Referee #3

This paper provides a very detailed account of the IPL composition in the waters of the oxygen-deficient zone (ODZ) of the Eastern Tropical North Pacific. It adds to the growing inventory of IPL data. The authors claim that it contributes to our understanding of these systems. I am not entirely convinced. There are also a substantial number of issues related to the analytical methodology that need to be resolved before this paper can be published.

The general aim of the paper is "to evaluate the microbial ecology and ecophysiological adaptations that affect organisms inhabiting the OMZ in the context of biogeochemical cycles" (line 30-32). However, when you read the remaining part of the abstract this does not materializes. It ends with a rather vague statement about potential phosphorous limitation, which in the light of the analytical constraints (see below) may be even weaker. I strongly suggest to remove these kinds of claims from the text and just focus on what the paper is about: an inventory of IPLs in the water column of this region.

We thank the reviewer for the constructive comments. We believe that this study contributes to our understanding of the microbial ecology of the system, as (1) a large proportion of the microbial communities in OMZ currently remain grossly uncharacterized (cf. Podlaska et al., 2012) and IPL analysis can at least help evaluate bacterial vs archaeal sources, and (2) we report on a variety of changes in the lipid composition that relate to changes in biogeochemical zones and may represent ecophysiologic stress adaptations. We thus believe that our statements in lines 30-32 are justified. Nevertheless, we slightly modified other instances in the text where our interpretations might have been too bold (e.g., lines 468-470). While we agree that this study may not be able to fully explain the observed IPL distributions, we still provide ideas of what their functional roles or biological sources may be. Thus, this study lays the groundwork for future investigations that may probe the suggested sources and functions of the lipids we report.

In terms of trustworthiness of the presented data due to analytical issues: Even if there may be some issues with respect to providing absolute numbers, the detection of these 'unexpected' lipids in the ODZ cannot be dismissed and remains to be explained. Secondly, we do not believe that the presented data is prone to more bias than is usually the case for any lipid-based quantification. Finally, this is not the first and only study that observes non-phosphorus lipids in anoxic environments, therefore we are encouraged that our data is credible.

While revising the manuscript, we considered to not make bold claims that we cannot substantiate and aimed at making our main conclusions and take home messages clearer. Furthermore, we addressed potential analytical biases (see section 2.3)

The authors do report absolute IPL concentrations (Fig. 3b) which show an order of magnitude decrease over the first 100 m of the water column. This is in line what would be expected since this is the zone where primary production is taking place (as is also revealed by the pigment concentrations) and the data would allow to discriminate IPLs produced by phytoplankton in the photic zone from IPLs (produced in much lower concentration) by prokaryotes residing below the photic zone and within the ODZ. However, the focus in the paper is too much on relative abundances of IPLs (e.g. Fig. 5) for unknown reasons. The discussion should be more focused on the zone where IPLs are primarily produced (i.e. the photic zone) vs. the remainder of the water column that may be influenced by IPLs in descending particles (i.e. produced in the upper water column) and additional production by prokaryotes. In this discussion, it should be taken into account that PUFA-containing IPLs may degrade faster than other IPLs. Now, the total inventory of IPLs is too much discussed in terms of relative abundance in relationship with nutrient profiles and other environmental parameters over the whole water column, which is too simplistic.

The main focus of the paper lies in observing shifts in community composition by looking at relative changes in IPL composition and we are furthermore putting this into context to the different biogeochemical zones. For this reporting the absolute concentration is only of secondary importance.

It is true that quantitatively IPLs are most abundant in the photic zone and that most of the export of IPL therefore likely occurs from the photic zone (see also Kharbush et al., 2016, OGC 100, 29-41 for more details on this). However, this observation is not very relevant for this paper, because here we focus on the suspended particulate material (SPM, i.e. the in situ living microbiota) and are not looking at exported material as we are filtering out larger particles of marine snow.

Since we see shifts in the relative IPL composition (both in the head group and core lipids) within the different zones we are fairly confident in stating that IPLs are produced throughout the entire water column by different microbial communities and not primarily in the photic zone as stated by reviewer #3.

We would like to highlight here again the importance of microbial processes in ODZ (as introduced in our introduction) and that many important global element cycles, such as the nitrogen cycle are affected by microbes living BELOW the photic zone. Therefore, we see no reason why in this study we should solely focus on the photic zone. This would be completely beside the point we want to make with this study.

In addition, higher absolute abundances do not equal a higher significance: Since small sized prokaryotic picoplankton may not produce as abundant biomass as phytoplankton does, diagnostic IPLs that may be indicative of either changes in the microbial community composition or their adaptations to environmental conditions may only be present in small amounts. For instance, ladderane lipids that are diagnostic for anammox bacteria are only present in very low amounts in the ETNP water column (pg L-1; Rush et al., 2012, OGC 53: 80-87), but this does not mean they are insignificant as these organisms are responsible for large scale denitrification in the global ocean.

In summary, we disagree with reviewer #3 that our approach is too simplistic, instead we would argue that only looking at absolute concentrations and consequently interpreting IPLs in the deeper water column to be mainly derived from exported material out of the photic zone is too simplistic.

The paper also suffers from too much detail. It is very hard to follow because every tiny IPL detected is described without a clear environmental implication. The authors should formulate specific research questions (i.e. not understanding the "functioning of OMZs throughout the world ocean" by studying IPLs) and address these. Not every minor IPL detected has to be described!

The manuscript has now been substantially revised, the discussion has been shortened by ca. 6 pages and we made sure to better formulate our research questions (lines 135-143) and the summary of the main results (lines 708 and 731).

There are also a number of issue related to the analytical methodology of analyzing the IPLs. Adequate answers should be provided on all issues since this may seriously affect the interpretation of the data.

We have revised the manuscript accordingly and addresses potential issues with respect to the analytical methodology (see section 2.3).

1) Filtration. The authors used 0.7 micrometer glass fiber filters for filtration. The limitations of the use of this filter size has extensively been discussed elsewhere and the authors acknowledge the limitations of their approach by admitting that they might be missing smaller cells. However, they should also mention that the pore size will decrease during the filtration process and thus will recover a (variable) part of this material. More importantly, they should stress that this does not only lead to "minimal values" for IPL concentrations but that it may also affect the distribution of IPLs that they report (as percentages). Furthermore, they used a prefiltration device to "remove most eukaryotes"

(line 169). It is hard to believe that this will remove most of the algae; if so this would also strongly affect their interpretations.

We added a sentence in the methods section to address variable material recovery due to shrinking filter pore size (lines 165 and 166). We revised the sentence on pre-filtrations and changed 'most eukaryotes' to 'larger eukaryotes'.

2) Extraction. The IPLs have been Soxhlet extracted (i.e. boiling DCM/methanol for 8 h). Although this is a common method for extraction of less labile lipids, for IPLs it is rather unusual since these are very labile and the commonly applied method for this type of work is Bligh Dyer extraction at room temperature and controlled pH conditions. The authors should present data to show that their extraction method does not alter the IPL distribution (i.e. their main target of study) due to the fact that some IPLs are more labile than others (e.g. in the ratio of phospho vs. non-phospho IPLs, which plays an important role in the discussion).

We would like to refer here to our comments to the second reviewer Dr. Bale as she raised similar concerns. As stated above (and now also in the manuscript, lines 195-197), Lengger et al. (2012, OGC 47:34-40) compared the typical Bligh and Dyer extraction with Soxhlet and ACE extraction. While overall the IPL yields were best with the Bligh and Dyer extraction, Soxhlet extraction did yield also more presumably labile phospho-IPLs (such as the HPH-GDGT). We are encouraged that we did not greatly discriminate against these allegedly more labile compounds as we abundantly detect HPH-GDGT in our samples. Having said this, our own (unpublished) experiences in comparing various extraction techniques often tell us one thing: extraction efficiencies often predominantly dependent on sample type and even the same extraction techniques may yield different yields in different environments. Extraction efficiencies thus affect any lipid study and are simply part of the known quantification biases that are associated with lipid-based studies (as is also the case for any quantitative gene-based technique for that matter).

We have raised awareness of the 'extraction efficiency issue' in the methods section so that the reader is aware that this problem exists (lines 195-197).

3) Analysis. The experimental description indicates that the IPLs have initially been analyzed by HPLC-ESI-IT-MSn using the same system as described by Schubotz et al. (2009). In the meanwhile, this group has developed improved methods of analysis of IPLs (e.g. Wormer et al., 2013) and the question arises why these "old" results are still used since the samples were also re-analyzed using these new methods. It does not become clear how IPLs can be quantified with one method and identified by an- other method (lines 206-207) using an entirely different separation system. One issue that should also be addressed is the timeline of all these analyses. Once extracted, IPLs cannot be stored for a substantial time without significant alteration/degradation. Schubotz et al. (2009) in their very much related work of IPLs in the Black Sea stated

"Three years after primary qualitative analysis the samples were spiked with C16-PAF (1-O-hexadecyl-2-acetoyl-sn-glycero-3-phosphocholine) as injection standard and re- run for quantification. Slight changes in the relative distribution of IPLs were observed within the two runs. In particular the differences were identified as a selective loss of the glycolipids Gly-DAG, Gly-Cer, Gly-GDGT and 2Gly-GDGT (data not shown). We interpret this loss as a sign of selective degradation of glycolipids during storage." So, an important question is how much time evolved between these two analyses and can the results still be compared?

We apologize that we were not very clear in describing the sequence of IPL analysis in this study. We have now substantially revised the methods section 2.3 to clarify when which samples were run and why. Unfortunately, for our second IPL runs in 2015 we are not able to provide absolute quantitative numbers since in between 2010 and 2011 the samples were used for multiple other analyses (published

in Zhu et al., 2016, EM 18: 4324-4336) where overall total lipid extract losses could not be accounted for anymore.

With respect to the selective degradation of IPLs, we would also like to refer to our comments made to reviewer #2 and in addition we would like to state that we did not observe a similar selective loss of glycolipids between the runs that were performed in 2010/2011 and 2015.

4) Quantitation. It does not become clear from the experimental description if the commercially available standards were run with the HPLC-ESI-IT-MSn system that was used for quantitation. If so, the results (response factors) should be reported. If not, there is a serious problem since response factors should be determined on the same system. The whole procedure of quantitation should be made clear. An "injection standard" is mentioned but it remains unclear what was done exactly. Why didn't the authors use an internal standard that was added to the extract? Such a standard would at least have been exposed to the same conditions as the IPLs of the samples (e.g. during storage). The authors should also report the analytical errors of their determinations. The data they now report (e.g. Table S2) are extremely accurate (e.g. a relative abundance of 16.68 % SQ-DAG in station 1 at 35 m). I would expect that the analytical error is 10-20%, so the reported data are far too accurate unless the analytical error is extremely low. This also holds for many of the other data: the reported accuracy of absolute pigment concentrations is also (far) too high and so is the data reported in Table S3 (if the SD is larger than the figure itself is odd to report three or four significant numbers).

Again, we apologize that our quantitative approach was not clearly formulated. We have now revised section 2.3 extensively. We can assure reviewer #3 that the response factors were determined for the same instrument that the samples were run on – anything else would be unacceptable. For quantification of IPLs response factors have to be newly determined every time samples are being run, because not only do they change from instrument to instrument but also over time for the same instrument due to re-tuning after every cleaning. We therefore determined response factors for the 2010 and 2011 runs that were performed on the ion trap system and for the 2015 run on the HPLC-QTOF-MS.

Since response factors are so variable and not only change from instrument to instrument but also within months, we are apprehensive in providing the absolute response factors as this will not be of value to anyone. Instead, we are providing the range of standard variation (line 243) and we have also now provided detailed information on the used standards for response factor correction for the different instruments in a table in the supplementary material (Suppl. Table 2).

Additional comments:

Line 40: It is useless to compare relative trends in IPLs with absolute trends of environmental parameters. To this end, absolute concentrations (like you have for pigments) need to be used.

As already stated above, this study focused on relative abundances of IPLs in order to track (relative) changes in microbial communities and their ecophysiologic adaptations. Solely looking at changes in absolute abundances would not provide this level of information, therefore we disagree with the reviewer that looking at relative trends is useless.

Lines 68-69: . . .. . .but not provided in this paper, so remove this sentence.

This sentence is a general statement and we don't see the need to remove it just because we cannot provide the answer for this in our study.

Lines 117-119: strange sentence

The sentence was revised.

Lines 119-124: I think this overview should be limited to papers describing intact IPLs in the water column. For example, the data presented by Lincoln et al., 2014 are not really solid IPL data. Turich and Freeman, 2001 and Hurley et al., 2016 present only core lipid data.

This section has been revised and references have been removed.

Line 125: It is not discussed how IPLs can be used as taxonomic indicators. This seems pretty relevant information for this type of study

We believe that we have extensively introduced the concept of using IPLs as taxonomic markers in the paragraph above this referenced sentence, e.g., glycolipids for phototrophic organisms, betaine lipids for lower plants and algae and specific phospholipids for either phototrophic organisms or bacteria.

Line 139: in what way is this approach "complementary"?

This whole paragraph has been revised (lines 135-143).

Line 146-147: here the authors promise that we should learn a lot (deeper insight into biogeochemical cycling, functioning of OMZs throughout the world ocean) but this is grossly overstated.

This sentence has been revised according to the statements we can actually make (line 142-143).

Lines 152-152: data on coordinates of sampling stations is incomplete.

Revised for completeness (lines 148-149).

Line 166: provide details on sampling volume

Pore water volumes are now listed in Supp. Table 1

Line 186: provide details on pore size of filter

Done.

Line 200: storage at -20 degrees C is not sufficient to prevent alteration; IPL extracts should be stored at -80 degrees C and even then, distributions may change. How long were the extracts stored before analysis for IPLs?

We wonder on which study the reviewer bases this statement. To our knowledge no literature data exists that tested IPL labilities at -20 vs -80°C. We have acquired knowledge on IPL analyses and stabilities for more than 15 years and based on our experience (re-analysis of the same samples and standards, some of which have been stored at -20°C and -80°C), IPLs are not as labile as they were originally assumed to be.

Since both reviewers #2 and #3 have expressed concerns on the storage and analysis of IPLs we addressed these issue in the revised manuscript. (lines 202-204).

Line 210: different column than in other analysis. Why?

Because the method has improved (as the reviewer has already noted in comment 3 above regarding separation of compounds, peak shape, etc.) and this is the current state of the art of analysis of IPLs and we wanted to analyze the samples with the best available method. This has been clarified in the revised methods section 2.3.

Lines 217-219: provide more details on these standards. What are the acyl moieties of these standards? How are the response factors affected by unsaturations in the acyl moieties? What was the time between the arrival of these standards in the lab and their measurement? How were they stored? It is known that these standards degrade over time upon storage. How often were these standards run? Before each batch of analyses? How did the response factors vary over time? Answers to these questions are essential for getting a feel for the confidence we can have in the reported relative abundances of the IPLs.

We now included a Table in the Suppl. Material (Suppl. Table 2) that contains all the detailed information on the standards that were used. With respect to the effect of the acyl moieties, yes there are differences in the response according to the chain length, but the effect of acyl chain length does

not affect the response factor as much as the head group does (see also Popendorf et al., 2012, Lipids, 48: 185.195 for details on this). We hereby would also like to point out that it is impossible to have a standard for every single analyte (with different chain lengths and/or unsaturation/ring) that exist in a sample, considering that we are identifying and quantifying 600 to 800 compounds in each sample. Firstly, many of these compounds are not commercially available and secondly it would be utopian to prepare the missing standards in the lab due to the required biomass and necessary time. Due to these mainly practical reasons, it is generally accepted in the IPL community that response factors are merely corrected for the different head group classes.

