Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2017-558-RC2, 2018

Biogeosciences Discussions

Interactive comment on "Iron triggers colony formation in *Phaeocystis antarctica*: connecting molecular mechanisms with iron biogeochemistry" *by* Sara J. Bender et al.

Anonymous Referee #2

Received and published: 8 June 2018

1. OVERALL MANUSCRIPT REVIEW AND MINOR CORRECTION REVISION:

Bender et al. report the effect of iron on colony formation in P. antarctica through a series of "-omics" scale measurements in the laboratory and directly from the field (in the Ross Sea). As the authors mention, this area of research is garnering increasing interest among a multidisciplinary audience- and rightfully so. The work is a discovery-driven report describing proteome/transcriptome dynamics in response to iron in P. antarctica. The authors have thoroughly described their approaches in the Supporting Information and should be commended for their transparency in data processing, as well as making their data available in multiple public repositories.

I DO recommend acceptance of this manuscript with MINOR revisions, with 3 reservations that I hope the authors consider:

A. The volume of work could greatly benefit from a visual scheme (or diagram or flow chart) in the Supplementary Information section (at a minimum). This will help the reader discern laboratory measurements vs. those collected from the field, while also enabling the depiction of the experimental replication scheme (for both transcriptomics experiments and LC-MS/MS experimental replicates). While the presentation of the data are in order as written, it was at times difficult for me to discern what experiments were conducted where, and how those specific measurements were being placed into a potential mechanism.

As presented, the work encompasses a very large undertaking, spanning approximately 15 years. Samples were collected in the Ross Sea in December 2005, mass spectrometry-based proteome measurements (LC-MS/MS) were acquired in late 2013 to early 2014, and transcriptomic/proteomic data analysis performed between 2014 to the time of manuscript submission in late 2017. One figure to show the body of work- particularly in a Supplementary Figure- would do the amount of work reported here justice.

B. Please "soften" some of the semantics in the discussion of the results. There are many instances itemized below that I think could help the audience interpret the measurements more accuratelyâĂŤparticularly the proteome measurements.

For instance, line 31 in the Abstract: ". . .327 and 436 proteins significantly different between low and high iron strains. . ." Analytically, this is just what was able to be measured in the complex matrix of the samples in question, so perhaps rephrasing using the specific words "detected" or "measured" are more accurate.

Reporting the results to a multidisciplinary audience is challenging and many different fields report measurements differently, but presenting the measurements in context to what they are could- in general across the paper- assist with what the authors are truly

describing as support for a more focused hypothesis of a mechanism (which would be presented in a subsequent paper, for example, as this work is very much discovery driven).

C. A citation should be added on page 21 (lines 476-477) where the reference to "...prokaryotic components of natural communities..." is cited, first by Ram et al. 2005. An Analytical Chemistry publication in early 2005 demonstrated the proof-of-principle that was used by Ram et al. in a significantly more complex sample matrix consisting of both eukaryotic and prokaryotic proteins.

VerBerkmoes NC, Hervey WJ, Shah M, Land M, Hauser L, Larimer FW, Van Berkel GJ, Goeringer DE. Anal. Chem., 2005, 77 (3), pp 923–932. DOI: 10.1021/ac049127n https://pubs.acs.org/doi/abs/10.1021/ac049127n

The paper above showed that commercial mass spectrometry instrumentation and software were capable of measuring complex metaproteomes and reported the limit of detection for E. coli across a range of protein extracts from multiple model organisms (Arabidopsis, the yeast, and at least 3 other microbes). This work also presented peptide-spectrum matching (as used in Bender et al.) in a complex system.

2. SUGGESTED MINOR REVISIONS FOR CORRECTION:

Line 34" "...dynamic in range..." - is this referring to dynamic range of plastocyanin or the complexity of the sample matrix (analytically). I think that this could be worded more clearly for the reader.

Line 40: EF domains - the acronym for "EF" was not presented in multiple descriptions of these proteins- that needs to be made clearly to the reader.

Line 70: One of several instances of a spaced needing to be inserted between the end of a sentence and/or a reference.

Line 149: Microsoft Office's "auto-incorrect" feature changed the abbreviation for acetonitrile (ACN) to "CAN." The abbreviation should be changed to ACN, MeCN, or an-

other abbreviation of the organic content in solvent B.

Lines 156-157: "Normalized spectral counts were generated from Scaffold..."

Line 170: please make the word "transcriptome" plural; the methods describe processing 6 of them.

Line 176: Please include the version of CLCbio (and the vendor, Qiagen) in the Supporting Information.

