impressive. I think it's exciting that they're seeing consistent patterns of N and P stress genes. I found the discussion of the mismatch in the protein/transcript pools especially insightful. I do note that there appears to be a word missing in the 5th from last sentence on page 9 (starting "For example NRT proteins"). I also think that the caveats associated with the quantification should be highlighted in the main manuscript (the timing of the addition of standards).

I do have several comments regarding figures and figure legends that I feel should be addressed as well:

• In all the relative abundance figures (figs. 2, 3, etc.), "site" should be replaced by "station." For what it's worth, I like that the separate groups are included as supplemental figures – I feel that those are 100% worth including.
• Figure 2 legend: the legend for "C" doesn't explain what the figure is showing. Also, the last sentence belongs (is) in the paper not in the figure legend and it should be "relatively" not "relaticely."
• Figure 4: This is a really busy figure. What are the "groups" based on? Prior knowledge? Add something about that in the legend.
• The figure legend for supp. 19 is lacking. Need to explain that they are the modules and what A vs. B are showing. I assume C is the correlation between patterns observed in each module and the env. variables? The figure 5 legend is much clearer.
• The layout for figure 6 is a bit odd with the 2nd ferritin making a 3rd column.

Reviewer #2 (Remarks to the Author):

This study analyzes the distribution of protists horizontally and vertically over a transect in the North Atlantic Ocean, spanning gradients of oligotrophic, continental margin, and

productive coastal ecosystems. The sampling is performed using an Autonomous Underwater Vehicle which allows to obtain a finer granularity in the sampling with in situ filtration, particularly useful for collecting microbial biomass.

The main novelty is the direct analysis of the proteins from the whole community, and the coupling with the transcripts expressed, obtaining a crucial dataset for the understanding of the protist community. One of the main hurdles of any omic approach is that we obtain just a specific point of view with inherent particularities of the technique used, and the combinations with novel approaches create a new window, gaining a more robust picture. Following the metaphor of the blind scientists in the room, we are starting to see the elephant, and studies like this one pushes this challenging endeavor. That is, if we do not get completely paranoid/crazy about the discordances between methods.

The introduction is direct and very well written. The Results and Discussion section are also very well detailed considering the amount of information that the study is trying to unpack (metatranscriptomic, protein spectrums, pigments and some images).

Given the amount of back to back in the taxonomic composition, concordance between transcript/protein and the quantitative estimates of biomass sections, it could be interesting including a schematic figure of the main processes behind these differences. I am imagining a table with a description of the effect, the references supporting it, and the taxons affected. Even if many of them are hypotheses, it would make the whole comparison more navigable. In the present form is already a good summary, but the inclusion of this would make it top-notch. Although maybe the authors are already considering performing a separate review on the topic. In any case, either a reordering of the text, a table summary or something in this fashion would be useful for the reader to better understand the contrasts. Overall I consider that this manuscript is valid to be published in Nature Communications and congratulate the authors in the effort to make it summarized and thorough at the same time.

General considerations:

-The discussion of the concordance between transcripts and proteins is difficult to keep in the head given that there are multiple situations with different explanations. This is expected given the amount of factors to take into account for a whole community discussion over long gradients with a huge variation of environmental conditions. I would propose to rewrite this section. One of your conclusions is that the fact of having a secondary perspective is helping in the process of 'seeing the elephant'. Before entering specific comparisons, maybe I would start with your expectation of not seeing that many similarities (L321 to 333) and discuss similarities first, and afterwards discussing the cases of discordance. I would also try to include a summary table of the discrepancies and your train of thought.

Small considerations:

-In Figure 2 I would split the bacterial supergroups from the protists in the legend of A and B.

- L246 to 252. If I understood it correctly, in this section the authors have summed all the TPMs over all the transect and have plotted the linear relationship. I worry that the approach is too reductionist, with data overly summarized. A more convenient approach would be to calculate the mean of the TPMs for a specific module, even if you threw away all the samples that present a 0 (in the sum approach you are doing it too). That is, the points in Supplementary figure 14 would be the mean TPMs over all the transect vs the means of the protein spectral points.
Figure 4 could highlight some of the discussions you are arising in the text. Although it is correct as it is, it does not help that much to understand the patterns.

- Figure 5 has too much information. Sticking only to the examples highlighted with a stroke and removing the rest (keeping it in the supplementary) would be a better approach since right now it's not possible to differentiate between functions.

