Response to reviewer 1

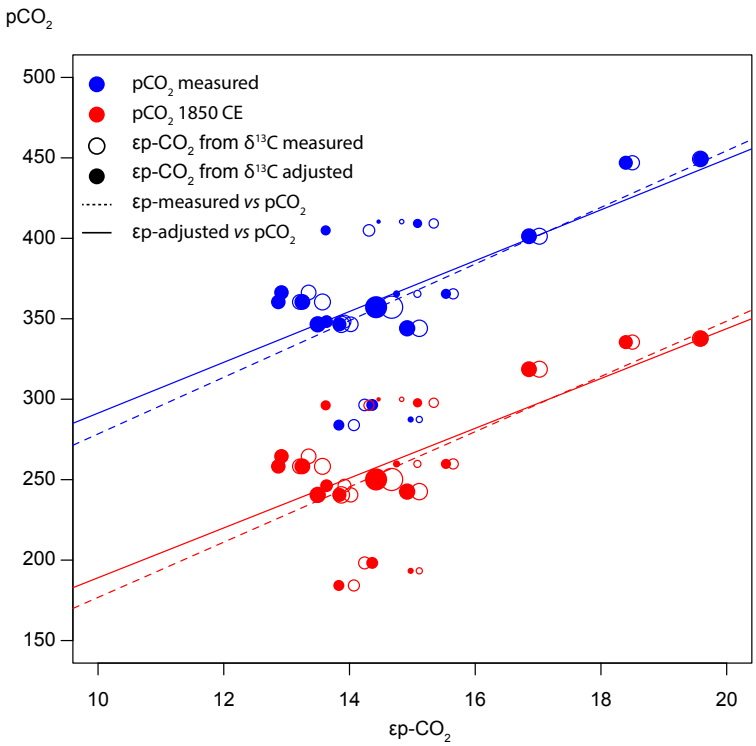
General Comments
Frieling et al., present records of carbon isotope fractionation from the resting cysts of dinoflagellates to investigate their utility in reconstructing ancient atmospheric CO2. This record of core-top material advances earlier work based on laboratory cultures (and based on sound theoretical basis) and so brings the community closer to confidence that this proxy may work in environmental settings. They show there are differences in carbon isotope fractionation between different species, emphasising the importance of single-species records, and show greater 13-C depletion in their core-top samples compared to cultured, motile organisms. The paper is interesting and makes an important contribution, but some of the analysis is unsatisfactory due to uncertainty about the age of the individual cysts in the "core-top" samples (detailed below). Therefore without a thorough treatment of that uncertainty (which is currently lacking) it's difficult to know whether this proxy has utility. There are certainly hints that it does, but unfortunately this paper does not yet demonstrate that compellingly.

Author response:
We thank the reviewer for recognizing the potential importance of our work and the constructive criticism. In the response below and in our revised manuscript we have further clarified (1) how the carbon isotope data from individual cysts has been treated and (2) we further elaborate on the uncertainty in the age dating of the core-tops.

Specific Comments
The problem with using core-top samples is the substantial increase in atmospheric CO2 since the industrial revolution. As the authors note, it is highly uncertain whether the cysts are from the last week, the last year, decades or even centuries ago. The uncertainty around the contemporaneous CO2 is potentially very large. The "rough correction" to 1850 isn't really a correction at all, but an assumption which is not well supported, at best highly uncertain, and not really dealt with satisfactorily in the later analysis. The best approach (although expensive) would be to 14-C date some of these samples to see when this material actually dates to. The cheaper, and for this present study, more plausible approach would be to propagate through what is a really quite large uncertainty and see whether the conclusions still hold. Lines 147-8 state that "With the exception of pCO2, we hence assume all parameters (SST, SSS, nutrients) to be constant over the period the core top samples represent." A fundamental problem here is that the authors have little information (or at least present little data) about how long a period of time the core top samples do in fact represent. I'm not sure that the approach taken to this, systematically removing the most 13-C depleted samples is appropriate. Whilst it is certainly plausible that these individuals represent modern samples, the evidence is fairly circumstantial, and they could represent another confounding variable. What is the impact on the analysis if these samples are not removed?