We cannot stress enough that IPLs are not as labile as reviewers #2 and #3 point out. It is unclear what the reviewer means with "it is known"? We are not aware of a reference supporting this statement. We run standards every time we run samples, firstly to get an understanding of the performance of our system over time and secondly to use them to correct for response factors. The standards change over time as the instrument changes over time, this lies in the nature of every instrument and this is why we run standards to correct for these natural instrumental fluctuations (as should be done for any instrument and method).

Line 224-224: why would the unknown response factors be in the range of the measured standards? This is not a scientifically acceptable statement in this way. Just say that they are unknown and what you have assumed to be able to calculate a concentration.

The sentence has been revised.

Lines 258-262: provide references to indicate that pigments can be used to reveal this information even at the species level (e.g. Rhizosolenia).

Pigment assignments were done according to DiTullio and Geesey, 2002 (the reference is provided in the manuscript) as is usually the case for pigments in POM samples.

Line 267: secondary maxima are not revealed in Fig. 1e.

Between 300 to 400 m there are lighter orange shadings compared to the red colors above and below.
Based on the reviewers comments we exchanged the section plots with XY-plots. Hopefully now the secondary maxima can be better seen.

Lines 268-269: Not evident from Figs 2a-b.
Based on the reviewers comments we exchanged the section plots with XY-plots. Hopefully now the described trends are better visualized.

Lines 273-276: So, the whole system is NOT P-limited!

Yes, and no, as this is exactly the point we want to make: From a common stand-point of oceanographers the whole system may not be P-limited (nutrient limitations are based on growth stimulation with nutrient amendment experiments and micromolar phosphate concentrations are concentrations where addition of phosphate may not stimulate additional growth). However, when one looks at bacterial cultures, these are exactly the phosphate concentrations where bacterial cultures replace phospholipids with glycolipids and are indeed growth limited. We have revised the text to clarify this (lines 701-705).

Line 282: How were absolute concentrations obtained?

See methods section 2.3 for this information.

Line 284: Secondary maxima are not observable in Fig. 3d.

Based on the reviewers comments we exchanged the section plots with XY-plots. Hopefully now the secondary maxima can be better seen.

Line 289-291: It would be logical to introduce first all IPL classes observed. Why are absolute concentrations of IPL classes not described and is the manuscript concentrated on relative abundances?

We opt to describe all the IPLs in detail later in the text because we find it logical to first provide a general overview of the grouping of glycolipids, phospholipids and aminolipids with depth. Based on the reviewers' suggestion we now mention the diversity of different IPL classes first and refer to Figure 4. Absolute concentrations of IPLs are described (lines 300-304). The reasons why we then focus more in detail on the relative abundances is explained above.

Line 292: "select substitute lipid ratios" is not introduced. It should be introduced in the discussion not in the result since it is an interpretation of the data presented.

We already introduced the (by now well-established) concept of substitute lipids in the introduction and we rather view the substitute lipid ratios as result that should be at least mentioned in the results section. We further discuss substitute-lipid rations in the discussion section 4.2.3.

Line 293: "total IPLs"? Does this now include archaeal IPLs or not (see line 291)?

We revised the sentence to make it clear that total IPLs excludes archaeal IPLs (line 312 and 316).

Lines 303-317: This section should be moved to the discussion. See also earlier comments on the distinction between the photic zone and remainder of the water column.

This section has been significantly revised and shortened.

Line 319 and Fig. 4: The whole distinction between major and minor IPLs is rather artificial. It becomes especially confusing when minor IPLs are normalized on their sum which is a variable part of the IPLs as a whole. It is entirely unclear why this is done other than for "stamp collection" purposes.

As explained above we report relative abundances of IPLs as we want to get an understanding of how the IPL composition changes with changing geochemical zones and microbial community composition. For this, minor lipids can be just as important as the major lipids. One example are ladderane lipids, which are highly diagnostic but are only present in very low abundances in the natural environment. If we would have not made the distinction of minor lipids (with an extra plot showing their distribution in Fig. 4) changes in their relative abundance would have not been evident due to their low abundance.

To answer the reviewers question: the distinction between major and minor IPLs was primarily done for visualization purposes in order to identify potential depth-related trends. Furthermore, this distinction also makes it clear which IPLs were used to determine the absolute amounts (as only major IPLs were seen with the ion trap system, while minor IPLs were additionally detected with the QTOF system, see also revised section 2.3).

Line 332: it would be much more helpful to report absolute concentrations. In that way it can be directly compared with the abundances of archaeal cells. Now, it is normalized to something where it is not at all related to and which varies by more than an order of magnitude (Fig. 3b).

As explained above the purpose of this study was to get an understanding of how the IPL composition changes with changing geochemical zones and microbial community composition. For this we do not need to examine changes in absolute concentrations but rather in relative concentration. Furthermore, absolute archaeal concentrations are already reported in Xie et al. (2014).

Line 352; formally this statement is incorrect: the chain length was not measured but the number of carbon atoms in the acyl chains. One cannot discriminate between branched and straight chains. The number of double bonds was also not determined since one cannot discriminate between a double bond and a ring. Needs adjustment.

The reviewer is correct with this and we revised our definitions accordingly. Chain length was changed to number of carbon atoms in the hydrophobic chain and double bonds changed to double bond equivalents (DBE).

Fig. 6 does not really show a lot of useful information since the variation observed is not extensive. I would consider to drop this figure.

As described in the text (lines 353-360), we believe there can be significant variation observed in this figure and would therefore like to keep it as one of the main figures.

Line 366-367: this definition and the one at line 320-321 does not exclude that some groups are both minor and major lipids. Anyhow, this distinction is extremely confusing as mentioned before.

We agree that at some depths some of the major lipids may be minor (but not the other way around). However, as already mentioned above, this distinction was (a) done for visualization purposes and (b) to distinguish between the ion trap and QTOF runs and we would therefore like to stick to this distinction.

Lines 365-406: only describe IPLs that are useful in the discussion.

The discussion has been substantially streamlined and shortened.

Line 383: One cannot make the distinction between an OH group and an epoxy group with the methods applied. Can the authors exclude that these components are formed upon storage of the extract?

Based on the MS2 data we are quite confident that we can make this distinction. Whenever we are observing the loss of water in the MS2, we see this as an indication for the loss of a OH-group. Even if a ring opening and loss of the oxygen from the epoxy-group would occur in the MS2 (which we doubt because this would require the addition of a nucleotide), all of the observed MS2 fragments could not be explained (due to the missing double bond that is formed during the loss of a OH-group). Whenever we do not have mass spectral evidence for the loss of water we do state that it could also be an epoxy group (line 531).

We do not have evidence that these oxygenated fatty acids are formed during storage: (1) Firstly, we do not see that these lipids appear (or increase in abundance) in other lipid extracts that we have stored over several years at -20°C. (2) Secondly, we know from the literature that OH- and epoxy-lipids are common components of algae and other organisms under stress conditions, therefore it is not surprising that they would be present in these samples. (3) Finally, if they were formed during storage, why would they not be equally present in all samples and why would they only form in some IPLs and not in all fatty acids in similar proportions? Certainly, additional investigations need to be conducted to further identify their structures and potential sources.

Line 408: the authors should make clear why it would be useful to perform statistical correlations between environmental variables and relative abundance of IPLs. This remains entirely unclear. I suggest to skip this entire section.

We respectfully disagree with the reviewer. This is a concept that has been applied before in many environmental studies, including gene-based studies (e.g., Legendre and Gallagher, 2001, Oecologia 129: 271-280; Ramette 2007, FEMS Microbiol. Ecol. 62:142-160; Rossel et al., 2011 GCA 75:164-184). Reviewer #3 seems to have the misconception that changes in communities can only occur in absolute amounts but not relative to each other. In a simple example: if there is quantitatively more light one can expect a relatively higher abundance of phototrophic organisms (which in turn synthesize phototrophic lipids) compared to low-light areas where there will be relatively lower abundances in phototrophic lipids. This is a simple correlation between relative abundances of IPL (or organisms) to absolute concentrations of environmental parameters that can be statistically evaluated, and we therefore do not see the justification to skip this section as it is an essential part of the paper.

Line 433: The discussion is extremely lengthy (19 pages of text) and should be focused on the important observations taking into account all the comments made so far. It should be cut by 50% or

so. It is, therefore, not useful to provide detailed comments and I will restrict them to more general comments.

We substantially revised and shortened the discussion by one third from the originally 19 pages to 13 pages.

Lines 434-487: General overview which is not connected to IPL dataset at all. Requires substantial shortening.

Done.

Lines 490-492: This statement needs to be proven by showing some kind of correlation.

This obvious trend can now be clearly observed in Figure 2.

Lines 492-497: First time export of IPLs is mentioned; this should be introduced in the introduction since it is important for the reader to understand that IPLs at depth comprise a mixture of exported and newly produced IPLs.

We actually do not expect the reported IPL signal to represent exported matter from above since we are looking at suspended particulate organic matter and not larger aggregates of exported material from above. Therefore there is no need to introduce this concept in the introduction.

Lines 497-499: Bold statement that is not (yet?) backed up by any data. Does not belong here.

We do not think that this is a bold statement since it is indeed backed up by our data: as described in the results section we observe 24 different IPL classes that also exhibit changes in core lipid composition with depth. What else should this IPL diversity reflect? Nevertheless, we toned the statement down (lines 468-470).

Lines 502-505: Repetitious.

This section was revised.

Lines 508-592: Very lengthy discussion assessing nothing really new: the IPLs in the photic zone derive from algae, cyanobactera, and heterotrophic bacteria. This was to be expected and has been shown previously.

The section was revised and shortened.

Lines 595-641: What I miss here is an answer to the intriguing question: how much of the IPLs detected in this zone can be derived from settling from the photic zone.

Not that much, this is why we show the changes in core lipids (carbon atom number and DBE), this is also stated at several instances in the manuscript (lines, 541-545, lines 573-574, lines 588-591).

Lines 631-641: Again, nothing new here. Have the authors evidence for the presence of specific GDGTs derived from Thaumarchaeota (i.e. crenarchaeol)? It would be fair to refer to the original literature for the detection of HPH GDGTs in Thaumarchaeota.

Yes, as stated in Xie et al. (2014) we detected crenarchaeol and are now also showing the distribution of the core GDGT composition for the individual IP-GDGTs in Supp. Fig. 2. We revised this section according to the reviewers' suggestion (line 562) and are also now citing Schouten et al. (2008).

Line 648-653: It is highly unrealistic to suppose that PUFA-containing IPLs would be produced in-situ in the ODZ. It seems the authors agree but the text is really confusing.

We do not agree that this is highly unrealistic, PUFAs have been shown to be synthesized by marine bacteria (see references cited in the text, lines 582-586). PUFAs could thus very well be produced in situ. Section 4.1.3 has been substantially revised and hopefully it is less confusing now.

Line 655: "due to the increase in bacterial abundance"? I guess bacterial abundance is still highest in the photic zone. The authors seem to forget that they are looking at relative abundances but when they would calculate absolute abundance a completely different picture emerges.

As already explained above, we do not understand the reviewers concern that we cannot use relative abundances of IPL to describe changes in relative community composition and/adaptations to environmental conditions. We disagree that a completely different pictures would emerge when only reporting absolute abundances.

Generally, bacterial abundance is highest in surface waters (with some exceptions at stations 1 and 2, see Fig. 2). However, what the reviewer seems to miss is that relative bacterial abundance is not highest in the photic zone, instead biomass (and IPLs) derived from eukaryotic phytoplankton dominate. In the ODZ, where phytoplankton abundance decreases (or is absent) relative bacterial abundance increases. This is also why we have more IPLs that are derived from bacteria within the ODZ and below.

Lines 659: these genes are much more widespread in the bacterial kingdom.

Yes of course, but we are discussing the IPLs here only in light of what is known from the FISH analyses performed at these sites.

Line 662: "chain" is incorrect

The term was replaced.

Line 674: the comparison of the IPL dataset with the FISH dataset is underdeveloped in this manuscript.

Where possible and applicable we referred to the existing FISH dataset (lines 457, 552, 566, 578 and 637-639).

Line 693: for PUFAs I would make a clear distinction between C20 and C22 PUFAs and the C18 ones, otherwise the text will become confusing.

The distinction we made for this was to call the C20 and C22 PUFAs 'long-chain PUFAs', we tried to be consistent using this distinction.

Lines 702-703: This strongly depends on the core of the GDGT IPLs detected. Crenarchaeol has only been detected so far in cultures of Thaumarchaeota. Suggestions that it may derive from euryarchaeota or crenarchaeota are only based on very indirect evidence and quoting these references (and not many other literature) in this context is only confusing the issue. In fact, one way to shorten this paper is to take out all the data related to isoprenoidal lipids. Part of this data has been published before (Xie et al., 2014) and the data reported here do not provide any new insight.

We disagree with the reviewer that our isoprenoidal data does not provide any new insight as now we also report on the presence of the important thaumarchaeal marker HPH-GDGT, which was not done in the previous Xie et al. (2014) data set. We revised this section of the manuscript and are now also providing supplementary information on the ring distribution of the individual IP-GDGTs (Suppl. Fig. 2).

Lines 712-825: Extremely confusing title. We just had a very extensive description of how species composition could influence IPL distribution. This section seems to start all over again (lots of repetition). The statistical data treatment is doubtful as explained before. With all the major changes in environmental parameters (light level, oxygen concentration) and its consequences for species composition it is impossible to link changes in nutrient concentrations to differences in IPL distribution. Such studies should be done first with microbial cultures and then, perhaps, these data can be used to interpret environmental datasets like this one.

We substantially revised the discussion, including the headers and the order of the different sections. The purpose of the three sections is to provide an overall synthesis and summary of the above described IPL species. As explained above we do believe our statistical evaluation to be relevant and necessary as it does provide additional insight into the zonation of IPLs into different geochemical

zones. We hope that by shortening and streamlining the discussion our main points we want to make are now expressed in a clearer way.

Figures 1: Explain how (software; kriging method) plots b-e were produced. The figure suggests (but the caption does not explain) that only at stations 1, 2, 5 and 8 full CTD casts have been obtained. Stations 1-8 are almost 3000 km apart. Is it statistically significant to perform interpolation between these stations for the deep (>200 m) waters? One can observe all kind of changes (as shown by colour changes) that are hard to understand from having only 4 data points over 3000 km. Specify the depth scale; it does not seem to be linear but it is not specified in the caption. Most of these comments also apply to Figs. 2,3, and 5.

Based on the reviewers comments we exchanged the section plots with XY-plots.

Table S3: The authors cannot report the number of double bonds; they determined the DBE number but they cannot discriminate between a ring or a double bond as far as I can tell. The table should be carefully checked; there are instanced where the FA combination says 18:4/18:4 but the number of bonds in nine. It is also not clear why sometimes they report the FA combination and sometimes not even for the same type of IPL. They should also specify where the relative abundance is normalized upon. It would be more convenient for the reader when these values are followed by a plus/minus sign and then the SD.

We changed the terminology and call it now double bond equivalents (DBE) and also revised Suppl. Table 5 according to the reviewers' suggestion.