- Figure 7 is an amazing way to convey the results. I would however change the coloring to give information regarding the ecological distribution of the species. Each color could be a

distribution, even if they are approximate: surface, deep, DCM, coastal, generalists, etc. In the present form it is difficult to discern the main message of the figure, how these groups are different in its biogeography.

Reviewer #3 (Remarks to the Author):

This is a well-written and interesting paper in which multiple methods have been used to characterize microeukaryotic ecology. I recommend this paper for publication with suggestions below that would be further beneficial to the manuscript. The main comment is that the utility of and confidence in the observing approach would be greatly increased if specific examples of discordance could be tested against the various theories for such.

Note: Detailed review would have been facilitated by inclusion of page numbers.

Pg 9: “where cells were investing in” nitrogen uptake: consider rephrasing to remove anthropomorphic tone. “driven” by water column depth; consider rephrasing; “driven” is too strong of a term for the data provided.
Pg. 13: “whereas the transcript data selected against prokaryotes…” directly contradicts page 4 that says “unexpectedly… belonged to prokaryotes…” This should be rectified.
Pg 8: difference in size fraction is mentioned for omics versus UVP but it is not mentioned that there is also a difference between omics and proteins.
Pg. 11. “unexpected difference in the proportion…” Is 21% vs 38% statistically different?

Overall, the manuscript does a good job describing the different reasons for concordance and not. Nonetheless, by the time you get down to Fig. 6 and other places in which discordance is clearly shown and the issue is no longer theoretical, the reader is left with a feeling of dissatisfaction with regard to knowing whether or not the expense of a multi-omic approach is recommended or useful. There are a couple of places that the manuscript tries to “underscore the utility…” of the approach (pg 20; pg 12), but that utility is not clearly demonstrated. In fact, pg. 13 says that data could suggest that limited ability of omics-based approaches to reflect bulk changes in biomass. One note on biomass: if the authors feel like proteins are a true gold standard for biomass, please cite. Also curious about the correlation

between proteins and visual estimates of biomass.

Acknowledging that I do not have an easy remedy to suggest, I nonetheless challenge the authors to see if they can address this issue better in some way. Although there is a good story about nitrogen and phosphorous stress biomarkers; it feels like the story could be mined because it is consistent with what is already (thought to be) known. The question is – did this dataset teach us anything new? Was there any information provided by only one technique? If you could only pick one – which would it be? (In global ocean observing, observing system simulation experiments (OSSEs) are performed to ask the question – can we sample less?).

With these thoughts in mind, might there be a few specific examples in which the authors could drill down into the data to test the different theories about mechanisms of concordance and discordance? For example: Can the PIT data in Figure 6 be explained by transcript versus protein degradation rates? Bias against transcript sequencing? Known database discrepancies? The utility of and confidence in the paper would be greatly increased with a few examples in which the various possible explanations for match/not match could be demonstrated by information independent from the correlations themselves.

Reviewer #1 (Remarks to the Author):

Cohen et al.'s manuscript "Microeukaryote metabolism across the western North Atlantic Ocean revealed through autonomous underwater profiling" is an impressive work highlighting not only the exciting opportunities associated with the Clio sampling platform, but the power of using a multi-omics approach to examine metabolism. Overall, this manuscript is in great shape, but there are a few issues that are critical to address before publication.

The noteworthy results are the insights related to both surface and deeper ocean microeukaryote metabolism that are linked to macro- and micro-nutrient concentration data. This manuscript is a giant leap forward in terms of the vast quantity of multi-omic data presented. It's also one of very few manuscripts that has coordinated multi-omic and micronutrient concentration data and none that I am aware of have done this at this resolution and scale. For the most part, the work supports the conclusions and claims (see comments of main issue below), the methodology is sound, and the work meets the expected standards in the field. Also, the methods have the appropriate level of detail.

My main issue with the manuscript is in the presentation of the prokaryotic transcriptome data, which is not the main crux of the manuscript, but because it is right up front, it detracts from the rest of the manuscript. Here's my issue with this section, these metatranscriptomes were generated using poly-A selection. The fact that prokaryotic transcripts were identified is not surprising, as prokaryotes have been found to polyadenylate, however the role of poly-A in prokaryotes is not 100% known. What is known is that the role of polyadenylation in prokaryotes is not the same as what occurs in eukaryotes. The tails are shorter, for one. It seems to be more about marking RNA for degradation, not stabilization. That said, the research in that area is not fully developed, it's not clear that all prokaryotes polyadenylate and it's not clear that it's always done for marking degradation. Since this isn't really the point of the article, I would make it a supplemental discussion (if it's discussed at all). That supplemental discussion would have to dive further into the literature about polyadenylation in prokaryotes. Given that this doesn't add much to the story and would involve an in-depth literature review, it is maybe less worthwhile to include. I would move figure 2 into the supplemental (or at least 2A) – certainly it shouldn't be the first figure showing multi-omic data in the manuscript.

We thank Reviewer #1 for their insights on the utility of this work and for the constructive feedback on the analysis. Since the prokaryotic findings were distracting from the microeukaryote-focused angle of this work, we have moved "whole microbial community" meta-omic community breakdown to the supplementary section (previously Fig. 2) as suggested.

Polyadenylation of prokaryotic mRNA is an interesting aspect to consider when interpreting our metatranscriptomes. It is possible our metatranscriptomic assembly is only reflective of bacterial groups with mRNA containing poly-A tails related to degradation. We agree this is not the focus of this manuscript, but this contextualization is important and we have added a few sentences to the main manuscript to clarify this point.

Lines 123-135 now read:
Transcripts were assembled using sequences derived from poly-A-selected mRNA, which enriches for eukaryotic members of the community. Approximately 23% of the annotated open reading frames (ORFs) belonged to prokaryotes (Supplementary Fig. 2C), indicating that bacterial mRNA was not excluded in the library preparation process. Up to 50% of bacterial mRNA can undergo polyadenylation, and this process may be associated with a variety of functions including mRNA degradation, stabilization, or translation, in contrast to the eukaryotes which polyadenylate most of their mRNA primarily to carry out translation 30,31. This prokaryotic poly-A selected mRNA in this analysis may therefore not be reflective of the true prokaryotic community, and was instead used as a qualitative assessment of prokaryotes along the transect and depth zones. The poly-A based metatranscriptome assembly enabled identification of bacteria in the metaproteomes (Supplementary Fig. 2B). Eukaryotes were not enriched during protein processing as they were during mRNA processing, and the metaproteomes therefore more accurately reflect the natural proportion of prokaryotes relative to eukaryotes in the water column (Supplementary Fig. 2B).”

My other comments on the manuscript are generally positive. The authors aren’t overstating the significance of their dataset, which as I said earlier, is very impressive. I think it’s exciting that they’re seeing consistent patterns of N and P stress genes. I found the discussion of the mismatch in the protein/transcript pools especially insightful. I do note that there appears to be a word missing in the 5th from last sentence on page 9 (starting “For example NRT proteins”). I also think that the caveats associated with the quantification should be highlighted in the main manuscript (the timing of the addition of standards).

We are glad to hear the manuscript adequately conveyed the main findings. The important detail about RNA standards being added after extractions were performed has been added to the main text.

Lines 479-481 now read:
“Note that internal standards were added after RNA extractions and prior to sequencing. As a result, this estimate assumes efficient yield of RNA during extractions and are likely high estimates (see methods).”

I do have several comments regarding figures and figure legends that I feel should be addressed as well:

• In all the relative abundance figures (figs. 2, 3, etc.), “site” should be replaced by “station.” For what it’s worth, I like that the separate groups are included as supplemental figures – I feel that those are 100% worth including.

“Site” has been replaced by “station” in all taxonomy barplots.

• Figure 2 legend: the legend for “C” doesn’t explain what the figure is showing. Also, the last sentence belongs (is) in the paper not in the figure legend and it should be “relatively” not “relaticely.”

This has been clarified and the figure caption now reads:
“(c) The relative abundance of taxonomically annotated eukaryotic ORFs contained in the translated metatranscriptome database is displayed as a single stacked barplot.”

• Figure 4: This is a really busy figure. What are the “groups” based on? Prior knowledge? Add something about that in the legend.
• The figure legend for supp. 19 is lacking. Need to explain that they are the modules and what A vs. B are showing. I assume C is the correlation between patterns observed in each module and the env. variables? The figure 5 legend is much clearer.
• The layout for figure 6 is a bit odd with the 2nd ferritin making a 3rd column.