Author response:
The reviewer comments on the potential of age-mixing of individual cysts in core-tops. This is very much a valid concern as we acknowledged in our original manuscript (lines 196-201). The main challenge here is that core-tops (the top-most 2 cm of sediment) contain individual sedimentary components with a range of ages. While it might be possible to ^{14}C-date carbonate, bulk organic matter or even specific chemical components, these materials will derive from different times in the past not necessarily the same as the dinocysts analyzed here. Hence, a single measurement will not show the range of ages of the individual dinocysts

(i.e. the age-distribution of our individual dinocysts). Ideally one would date single-cell dinocysts using ^{14}C analyses but that is technically not feasible (yet).
As the reviewer correctly points out, alternatively one needs to show potential impact of the data-treatment, which we now include in a supplementary file to the revised manuscript (Supplementary Figure 1). In general, the impact of removing analyses based on exceptionally low amounts of C and exclusion of outliers has no appreciable impact on the regression parameters. The largest uncertainty indeed comes from the recent addition of anthropogenic carbon, the Suess effect. Comparing our calibration including and excluding the Suess effect is to our opinion therefore the best approach to estimate the maximum uncertainty in the regressions (lines 211-239).



Supplementary Figure 1. Effects of data treatment on the difference between measured and adjusted pCO_2 and $\delta^{13}C$ ($\varepsilon p\text{-}CO_2$). Open symbols indicate measured $\delta^{13}C$, closed symbols represent data after eliminating small signals (<0.2 Vs) and outliers. Blue dots represent measured CO_2 values and red dots indicate the CO_2 around 1850 CE. For each dataset a simple linear regression, weighted to the number of measurements, is given. Dashed lines utilize measured $\delta^{13}C$, solid lines are from final $\delta^{13}C$ data. The red solid line is used in Figure 5A.

Supplementary Figure 1 illustrates the maximum error that may result from the Suess-effect by constructing a calibration with uncorrected data for each of the parameters (measured $\delta^{13}C$ and CO_2) and we compare this to the original calibration (solid red line). The main difference is in the intercept of the calibration, which is offset by the average difference in atmospheric pCO_2 between 1850 and the measuring date (~2000 CE; *ca.* 100 µatm) resulting from adjustment of measured CO_2 levels to pre-industrial times. The slopes of regressions using the data after drift-correction and outlier-removal are slightly shallower than those from the measured $\delta^{13}C$, but slopes of all regressions are statistically indistinguishable (lines 300-303) and the difference between the intercepts of the calibrations represents a worst-case scenario.

We also fully agree with the reviewer that it is critical to show that the data-treatment is free of potential preconceived biases regarding which values to include or exclude. We therefore further clarify in our revised manuscript how we first corrected for instrument drift at low C. This correction is exclusively based on repeat measurements of the IAEA-PE ($\delta^{13}C$ certified -32.15 ± 0.05‰) standard which show a convergence towards ~ -27‰ at the lowest intensities (original manuscript lines 172-174; see also Van Roij et al., 2017). After performing the drift correction, we subsequently identify outliers within the species-specific populations measured on a single sample (revised text line 226-231):

Instead, we therefore illustrate the influence of $\delta^{13}C_{DINO}$ data treatment and pCO_2 correction (Supplementary Figure 1). For this, we compared both measured pCO_2 and pCO_2 around 1850 CE (see section 2.2) to ε_p calculated using both our raw $\delta^{13}C_{DINO}$ data and the $\delta^{13}C_{DINO}$ data after drift-correction and removal of statistical outliers identified within the sample-specific single species populations.

We find a greater proportion of negative outliers (original manuscript line 215) compared to positive outliers (hence a skewed data set; original manuscript Table 1). Since we correct using the PE standard and exclude both negative **and** positive outliers there is no assumption that the most negative are the most recent cysts (which one might assume based on the Suess effect but would be an *a priori* interpretation).

The reason why we think our data-treatment has some of a Suess-effect related bias which may exist between samples is as follows:
If we assume that (a) statistical distributions of $\delta^{13}C$ cyst populations are dominated by both pCO_2 and $\delta^{13}C$-CO_2 trends and that (b) these populations include a portion of both pre-industrial and recent times – the trends in both pCO_2 (higher pCO_2 > higher εp > lower $\delta^{13}C$cyst) and $\delta^{13}C$-CO_2 (lower $\delta^{13}C$-CO_2 > lower $\delta^{13}C$cyst) would impose a (negative) skew on the $\delta^{13}C$cyst populations. Prior to correction for skewing at low intensities and outlier removal some, but not all, samples show a negative skew, while after outlier-removal the vast majority is statistically indistinguishable from a normal distribution (Table 1 in the original manuscript). In addition to the supplementary figure showing the minor impact of excluding outliers (see above), we have now clarified this rationale in our revised manuscript (line 226-239).

Ie L 213-4 “We assume these assemblages are representative of ocean conditions prior to the massive increase in anthropogenic carbon emissions.” This is very sizable assumption which whilst plausible is not currently supported by very much evidence.

Author response:
Please also see response above – we now provide a more elaborate reasoning why we assume that the vast majority of the analysed cyst populations may be considered to represent pre-industrial times or experienced only minor influence of anthropogenic CO_2. We will clarify in our revised manuscript that this is also based on the assumption that core-top material includes an age range, likely centuries up to millennia. This implies that in a core top sample only a very limited part of the population is derived from the most recent, anthropogenic times. For core-top samples a major impact of the Suess effect may be detectable only in the relatively few cysts that were formed in the past ~70 years (lines 211-225).

It appears from Figure 5a that no uncertainty at all has been applied to the assumed CO2 value – is this correct? This is not a fair assumption given the uncertain age of each cyst. In fact a plain reading of Figure 5a suggests that, rather than supporting a function between CO2

and ep apart from at <240 uatm CO2, ep is effectively constant, and only slightly higher above 310. Why therefore has 240 uatm been emphasised?

Author response:
The reviewer's comments prompted us to carefully revisit our analyses and numbers and led us to repair a few small errors and inconsistencies. As the data correction itself is a relatively minor adjustment (lines 236-239), any changes to the calibration equations and table 2 do not affect the final conclusions.

We clarify in our revised text that we do not expect any other major biases in the CO_2 gradient (see also Figure in previous response) after instrument drift correction and removal of outliers. The difference illustrated in the new supplementary figure 1 arguably shows a worst-case (maximum) effects of fossil-fuel derived CO_2 on the calibration. We therefore argue, regardless of uncertainty in the actual CO_2, the slope of the relation would not significantly change (lines 300-303). However, the reviewer is correct in pointing out that this may influence our interpretation at what value of CO_2 εp might become (in)sensitive to CO_2 changes. Accordingly, we added further nuance to statements on CO_2-insensitivity in our revised manuscript (lines 399-405) and better explain our approach regarding data treatment (paragraph 3.2.2 lines 206-250). We also make it clear now that the 240 µatm – level should be seen as the lower limit of (in)sensitivity and that, also for practical reasons, the quadratic calibration should not be used below this level (lines 299-300).

In addition, we clarify that for our original Figure 5D and 5E we propagated (via Monte-Carlo analyses) a 5% error on the measured CO_2 and nutrients, as well as the standard error on the mean dinocyst $\delta^{13}C$. We erroneously omitted error bars in Figure 5A-C and we did not provide a clear explanation of how errors were propagated for Fig. 5D and E. The revised figure now includes the error bars and the error propagation is properly explained in the Figure caption.

Whilst the data presented here are interesting and important, the analysis at present is not sufficient to support the conclusions drawn robustly.

Author response:
We hope with the clarifications above, the calculated (maximum) uncertainty and the proposed adaptations of our revised manuscript we have alleviated the reviewers' concerns.

Technical corrections
22 use of "significantly" if this is meant in the statistical sense, please add p and n values, else reword.
Author response:
Changed to 'appreciable' as these data are not directly statistically compared here.

24 ibid "significant"
Author response:
We have retained this statement - given the number of comparisons (20 for both variance and mean) we cannot provide p or n values of each comparison in the abstract of the manuscript. For significance of each comparison we refer the reader to Figure 3 and keep the original generalised statement in the abstract.

40-42 This is a slightly eccentric choice of papers to cite here. At a minimum add an "e.g." but better to make it clear why these papers or a more comprehensive survey of the pCO2 proxy literature.

Author response:

We have included "e.g." before the cited references.

43 "However, many of the organic compounds used for CO2 reconstructions are not related to a single species, genus or even group." A fairly sweeping statement here not supported by any references. Which records and compounds are you referring to?

Author response:

We have clarified this statement in our revised manuscript. The statement refers to proxy substrates such as phytane and alkenones mentioned in the previous paragraph (lines 43-45):

"However, many of the organic compounds used for CO_2 reconstructions such as phytane (Witkowski et al., 2018) and alkenones (Pagani, 2013) are not related to a single species, genus or even group of organisms."

54 "extremely long-ranging" in space or time? Please be specific and it time list age range.

Author response:

We have add the statement below (lines 54-57) to clarify this refers to the geological record of *Operculodinium centrocarpum* and *Spiniferites* species as used here (see lines 68-70 for their age-range).

"The organic resting cysts from autotrophic species have excellent preservation potential, are often highly oxidation-resistant (Zonneveld et al., 1997, 2019; Kodrans-Nsiah et al., 2008) and several ubiquitous extant genera and species, such as Spiniferites spp. and Operculodinium centrocarpum, have extremely long geological records (Fensome et al., 1996; Williams et al., 2004)."

70 Should "cyst species" by cyst-forming species?

Author response:

To clarify we have rephrased this to (lines 72-74): *'Although $\delta^{13}C_{DIC}$ exerts a major control on dinocyst $\delta^{13}C$ (Sluijs et al., 2018), it remains uncertain whether the CO_2 control on ε_p of motile cells from controlled growth experiments can be translated to their cysts formed in the natural environment.*'

90 "Using standard palynological techniques" Please provide a citation.

Author response:

We now refer to Brinkhuis et al. (2003) who described the standard cold HCl/HF acid-digestion procedures employed to obtain palynological residues (line 93).

94 "ultraclean water" what is this? Ie quote a specific measure such as resistivity if reverse-osmosis teqnique has been used

95 "milliQ" is a brand name not a type of water. Please revise.

Author response:

We have changed both these to demineralised water.

L104-5 is the 0.3-0.4 permil number precision or accuracy? How has accuracy been determined.

Author response:
We thank the reviewer for pointing out this unclear statement. We have removed 'accuracy' here as this statement was meant to refer to precision.

L343-4 "Badger, 2003, 2021;" These are two difference Badgers. Check BG style but likely need to include initials (lots, because they share first first name initial too).
Author response:
We thank the reviewer for highlighting this – we have revised to "Badger, M.R, 2003, Badger, M.P.S., 2021".

L 385. I'm not sure this is sufficient to meet the journal data policy. Pangaea doi should be available at publication.
Author response:
We have included the Mendeley data DOI and will release the embargo upon publication.

Response to reviewer 2

The work by Frieling and colleagues is strong framework and a much-needed study that will open a new opportunity for applications of organic microfossil 13C analysis. Like single species foram analyses (the benchmark for modern carbon and oxygen isotope studies) single or several organic microfossil 13C limits the breadth of sources to sedimentary organic matter and limits the degrees of freedom in a highly advantageous way. This study is the gateway to the deeper geologic record that will allow broad application of the dinocyst proxy to ancient carbon cycle studies. The questions below are meant to enhance the discussion, but the work, as it is, stands on its own as it is presented.

Author response:

We thank the reviewer for the positive and constructive review of our work. The review highlights a number of aspects of the methodology which often do not appear in the published literature but in this case will be helpful to provide a baseline for further work.

From a methodological perspective I appreciate the details provided here. Controlling for size and process length is great approach but do you see relationships between $d^{13}C_{cyst}$ and cyst size?

Author response:

The reason we aim to exclude size-dependent $\delta^{13}C$ differences is that for e.g. foraminifera, coccoliths and living dinoflagellates a size-dependent ^{13}C fractionation has been observed previously (Burkhardt et al., 1999; Hoins et al., 2015). Unfortunately, a size-dependent $\delta^{13}C$ relation in modern ocean dinocysts is beyond what we can reasonably test with our method – because of the analytical uncertainty when measuring such small (30-40 µm diameter) individuals the number of required repeat measurements would become unpractically large. However, we fully agree this is a logical next step, once the methodology is sufficiently developed to achieve the precision needed to distinguish between individual cysts' $\delta^{13}C$ signature of modern species.

To what degree do you feel that the time averaging affected your data? Do you have access to any 14C dates of the surface sediments? From here you could potentially model the expected range of 13C values of DIC accounting for Suess Effect. More details in the manuscript on your rough correction would be helpful.

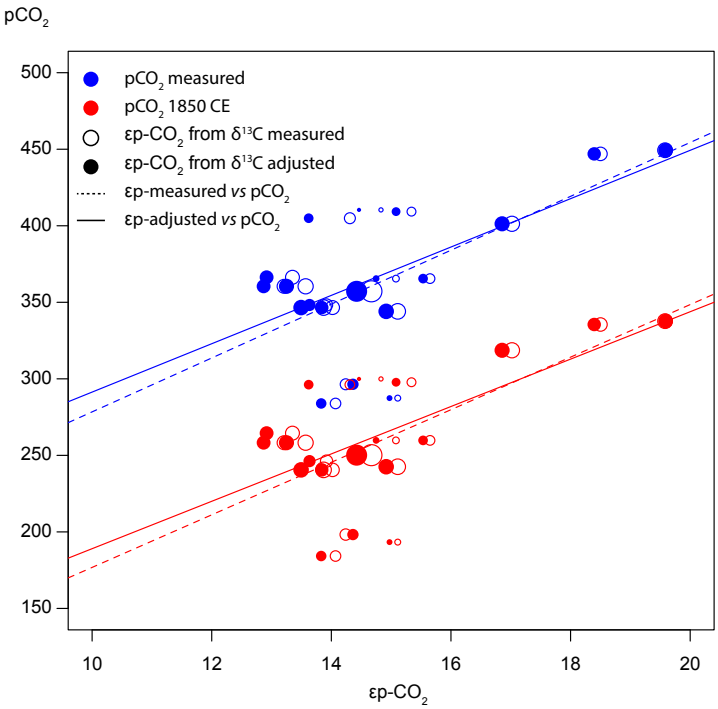
Author response:

This point is in line with one of the points raised by the other reviewer. We elaborate on our reasoning regarding age-control below and in the revised manuscript (lines 139-148).

We fully agree that better age-control on sediments, but especially of the individual cysts would be helpful and also modelling $\delta^{13}C_{DIC}$ (subtracting the Suess Effect) as the reviewer points out is a step we would like to take if feasible. However, the use of ^{14}C dates (often carbonate, otherwise bulk or macro-scale organic matter) is complicated as they cannot be measured on the cysts themselves – not even when concentrating large amounts of cysts. Therefore, these analyses cannot represent the true dinocyst age, and certainly not an age distribution as the individual cysts in our data represent. We feel that incorporating any

sample 'age' correction for now would mainly result in erroneous corrections and hence we prefer to illustrate the range in CO_2 and $\delta^{13}C$ corrections (a maximum error range), which we elaborate on in our revised manuscript (paragraph 3.2.2, lines 211-239).

Specifically, the maximum influence of the Suess-effect is assessed by constructing calibrations with uncorrected data for each of the parameters (measured $\delta^{13}C$ and CO_2). These are compared these to the original calibration (see Figure below, which we added as Supplementary Figure 1). We stress the difference between these calibrations is a worst-case scenario (i.e. maximum offset).



The above figure has been added to the supplementary information. The figure shows the offset between measured and adjusted pCO_2 and $\delta^{13}C$ ($\varepsilon p\text{-}CO_2$) values. Open symbols indicate measured $\delta^{13}C$, closed symbols represent data after eliminating small signals (<0.2 Vs) and outliers. Blue dots represent measured CO_2 values and red dots indicate the CO_2 around 1850 CE. For each dataset a simple linear regression, weighted to the number of measurements, is plotted. Dashed lines are based on measured $\delta^{13}C$, solid lines are from final $\delta^{13}C$ data. The regression plotted as a red solid line is used for Figure 5A.

What do you think is the background blank source? Is it from atmospheric aerosols that adhere to all surfaces regardless of precautions or is it from within the nickel plate? (Does the nickel plate show scoring from the laser?). Regardless, the approach to signal size to noise, considerations of the blank and other corrections seem reasonable. These consideration are important not only for your study and approach but for the future potential of this kind of analysis for sample return from Mars and elsewhere.

Author response:

We have expanded our methods section with a few statements covering these aspects (line 175-189):

The origin of the blank source is currently unknown and proved difficult to constrain. When setting up the system, we used a liquid N_2 cooled trap to pre-concentrate CO_2, before releasing it to the IRMS. With that system inherently the 'blank' $\delta^{13}C$ was also much larger, as it also concentrates the blank signal, which is why we returned to the current true continuous flow system. However, even with that concentrated blank it remained difficult to constrain its C-isotopic value and therefore also the opportunity to confidently identify the source (see also e.g., Van Roij et al., 2016). However, the trapping experiment provides some useful constraints on the potential size of the background contamination, which has been added to the revised manuscript (lines 185-189).

