March 31, 2016

Dear Editor;

We resubmit the manuscript “Compositional analysis: a valid approach to analyze microbiome high throughput sequencing data” formerly submitted as Manuscript ID cjm-2015-0821. This manuscript was invited by Dr. Chris Yost following a successful workshop on Microbiome Data analysis at last years annual meeting of the Canadian Society of Microbiology. The focus of the workshop was exploration of data and compositional data analysis. We were able to introduce the bare bones of this in the workshop, and this manuscript contains a more complete and comprehensive treatment of the issues and approach.

We appreciate the reviewer’s constructive criticism and submit a revision that includes all suggestions and some additional re-writing. Changes made are identified in a point-by-point response attached to this letter.

Thank you for considering this revision.

Greg Gloor,

Professor of Biochemistry

U. Western Ontario

Point by point response to reviews:

Please not that the line numbers given are the line numbers in the revised manuscript with track changes on.

Reviewer 1:

General

The manuscript by Gloor et al. outlines the importance of the use of compositional data analysis tools for metagenome analysis. The authors use two case studies to illustrate how the application of these tools can impact the interpretation of data from published studies. This is a resubmission of a previous version.

I must commend the authors on the substantive editing and expansion of several sections, which has improved the manuscript markedly. In addition, I see tremendous value in the authors’ message and believe that this manuscript will be of use to many who will be searching for appropriate methods for metagenomic analysis.

I still struggled with the flow of the manuscript, and I would encourage the authors to try to improve the organization. Here is one suggestion that may help:

a) Move lines 67-78 to line 268, and change the somewhat awkward heading “Results and Use Cases” to “Case Studies” or something similar.

b) Move lines 79-90 to line 194

These changes have been made, and substantial re-writing was done. Lines 79-90 can be found at lines 320-334. Lines67-78 can be found integrated at lines 428-430. We believe that this has substantially aided the flow of the document and reduced redundant information.

For the first case study, line 276 – “First worked example”, I would suggest “Case 1 – Vaginal Microbiome” or something similar. Also, it might be worth providing an overview of the study, and describing the publication's primary approaches and the findings before the authors launch into their approach.

The wording of all headings was completely reworked to increase readability. See line 437.

The second case study is 16 lines, whereas the first is 6 pages. Can the second case study be expanded? As suggested above, it might be worth providing an overview of the study and the major findings before launching into the limitations.

We have nearly doubled the length of the description of the second case study to 58 lines of text. The first case study is necessarily long because it introduces all of the concepts as applied to the analysis. At least 3 pages were devoted to explaining how to interpret the biplot in Figure 3, it would be completely redundant to recapitulate this information and the second analysis can be shorter because the reader can simply translate the concepts. We have also included additional information in the second case study that will help the reader decide whether it is worth proceeding after the initial exploratory analysis.

Table 1 – needs to be able to stand on its own. All of the abbreviations in the column headings should be self-explanatory or defined in footnotes (or in the text at minimum).

Definitions for the abbreviations are now included below the table

Some other small edits (not comprehensive) – the authors need to proof-read this manuscript carefully:

Line 35 – “clusters of microbes” - microbial consortia?

Line 56 – culturing techniques

Line 70 – delete “on”

Line 74 – “dangers of improper understanding of data analysis in this area” – rephrase (delete “understanding of”?)

Line 79 – “The goals of analysis: Any analysis of a microbiome dataset usually has three

80 major goals:” – pick one, although I am not sure I agree that these are really goals as they are presently phrased

Line 81 – “Does the data” - Do the data (fix throughout)

Line 91 – Awkward heading

Line 98 – “can affect the results” - can have.

Line 182 – “per taxa” – per taxon

Line 197 – Capitalize “is” in the subheading

Line 312 – “L. iners”, “L. crispatus” should be defined on first use, and it is important that the reader can distinguish between species of Leptotrichia and Lactobacillus when they are abbreviated.

Line 453 - Discussion and Summary – awkward heading.

Line 469 – programming? Not sure what this sentence means.

Line 489 – move “by Pawlowsky-….2015)” to after “comprehensive book” in Line 488

Line 498 – “making claims that are later disproven” – suggest rephrase or delete

Line 505 – “filed”?

Line 507 – “there”?

Line 511 – is the exclamation mark necessary?

All these minor edits have been completed, and additional changes have been made to improve readability, and the manuscript was given a complete proofreading. In particular sub-headings have been removed and replaced by a narrative introduction, and additional references have been added that support key points.

Figures are very nice. Quite the improvement. Please italicize all latin binomials in all figures, and I would suggest avoiding using fluorescent baby blue, which is barely visible (biplot, pg 32/91).

Thank you, we have changed the cyan to a purple tint, and have italicized genus and species names in the figures.

Reviewer 2:

All minor edits have been made.

Substantial comments (lines refer to the second revision marked up version) :

We have added ‘DNA’ to all instances of high throughput sequencing

The url on line 69, has been changed to a stable DOI at Nature, and the sentence rephrased

A citation to R has been added (line 79)

All sub-headings (underlined) have been reworded as narrative

The single citation to pearson has been expanded and reworded to acknowledge that this is a known problem in the microbiome field (lines 214-217)

The document has been reworded so that it is clear that the citations requested on line 196, are included in the discussion below (line 319). Including all the citations at this point would be unnecessarily cumbersome, and would not have any context. All extant tools are described and cited in this section from lines 316-426.

The new header for the use cases is now “Application of CoDa to Two Case Studies” on line 426

The colloquialism on line 304 has been removed

Thank you for catching the incorrect binomial name on line 435

We have expanded the description of the subsequent analysis substantially (R1 line 444)

(two minor comments,changes made)

The awkward paragraph in the discussion has been amended

Legend to Figure 5 has been revised