Suggested edits to figure legends and captions have been incorporated. The second panel in what was previously Fig. 6 has been moved to the supplemental section.

Reviewer #2 (Remarks to the Author):

This study analyzes the distribution of protists horizontally and vertically over a transect in the North Atlantic Ocean, spanning gradients of oligotrophic, continental margin, and productive coastal ecosystems. The sampling is performed using an Autonomous Underwater Vehicle which allows to obtain a finer granularity in the sampling with in situ filtration, particularly useful for collecting microbial biomass.

The main novelty is the direct analysis of the proteins from the whole community, and the coupling with the transcripts expressed, obtaining a crucial dataset for the understanding of the protist community. One of the main hurdles of any omic approach is that we obtain just a specific point of view with inherent particularities of the technique used, and the combinations with novel approaches create a new window, gaining a more robust picture. Following the metaphor of the blind scientists in the room, we are starting to see the elephant, and studies like this one pushes this challenging endeavor. That is, if we do not get completely paranoid/crazy about the discordances between methods.

The introduction is direct and very well written. The Results and Discussion section are also very well detailed considering the amount of information that the study is trying to unpack (metatranscriptomic, protein spectrums, pigments and some images).

Given the amount of back to back in the taxonomic composition, concordance between transcript/protein and the quantitative estimates of biomass sections, it could be interesting including a schematic figure of the main processes behind these differences. I am imagining a table with a description of the effect, the references supporting it, and the taxons affected. Even if many of them are hypotheses, it would make the whole comparison more navigable. In the

present form is already a good summary, but the inclusion of this would make it top-notch. Although maybe the authors are already considering performing a separate review on the topic. In any case, either a reordering of the text, a table summary or something in this fashion would be useful for the reader to better understand the contrasts.
Overall I consider that this manuscript is valid to be published in Nature Communications and congratulate the authors in the effort to make it summarized and thorough at the same time.

We are grateful for this thoughtful feedback, and for the recommendations to improve clarity of the transcript/protein concordance discussion. We have included a supplemental table (Supplementary Table 2) to outline the biological, technical, and bioinformatic factors influencing transcript and protein agreement in marine meta-omic studies. A second supplemental table is also now included documenting specific examples of disagreement between proteins and transcripts found in this study and the organisms affected (Supplementary Table 3).

General considerations:

-The discussion of the concordance between transcripts and proteins is difficult to keep in the head given that there are multiple situations with different explanations. This is expected given the amount of factors to take into account for a whole community discussion over long gradients with a huge variation of environmental conditions. I would propose to rewrite this section. One of your conclusions is that the fact of having a secondary perspective is helping in the process of 'seeing the elephant'. Before entering specific comparisons, maybe I would start with your expectation of not seeing that many similarities (L321 to 333) and discuss similarities first, and afterwards discussing the cases of discordance. I would also try to include a summary table of the discrepancies and your train of thought.

We thank the reviewer for this feedback, and have reorganized the "*Concordance between transcript and protein pools*" section (lines 236-372) for clarity as recommended, beginning with a description of expectations, evidence for agreement, disagreement, and ending with potential reasons for the disagreement. We have additionally expanded on why disagreement may have occurred using a few case examples, as suggested by Reviewer #3.

Small considerations:

-In Figure 2 I would split the bacterial supergroups from the protists in the legend of A and B.

This figure has been moved to the supplemental section, and prokaryotes/eukaryotes have been separated in the legend (now Supplementary Fig. 2).

- L246 to 252. If I understood it correctly, in this section the authors have summed all the TPMs over all the transect and have plotted the linear relationship. I worry that the approach is too reductionist, with data overly summarized. A more convenient approach would be to calculate the mean of the TPMs for a specific module, even if you threw away all the samples that present

a 0 (in the sum approach you are doing it too). That is, the points in Supplementary figure 14 would be the mean TPMs over all the transect vs the means of the protein spectral points. Figure 4 could highlight some of the discussions you are arising in the text. Although it is correct as it is, it does not help that much to understand the patterns.

We have redone this portion of the analysis with these suggestions in mind. As before, we need a way to consolidate count information from across sites/depths in order to perform the correlation. We still sum normalized TPM (transcript) and NSAF (protein) counts associated with ORFs across sites/depths to estimate total relative abundance (Supplementary Fig. 15A), but now we average counts among ORFs of the same KEGG ID to compare average KO counts between transcripts and proteins (Supplementary Fig. 15B). In contrast to the previous analysis where we only summed counts of the same KEGG ID (Supplementary Fig. 15D), there was no linear relationship between transcripts and proteins. We interpret this to mean that functional processes only appear congruent when aggregated across the ORFs, in which relatively abundant processes become apparent in both pools. Averaged counts among ORFs additionally may not be representative of protein levels since mass spectrometry data is zero-inflated. (Note that we sum counts across ORFs in our other analyses to generate an aggregated view of functional processes in these microbial communities [Fig. 4-6]). This exercise served as a useful reference point for us to examine specific cases of functional discordance, which are now described in the text. In particular, we found Rubisco and tubulin to show strong agreement, but other cytoskeletal processes such as desmin strongly reflected in the proteins, and ribosomal components highly expressed in the transcripts. Possible reasons for these are discussed in the text on lines 334-354.

- Figure 5 has too much information. Sticking only to the examples highlighted with a stroke and removing the rest (keeping it in the supplementary) would be a better approach since right now it's not possible to differentiate between functions.

We acknowledge this is a busy figure. Making most of the modules transparent was done to help the reader follow the major trends between surface and deep communities. We want to avoid making them even more transparent since that impedes interpretability of these other abundant modules. Instead, we have removed most KEGG BRITE functional annotations that were not relevant. (The full legend is still shown in Supplementary Fig. 20). We have also added asterisks to bring the reader's attention to 6 example modules highlighted that were correlated with surface and deep zones.

- Figure 7 is an amazing way to convey the results. I would however change the coloring to give information regarding the ecological distribution of the species. Each color could be a distribution, even if they are approximate: surface, deep, DCM, coastal, generalists, etc. In the present form it is difficult to discern the main message of the figure, how these groups are different in its biogeography.

We have updated the background coloring to differentiate between coastal vs offshore, and surface, mesopelagic, and bathypelagic zonations. An important message is that some

organisms change in distribution (e.g., forams in the mesopelagic and fungi in the bathypelagic), but others are ubiquitous, such as the diatoms and dinoflagellates. We have left the organisms themselves color-coded so this pattern can be identified.

Reviewer #3 (Remarks to the Author):

This is a well-written and interesting paper in which multiple methods have been used to characterize microeukaryotic ecology. I recommend this paper for publication with suggestions below that would be further beneficial to the manuscript. The main comment is that the utility of and confidence in the observing approach would be greatly increased if specific examples of discordance could be tested against the various theories for such.

We thank this reviewer for their helpful suggestions and constructive feedback regarding our presentation of protein and transcript concordance.

Note: Detailed review would have been facilitated by inclusion of page numbers.

We apologize for the omission of page numbers in the previous draft. They have been added to the revised manuscript.

Pg 9: “where cells were investing in” nitrogen uptake: consider rephrasing to remove anthropomorphic tone. “driven” by water column depth; consider rephrasing; “driven” is too strong of a term for the data provided.

These sentences have been modified as suggested. Lines 266-271 now read:

“Negative correlative relationships were also identified, and but were largely related to environmental distributions of where metabolic processes were occurring rather than intracellular biochemistry. For example, NRT proteins were detected in surface waters where nitrogen uptake to fuel photosynthetic growth occurs, and negatively correlated to ferredoxin and Cu/Zn SOD transcripts in deep water where they presumably are used in metal-dependent heterotrophic metabolism.”

Pg. 13: “whereas the transcript data selected against prokaryotes…” directly contradicts page 4 that says “unexpectedly… belonged to prokaryotes…” This should be rectified.

We have rephrased to clarify. Lines 389-393 now read:

“Observed disconnects among these measurements may be partially explained by protein and pigment pools reflecting abundant prokaryotes such as cyanobacteria (Supplementary Fig. 2B, Supplementary Fig. 12), whereas the transcript data enriched for eukaryotes by use of poly-A transcript selection (Supplementary Fig. 2A).”

Pg 8: difference in size fraction is mentioned for omics versus UVP but it is not mentioned that there is also a difference between omics and proteins.

Both transcripts and proteins in this analysis are derived from filtered cells that undergo extractions in the lab, although it is possible that soluble molecules from lysed cells are lost on the 0.2 µm filter during the *in situ* filtration process. Because this likely affects both molecular pools, it isn't mentioned in this section. However the important point regarding loss of soluble proteins and transcripts has been incorporated into Supplementary Table 2, which outlines factors contributing to discordance between transcripts and proteins in marine microeukaryotes.

Pg. 11. "unexpected difference in the proportion…" Is 21% vs 38% statistically different?

We performed a Student's t-test, and this difference was determined to be significant. Lines 315-318 now read:

"On average, 21 ± 6% of annotated metatranscriptomic communities (n=44) were associated with ORFs unable to be classified at the supergroup level, while an average of 38 ± 5% of metaproteomic communities (n=86) were associated with these lineage-conflicted ORFs ($p = 1.01 \times 10\text{-}34$, Student's t-test, t=17.03, degrees of freedom=128; Fig. 3: "Not Annotated (NA)")."

Overall, the manuscript does a good job describing the different reasons for concordance and not. Nonetheless, by the time you get down to Fig. 6 and other places in which discordance is clearly shown and the issue is no longer theoretical, the reader is left with a feeling of dissatisfaction with regard to knowing whether or not the expense of a multi-omic approach is recommended or useful. There are a couple of places that the manuscript tries to "underscore the utility…" of the approach (pg 20; pg 12), but that utility is not clearly demonstrated. In fact, pg. 13 says that data could suggest that limited ability of omics-based approaches to reflect bulk changes in biomass. One note on biomass: if the authors feel like proteins are a true gold standard for biomass, please cite. Also curious about the correlation between proteins and visual estimates of biomass.

Acknowledging that I do not have an easy remedy to suggest, I nonetheless challenge the authors to see if they can address this issue better in some way. Although there is a good story about nitrogen and phosphorous stress biomarkers; it feels like the story could be mined because it is consistent with what is already (thought to be) known. The question is – did this dataset teach us anything new? Was there any information provided by only one technique? If you could only pick one – which would it be? (In global ocean observing, observing system simulation experiments (OSSEs) are performed to ask the question – can we sample less?).

We appreciate the encouragement to directly make recommendations to the oceanographic community. We now incorporate a summarizing statement under the "*Future Directions*" section to emphasize that the methods are complementary, and in this analysis, both provided unique perspectives in the North Atlantic microbial community. Although both omics tools were valuable here, we are not recommending to leave out traditional lower cost approaches such as

microscopy, imaging, flow cytometry, and pigments, because this information is essential for ground truthing omics based relative community composition which have their own biases (Supplementary Table 2).

Lines 581-589 now read:
“This analysis demonstrated that both metatranscriptomics and metaproteomics are advantageous in answering ecological questions, providing complementary views into microbial community dynamics. Metatranscriptomics enabled a more in-depth examination with greater molecule coverage and clear view of nutrient status shifts along the transect (Fig. 5), but with metaproteomics offering insight into shifts into microbial communities with depth (Fig. 2) and more closely tracking with changes in community biomass (Supplementary Fig. 14, Supplementary Fig. 18). We recommend multi-omic approaches be used in future oceanographic surveys to the extent possible, given the power of complementary approaches to account for biases specific to one method (Supplementary Table 2).”

Proteomics derived community composition can be used to infer bacterial community biomass (Kleiner 2019; Kleiner et al. 2017), and in this analysis, extracted proteins correlated with total pigments (Supplementary Fig. 18), supporting that they can be used to assess changes in biomass. We do not have cell counts however in this analysis, and there would be tremendous value in equating protein copies/L using targeted quantitation methods (Saito et al. 2020) to imagery data to determine if marine metaproteomics can be used to track changes in specific taxonomic group biomass.

Kleiner, Manuel. 2019. “Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities.” *mSystems* 4 (3). https://doi.org/10.1128/mSystems.00115-19.
Kleiner, Manuel, Erin Thorson, Christine E. Sharp, Xiaoli Dong, Dan Liu, Carmen Li, and Marc Strous. 2017. “Assessing Species Biomass Contributions in Microbial Communities via Metaproteomics.” *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-01544-x.
Saito, Mak A., Matthew R. McIlvin, Dawn M. Moran, Alyson E. Santoro, Chris L. Dupont, Patrick A. Rafter, Jaclyn K. Saunders, et al. 2020. “Abundant Nitrite-Oxidizing Metalloenzymes in the Mesopelagic Zone of the Tropical Pacific Ocean.” *Nature Geoscience* 13 (5): 355–62.

With these thoughts in mind, might there be a few specific examples in which the authors could drill down into the data to test the different theories about mechanisms of concordance and discordance? For example: Can the PIT data in Figure 6 be explained by transcript versus protein degradation rates? Bias against transcript sequencing? Known database discrepancies? The utility of and confidence in the paper would be greatly increased with a few examples in which the various possible explanations for match/not match could be demonstrated by information independent from the correlations themselves.

Yes, and this was a helpful suggestion. We now go into detail regarding why the phosphate transporter spatial distributions do not match, and provide examples of transcripts/proteins that were divergently expressed between pools. This has been incorporated into a revised

"*Concordance between transcript and protein pools*" section to walk the reader through examples of agreement and disagreement, and to provide possible explanations.

Lines 332-354 now read:

"The PLS correlation analysis showed most transcript/protein pairings lacked strong linear relationships, including ferritin, Cu/Zn superoxide dismutase, and the inorganic phosphate transporter (Fig. 3). Phosphate transporter transcripts and proteins have been shown to be correlated in the marine diatom Thalassiosira pseudonana, with both upregulated under phosphate stress 69. The cause of the discordance here is likely related to differences in taxonomic coverage between the molecular pools, with few PHO84 and PiT proteins only detected in stramenopiles and haptophytes, and transcripts detected in a greater assortment of protists (Supplementary Tables 8-9). We further investigated specific functional processes that were highly divergent between molecular pools (Supplementary Fig. 15D), being abundant in the protein fraction yet weakly expressed in transcripts and vice versa. In the proteins, some of the largest relative abundance counts were associated with a protein related to the eukaryotic cytoskeleton (neurofilament medium polypeptide, vimentin and/or desmin), and a cilia associated protein (Bardet-Biedl syndrome 5 protein). In transcripts, high counts were associated with ribosomal proteins (large subunits 6 and 22, small subunit 2), and a calcium/calmodulin-dependent serine/threonine protein kinase. These molecules may have large inventories with more stable expression compared to dynamically regulated processes such as NRT 61. It is difficult to determine with certainty which sole factor or combination of processes influenced the patterns found here. It is possible cytoskeletal proteins are stabilized through post-translational modifications or by incorporation into a cytoskeletal interlinked filament lattice 70, and therefore enriched in the protein fraction. A weak negative relationship in mitochondrial ribosomal proteins and transcripts has also been documented in human lymphocytes 71, potentially related to ribosomal mRNA being more stable than ribosomal proteins 71,72. Thus, differences in mRNA/protein stability and inventory may both contribute to the cause of disconnect between transcripts and proteins detected here."

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

I have reviewed the revised version of Cohen et al.'s manuscript "Microeukaryote metabolism across the western North Atlantic Ocean revealed through autonomous underwater profiling."

As stated in my prior review, this is impressive work that highlights not only the exciting opportunities associated with the Clio sampling platform but also the power of using a multi-omics approach to examine metabolism. The noteworthy results are the insights related to both surface and deeper ocean microeukaryote metabolism that are linked to macro- and micro-nutrient concentration data. This manuscript is a giant leap forward regarding the vast quantity of multi-omic data presented. It's also one of the very few manuscripts that have coordinated multi-omic and micronutrient concentration data. None I know of have done this at this resolution and scale. The work supports the conclusions and claims, the methodology is sound, and the work meets the expected standards in the field. Also, the methods have the appropriate level of detail for the work to be reproduced. I appreciate the attention the authors paid to addressing reviewer suggestions. All of my concerns have been adequately addressed. While not my suggestion, the rewrite of the compare/contrast section was particularly helpful in improving the flow of the manuscript.

I am intrigued by the potential exploration of the prokaryotic transcriptome pool that may come in a future manuscript. As an aside, after reading the initial manuscript, I spent some time looking into the Clontech protocols and noted that some (but not smart-seq v4) do use random priming and thus would amplify the prokaryotic fraction (though if no rRNA removal step was included in your library prep, this would have certainly led to an overabundance of rRNA reads).

Reviewer #3 (Remarks to the Author):

Concerns have been addressed. thank you.