We considered several potential sources: (1) atmospheric CO_2 (air or particles that come into the system when opening the sample cell), (2) residual material from earlier measurements (wall sorption) and (3) possibly a minor amount of additional C from the water the cysts are isolated from. We assume that the first source are atmospheric particles the reviewer hints at. The nickel plate is not ablated with the low energy densities used and testing with clean plates (before any sample is added) shows the nickel plate itself does not add to the C signal, Similarly, repeated test with water droplets being added showed this to be a minor carbon source (< 10% of total blank). Etching or sample water also would not influence the PE standard as the ablation does not fully penetrate the plastic and no water is used for preparing the PE standard. In addition some micro leakage of the system at connections for the GC and/or the ovens could also add to the blank signal. Because it is currently impossible to constrain the source we here prefer to refrain from speculation on potential blank sources, other than stated above.

Regardless of the source, the combined contribution of these factors proved to be minor and stable, and hence we were able to correct for it (as can be seen in Fig. 2).

Line 280: From this discussion I think I favor your argument that intercyst variability reflects individual differences. One can envision that individual cells or cysts have significantly different 13C values owing to the randomness of cellular growth, changes in microenvironments of growth that also affect DIC and CO2 13C. Add in the time averaging from core top sample collection it is not a surprise that you see large variance. In fact, I would be worried if you did not. Your suggestion of controlling for size, as much as one is able, is a good idea.

Author response:

We thank the reviewer for their view – indeed this is also our preferred scenario for explaining the intra-sample variability. We have included the reviewer's points on potential for ^{13}C-impact of cell-microenvironments and growth-induced randomness to the $\delta^{13}C$ of the cyst in our revised manuscript (lines 318-319).

Line 280: For standards have you considered dissolving a standard material like caffeine in water and allowing it to dry onto a surface and analyzing that (you could spray it or something). At the very least here you could assume that the starting composition is isotopically uniform. I supposed 13C differences could arise from the drying process, but it may be better than PEF.

Author response:

We have been in search for a sufficiently homogenous standard with similar ablation and material characteristics as the polyethylene plastic currently used. We prefer a more or less similar material as potential differences in ablation characteristics could interfere with our method of standard bracketing in which we compare the signal of the ablated dinocysts with that of the standard. A standard material with similar ablation characteristics is also preferable from an operational perspective. For example, if the sample plate is covered entirely with a standard material (as indeed could be done with spray or by submerging), it becomes difficult to insert the sample without contaminating the system. Moreover, we prefer a standard with a relatively low vapor pressure as a somewhat volatile standard material in the sample holder could increase the blank signal. However, we also acknowledge that solid standard material (e.g., a film or foil) is inherently hampered by inhomogeneities on the scale required (~80 µm spot size). We are involved in the constant quest for better and improved standards (see e.g Boer et al., 2022* in which we developed a new standard using micro-milled powders), but this is challenging for solid organics. When cooling or drying (crystallisation) organic substances (e.g., glycerine, various corn-starch based products, monosodium glutamate) we observed the solid to structurally differentiate and no homogeneity could be reached. Work on a new standard continues (including the suggestion given by the reviewer using spraying) and, once successful, we will implement such a standard in our methods and report on it.

*Boer, W.; Nordstad, S.; Weber, M.; Mertz-Kraus, R.; Hönisch, B.; Bijma, J.; Raitzsch, M.; Wilhelms-Dick, D.; Foster, G.L.; Goring-Harford, H.; Nürnberg, D.; Hauff, F.; Kuhnert, H.; Lugli, F.; Spero, H.; Rosner, M.; van Gaever, P.; de Nooijer, L.J.; Reichart, G.-J. (2022). New calcium carbonate nano-particulate pressed powder pellet (NFHS-2-NP) for LA-ICP-OES, LA-(MC)-ICP-MS and µXRF. *Geostand. Geoanal. Res. 46(3)*: 411-432.

Line 300: Have you investigated the compositional differences between cyst and motile cells? I am familiar with the references you report on this issue but what specifically are the differences? What proportion of the carbon from the cell transferred to the cyst? Is this known?

Author response:

These are important outstanding questions and subject of currently running as well as planned work regarding cell compartment derivation of cyst molecules using LC-IRMS, cyst production – excystment experiments to assess cell to cyst fractionation. In short, for this we need to compare the core-top cysts to cultured motile cells to ensure that cells and cysts can be related one on one, which is a line of research in itself.