Line 177: Please see comment above about line 70; a space needs to be inserted after "FragGeneScan" and the reference.

Line 213: Please see comments above about line 70 & 177: a space is needed between "...(Thermo Scientific)..." and the reference to Eng 1994.

Lines 268-269: Please refer to the semantics (above) on how "...436 proteins were identified as significantly different in relative protein abundance between "low" and "high" Fe." In my opinion, they were detected or observed at different iron concentrations and the amount of iron MAY modulate expression of these proteins.

Line 281: please spell out what the EF-hand domain is again; please see comment about line 40.

Lines 282-285: How many of the proteins of unknown function were previously listed as hypothetical proteins (e.g. potential protein-coding ORFs that have not yet been identified at the protein level)? That number is also a potentially interesting result from the proteome measurements.

Lines 328-330: With respect to drawing a hypothesis from the measurements, it appears that the word "expressed" is missing before constitutive(ly)- unless I misunderstand the conclusions that are being drawn.

Line 347: Please insert the word "potentially" before "related" - this inference is based upon sequence homology for functional inference

Line 383: Please include the word “potentially” before “contribute” - the ECM-related adhesion of cell (potential function of the proteins detected) is based upon sequence homology for functional inference

Line 409: Please see the previous comments above on semantics about the verb “produced” - the similar protein families that are being referred to were measured (and then filtered, etc., etc.) before making this conclusion.

Line 419: Please add a space before the word “Additionally. . .” - the reference preceding this word needs a space.

Line 420: Please spell out the “EF” domain proteins that are being described here. By this point in the paper, I am presuming that this is the canonical domain that is in calmodulin; but, are the proteins that were measured EF-containing or EF-like, etc. Moving this up in the Abstract for the reader would make things a bit clearer.

Lines 451-452: Semantics of “relative newness” - perhaps novelty of omics application(s)? Relatively new maybe? Or perhaps not include the word “relative” at all, since spectral abundance measurements are included and discussed in depth.

Lines 476-477: Please see the request to cite VerBerkmoes et al. Analytical Chemistry, 2005, as this reference provided a benchmark (proof-of-principle) for the Ram et al. paper presented- in a more complex biological system at that. The Analytical Chemistry reference also shows that detection of eukaryotic proteins in a complex matrix by peptide-spectrum matching was plausible on commercial instruments and software $\sim$13 years ago.

Line 517: there is an inadvertent space between “colonies .” that needs to be deleted.

Lines 592-593: Could this sentence please be re-worded for clarity? Again, this goes to semantics, in my opinion- perhaps the measurements “Yield preliminary insight into structural remodeling process(es). . .” going into line 594. As written, the “provides a first window into the complex. . .” could be rephrased to help the reader know what is

being emphasized from the measurements.

Line 613 (Acknowledgements): Please insert a space to the person that the authors are indebted to & the NSF award number.

Line 629 (Table 1 legend): As written, the table appears to include a number of differentially expressed [modulated] transcripts; however, there are 3 columns in Table 1 and the number of transcripts appear to not be listed. I am not certain if this is a conversion error (from Word into a PDF, etc.), but the way that I am reading it it appears that values are missing for differentially modulated transcripts.

Line 663 (Figure 3 legend): “. . .PCA of the full proteomes for each condition. . .” - I am not able to discern if “full proteome” refers to the proteome profile that has been measured, all of the entries in the database(s) used, or something else. I believe that what is being listed is a PCA plot of the proteins measured, but “full proteomes” is semantically misleading.

Line 729: Please capitalize “ms1” as written in the Figure 10 legend. Given the amount of time that went into making these figures merits this small change.

Supplementary Information: For a multi-omits study, the Supplementary Information provide adequate information for a team of researchers to determine how and what were measured. However, as strenuously suggested above, a visual depiction or representation of the ALL of the measurements performed would greatly benefit the reader- especially since this work spans 2 omics measurements, multiple replicates, and BOTH samples measured from the lab & field.

With respect to extensive description of peptide-spectral matching in metaproteomes in the Supplementary Information (SI, pages 11-12), the authors have taken considerable time justifying an approach that is commonplace among the studies of microbiomes- particularly those from limited environmental origin. While novel applications are in active development to improve the state of proteome informatics, the

cross-correlation scores (and appropriate delta CN values) listed by the authors have been used extensively over the past 16 years (Tabb et al. J. Proteome Res. 2002. https://pubs.acs.org/doi/abs/10.1021/pr015504q) and the justification, as written, for the peptide-spectrum matching approach is widely used in metaproteomics.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-558, 2018.