Response to Reviewer 3 Comments (RC3):

We thank the reviewer for the constructive comments

Cantarero Biogeosciences Notes

RC3-1: Summary:

The authors use a combination of multiple linear regression, random forest algorithms, and classification trees to interrogate how physicochemical environmental factors and phytoplankton community structure drive the bulk intact polar lipid pool.

The authors have combined an impressive array of approaches to tease apart a complex system, and the application of these methods to environmental lipidomics is relatively novel (though this paper makes its difficulties clear). Models like these are an important aspect of how the field is currently thinking about interacting environmental drivers, and, with some work, this paper could shed light on how these variables influence the community lipidome and how the lipidome may influence community structure. The paper was generally well written, though the discussion was overly-long and I often felt like it lacked direction, so I would recommend reducing the amount of speculation in the discussion, especially with regards to unsubstantiable TAG production.

Response RC3-1: We thank the reviewer for their positive and constructive comments that will allow us to revise and improve our manuscript.

RC3-2: These are the areas where I think the paper could be significantly improved:

In the MLR section, the authors performed 1650 statistical tests (10 parameters x 165 IPLs), the results of which drove a significant portion of the discussion, but they did not mention controlling for false discovery rate. Additionally, the authors mentioned many IPL classes showed both positive and negative relationships (e.g. SQDG), confounding deep interpretation and leading to an extremely long discussion with many "possible explanations." Controlling for FDR would likely reduce the number of significant relationships, and may help the authors focus their efforts on more significant and powerful correlations and therefore drivers of IPL abundance.

Due to this lack of clear direction, paragraph and section structure in the discussion was not always clear—by that I mean, frequently, the authors suggest a potential cause of various differential IPL abundances, often with little or no evidence in the data and with a citation of a singly study, and then the idea is dropped after a couple of sentences without a clear tie to an overall narrative. Potential TAG production was often invoked as an explanation, despite no TAG data in the manuscript, and I hesitate to endorse such an overarching conclusion with no evidence to support it.

Response RC3-2: Thank you for this thoughtful and thorough suggestion. We agree that controlling for FDR simplifies the interpretation of these MLRs and we will include this reanalysis in the edited manuscript. As suggested, it does allow us to reduce the discussion in several instances where both positive and negative correlations previously complicated our

interpretations. This provides us an opportunity to streamline the discussion and focus on the most significant linear correlations between IPLs and environmental conditions.

We agree that without having measured TAG concentrations directly, the explanation for IPL distributions in several instances is limited and not conclusive. We will reduce the speculation of TAG production in the discussion and focus more so on the observations made directly in this analysis.

RC3-3: Additionally, there was little to no mention of bacteria, or of IPL abundance as a proportion of particulate organic matter—I am not an expert in the Peruvian Upwelling Zone, so it may be that they are an insignificant proportion of the biomass in this location. However, many of the IPLs in this study are prevalent in bacteria (e.g. PE, PG, DGTS, MGDG), in some cases moreso than in phytoplankton. (Twice, the authors suggest a higher IPL/chl a ratio in the subsurface is indicative of greater IPL contribution to phytoplankton biomass—however, this could be explained by either non-phytoplankton, i.e. bacterial, or non-living/detrital IPL pools.) This inquiry may help the authors ground their interpretations of the community response to environmental stressors. Without this understanding of how phytoplankton IPLs fit into the total organic carbon pool, many of their interpretations are incomplete.

Response RC3-3: Our previous work in the Humboldt Current System shows that the ratio of total IPLs to POC is highest at the peak of chlorophyll, as is the relative abundance of particle attached IPLs vs free living sources (Cantarero et al., 2020). This suggests that the majority of the biomass (and IPL content) measured at these high chlorophyll depths is likely to be derived from phytoplankton. We note that our mesocosm study compresses the depth of the chlorophyll maximum and oxycline into a 20m deep system, likely resulting in a greater contribution of phytoplankton lipids in ODZ waters than would be expected stretched over a deeper oxycline with reduced light availability. Bach et al. (2020) demonstrated that whereas light intensity decreases quickly with depth in these mesocosms, the chlorophyll a maximum remained in the upper 5m in the week after the ODZ water addition, but shifted to the intermediate depth range of between 5 and 15m thereafter and until day ~40.

Despite this, we agree that there is no perfect separation between bacterial and eukaryotic IPLs and that we can not rule out some contribution from the former to our data, which we consider to be predominantly but not exclusively phytoplanktonic in origin. As we highlight in section 2.5, our approach to minimize the contribution of bacterial IPLs includes the removal of molecules with odd and/or short chain fatty acids as well as certain headgroups exclusively found in bacteria). However, we recognize that there is certainly still a bacterial component in these IPL distributions and we will qualify that potential in the edited manuscript when suggesting the evidence for a greater proportion to phytoplankton biomass. This will include a detailed review of bacterial and phytoplankton distributions based on the reports from other publications in this special issue (Bach et al., 2020; Min et al., 2023).

We do value this comment as it also highlights the need for environmental lipidomics in order to parse the many biological sources and environmental forcings on the lipid pool. Indeed,

we have a follow-up manuscript in prep focusing on IPLs thought to be predominantly found in bacterial biomass (following the reverse filtering procedures stated above). We hope to contribute to this growing field in a subsequent publication focused on the biological sources and environmental drivers among bacterial and archaeal plankton.

RC3-4: Finally, throughout the Discussion, the authors suggest potential triggers that induce microbes to alter their IPL pool (e.g. light, pH, temperature, etc). However, the authors then jump to invoking them as the actual cause in the Conclusion, which I do not think is a valid conclusion from correlation analyses—i.e. correlation does not imply causation.

Response RC3-4: We agree that this is an overstep in the language used to summarize the correlations observed in these analyses and will clarify the degree to which we may assert these claims.

Specific comments

RC3-5: I hesitate to question data, but in Fig. 2D, Subsurface, on Days 12, 16, and 18, PG comprises almost 75% of all IPL's. I know of no phytoplankton or marine bacteria where PG makes up anywhere near that percentage—I typically think of them on the order of 5-10%, maybe 20%, in phytoplankton (e.g. Cañavate et al. New Phytologist, 2017; Popendorf et al. Org. Geochem. 2011). Anything higher than that would, to me, seem more indicative of marine bacteria, which, as I said, were not mentioned in the manuscript. I would recommend the authors check their quantification calculations.

Response RC3-5: We agree that these are high relative abundances for PG (up to 65% of the total IPL pool in some extremes) for predominantly phytoplanktonic biomass. We do analyze an exhaustive array of calibration standards to account for ionization differences between every IPL head group quantified in this analysis. In addition, we have several deuterated standards to account for potential matrix effects (including a deuterated PG standard). These method details are referenced in Cantarero et al., 2020 but will also be included in the edited manuscript per the request of Reviewer 1).

We recognize as stated in RC3-2 that there is likely some component of bacterial biomass in the form of IPLs that are common to both domains. While many of the more abundant individual moieties in this experiment (e.g. PG 32:0, PG 30:1, PG 34:2, PG 36:1) have shown strong correlations with high chlorophyll concentration in the ETSP (Cantarero et al., 2020), we do agree it is important to clarify that some of these trends could be in part driven by enhanced bacterial contributions to the IPL pool.

RC3-6: There are many citations in the manuscript where sources are not listed in the bibliography – e.g. Brandsma et al. 2011 (L 670), Abida et al. 2015, Urzica et al. 2013, Gordillo et al. 1998 (all L 81) – there may be more, but these are the ones I happened to check because I was interested.

Response RC3-6: These will be corrected.

RC3-7: L 861 – Where does the 2% N come from? (Citation?)

Response RC3-7: This was an estimation based on the stoichiometry of a typical betaine lipid with a molar mass of ~700 g/mol. We will clarify that we make this estimation in the text.