Reviewer’s report

Title: Composition and Diversity of the Subgingival Microbiome and its Relationship with Age in Postmenopausal Women: an Epidemiologic Investigation

Version: 0 Date: 27 Jun 2019

Reviewer: Egija Zaura

Reviewer's report:

LaMonte et al: Composition and Diversity of the Subgingival Microbiome and its Relationship with Age in Postmenopausal Women: an Epidemiologic Investigation

This is a crossectional study on subgingival microbiome of a large group of postmenopausal women (N=1219). The authors aimed to assess the extent in variation in microbiome composition and diversity with age. They conclude that they have identified associations between 12 bacterial species and age.

Strength of the study: obviously, its sample size and sample type. Not so many studies of this scale use subgingival plaque samples, and as authors state, in this population only small-scale, usually targeted studies are performed.

The sequences and subject metadata should be deposited in a publicly-available database such as SRA of NCBI. The access link should be provided in the text.

Weaknesses:

The authors must have hypothesized that biological age will be an important parameter, while age alone does not determine the microbiome, certainly not in adults. It is the senescence of tissues and functions, side effects of medication (such as lowered salivary flow), number of teeth, presence of dentures, oral hygiene habits, smoking, snacking, history and presence of caries, number of bleeding sites, and so on.

The authors ignore or are not aware of the possibilities of microbial profile data analyses, and perform traditional univariate statistics on individual OTUs (in their case, centered log2 ratio transformed OTUs). Multivariate analyses including the factors listed above should have been performed, to elucidate the relative contribution of age alone. Current approach is too simplistic and a pity to see the potential of the data not being used.

The presentation of the results: the huge and endless tables are incomprehensible for the reader and will not look better when published. These should be provided as excel files in the supplement, and only relevant (significant) data presented in the main text.
The title and introduction mention Diversity, while authors report that no difference was found and choose not to show the data. This information should be provided to the reader, even though the authors did not find any differences. Relating the microbial diversity to the other parameters mentioned above should be performed.

How was the diversity calculated? This is not mentioned in the methods part.

Based on the results (line 237 and onwards), the amplicons have not been mixed equimolarly. Extremely large variation in sequencing depth among the samples - minimum 3034 reads, maximum - over 1 million reads is observed. Authors state that sequencing depth increased with age. This is very alarming finding, and suggests sequencing bias in the data. Authors should test the read statistics for each sequencing run, and assess if they have randomized the samples of different ages among the runs. The information on this should be provided.

Huge variation (over 300-fold difference) in sequencing depth needs to be corrected for. The diversity analyses (both alpha and beta) are highly biased if they are performed on this dataset without correcting for the read depth differences. There are several ways for correcting this. The straight forward way is a random subsampling of all samples to an equal depth. The subsampling depth should be chosen based on plateau-phase of the OTU-rarefaction plots.

Throughout the text, terms like bacterial species, bacteria, in respect to which taxa were found or associated to something (eg, as in line 333), should be replaced with OTUs classified as… at least when these are mentioned for the first time. This is a very common mistake, and unfortunately is copied from paper to paper, since more and more researchers are not aware of the limitations of the methodology. The methodology used here does not allow bacterial identification, certainly not at the species level. Instead, it creates clusters with short 16S sequences similar at 97% level. Each OTU can and will contain different unique sequences (within 97% similarity), and the OTU-names are assigned to single sequence from the cluster by greedy algorithms, thus the assigned species names are just a representation but not exclusive to what is in the rest of the cluster (OTU).

The OTUs that did not match any of HOMD sequences should be assigned taxonomy using a general microbial database eg, SILVA.

Table 1 should include additional oral health parameters (caries, bleeding on probing, salivary flow, medication use besides hormonal therapy, etc), and the differences among the age-groups should be tested and presented in the table.

Figure 1: the data should be presented as boxplots, to provide within-group variation.

Figure 2: PCA on what? How was the dataset transformed for this? PCA assumes that data is normally distributed, which it is not. Therefore, the data needs to be transformed before that. The % of explained variance per component should be presented.

Line 274 and elsewhere in the text: Strep. Should be Streptococcus, and on repeat: S.
Line 294, 297 - what does the 'asterisk' mean?

Tables 2-6 - too long for the main text.


**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Unable to assess

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Unable to assess

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal