

Email

| | |
|-----------------------------|---|
| Manuscript # | mSystems00144-16 |
| Title | Health and disease imprinted in the time variability of the human microbiome. |
| Corresponding Author | Prof. Andrés Moya (Universidad de Valencia - Instituto Cavanilles) |
| Date: | 2016-11-14 18:48:06 |
| Last Sent: | 2016-11-14 18:48:06 |
| Created By: | Redacted |
| From: | gilbertjack@uchicago.edu |
| To: | andres.moya@uv.es |
| Subject: | mSystems00144-16 Decision Letter |
| | <p>November 14, 2016</p> <p>Prof. Andrés Moya Universidad de Valencia - Instituto Cavanilles Instituto Cavanilles de Biodiversidad y Biología Evolutiva c/ Catedrático José Beltrán 2 Valencia 46071 Spain</p> <p>Re: mSystems00144-16 (Health and disease imprinted in the time variability of the human microbiome.)</p> <p>Dear Prof. Andrés Moya:</p> <p>While the article represents an interesting dataset and a useful statistical analysis, there are major concerns about the interpretation of the statistics and the english language.</p> <p>I have selected Minor Modifications - BUT you must have this paper THOROUGHLY reviewed for English language edits prior to resubmission, we recommend you use a service to do this.</p> <p>Also, the statistical concerns of review 2 are legitimate, and FAILURE to address these concerns adequately will result in the paper being rejected.</p> <p>Below you will find the comments of the reviewers.</p> <p>To submit your modified manuscript, log onto the eJP submission site at Link Not Available . If you cannot remember your password, click the "Can't remember your password?" link and follow the instructions on the screen. Go to Author Tasks and click the appropriate manuscript title to begin the resubmission process. The information that you entered when you first submitted the paper will be displayed. Please update the information as necessary. Provide (1) point-by-point responses to the issues raised by the reviewers as file type "Response to Reviewers," not in your cover letter, and (2) a PDF file that indicates the changes from the original submission (by highlighting or underlining the changes) as file type "Marked Up Manuscript - For Review Only."</p> <p>Please return the manuscript within 60 days; if you cannot complete the modification within this time period, please contact me. If you do not wish to modify the manuscript and prefer to submit it to another journal, please notify me of your decision immediately so that the manuscript may be formally withdrawn from consideration by mSystems.</p> <p>To avoid unnecessary delay in publication should your modified manuscript be accepted, it is important that all elements you upload meet the technical requirements for production. I strongly recommend that you check your digital images using the Rapid Inspector tool at http://rapidinspector.cadmus.com/RapidInspector/zmw/.</p> <p>For a corresponding author who is an active member of ASM at the Contributing or Premium level, the mSystems article processing charge (APC) is \$2,250 for Research Articles, Resource Reports, and Methods & Protocols and \$1,125 for the shorter Observation and Opinions & Hypotheses formats. For a corresponding author who is a nonmember or Supporting member, the mSystems APC is \$3,000 for Research Articles, Resource Reports, and Methods & Protocols and \$1,690 for the shorter Observation and Opinions & Hypotheses formats. Nonmember corresponding authors or Supporting members may <a</p> |

[upgrade or join ASM](http://www.asmscience.org/content/membership/all) to obtain discounts on APCs. Minireviews, Commentaries, Perspectives, Editorials, and Letters to the Editor are not subject to APCs. **APCs are subject to change without notice.**

Authors are charged a flat fee for posting supplemental material as an adjunct to their published article. For a corresponding author who is an active member of ASM at the Contributing or Premium level, the supplemental material fee is \$200. For a nonmember or Supporting member corresponding author, the supplemental material fee is \$300. Nonmember corresponding authors or Supporting members may [upgrade or join ASM](http://www.asmscience.org/content/membership/all) to obtain discounts on supplemental material fees. **Supplemental material fees are subject to change without notice.**

Thank you for submitting your paper to mSystems.

Sincerely,

Jack Gilbert

Editor, mSystems

Journals Department
American Society for Microbiology
1752 N St., NW
Washington, DC 20036
E-mail: peerreview@asmusa.org
Phone: 1-202-942-9338
Fax: 1-202-942-9355

Reviewer comments:

Reviewer #1 (Comments for the Author):

This is a very interesting paper exploring temporal variability of the human gut microbiota in the context of Taylor's Law. The authors explore temporal variability of genera in the context of host health and disease, and based on 16S and shotgun sequencing profiles. My detailed comments follow.

The manuscript needs editing for English language. These issues are severe, and make parts of the manuscript difficult to understand. I won't provide detail, except in cases where I notice a term that seems to be misused.

Figure 1: Axis labels should indicate that these are mean/std relative abundances. Clarify the meaning of the error bars in the figure legend.

Line 56: expand SMS acronym as this is the first use in the main text. It doesn't seem to be used frequently in the text, so it might not be worth using at all.

Line 98: It would be very useful to have physical descriptions of beta and V at this stage - what do each of these values tell us about the microbial communities?

Figure 2: The points in this plot should somehow indicate which represent the "healthy" and "unhealthy" samples from each study. As presented, it looks like you're labeling Kwashiorkor samples as healthy, since they're in the healthy zone. (I notice that these may only be the unhealthy samples in this plot. On line 115, the authors mention that these are the samples for "individuals whose gut microbiota is altered", but the Figure 2 legend says "all the data studied in this work". Please clarify this.)

Since you're trying to define a healthy zone in beta-V space, it's important to quantify your prediction accuracy. For example, if you were to leave some studies out (e.g., the Kwashiorkor study) and regenerate Figure 2 without it, how many of the unhealthy samples from the Kwashiorkor study would fall in your "healthy zone", and how many outside of the "healthy zone" (and the same questions for the healthy samples from that study). Without these types of comparisons, it's not possible to know whether the results presented here would extend to other data sets or whether the parameters of the healthy zone have been overfit to these data. I think these "leave-one-out" analyses should be performed on a per-study basis (not a per-sample basis) and added to this work.

Line 115: I don't think "whose gut microbiota is altered" is the correct phrase here - maybe "compromised" instead of "altered", as the latter suggests that the microbiota has been

Email

intentionally modified.

Figure 4: Color coding of RSI percentage column makes those values difficult to read (and I don't think they're color blind safe). Could this be represented with a heatmap-style coloring instead, and all text in black (so similar to the heatmap in the same figure)?

The authors illustrate RSI values for two samples in Figures 4 and 5. These should either be presented in the same figure to facilitate comparison. It would also be useful to show the RSI percentage columns for other healthy and unhealthy individuals. As it stands, it's hard to see exactly what the difference is across these two vectors (but also see my comment above on the color coding of the RSI percentages).

Figures 4 and 5: The x-axis should be labeled better. These look like they might be sample identifiers, but it would be very helpful if it was made more clear that this axis represents time, and how much time passed between each pair of adjacent samples.

Line 192: "Both samples..." which samples are being referred to here? I don't see this quasi-periodic behavior in figure 7, which is what is most recently referred to. Is this in reference to figure 6?

Figure 8-9: The key and axis text is unreadable (too small). These could also be moved to a supplementary file if the number of figures needs to be reduced.

Line 231: "flips or jumps" should be replaced with a more specific description.

Line 254: "sensible" should be "sensitive"?

Line 384: "and 16S" - are the 16S referred to here those are included in the shotgun metagenomic data, or is this in reference to 16S amplicon data sets. If the former, this parenthetical statement should probably be dropped as it's confusing. If the latter, more detail is needed.

Line 389: "(0.5 TiB, that is, 512 gibibytes)" this clarification is not necessary.

Line 399: Where are these scripts? Are they necessary to reproduce these analyses? If so, they must be made publicly accessible (e.g., ideally in public revision control such as GitHub, or alternatively as a supplement to this manuscript).

Reviewer #2 (Comments for the Author):

The authors present a manuscript describing dynamical features of the human gut microbiome in healthy and diseased people (97 time series of varying lengths and densities). They find that relative abundance trajectories for individual bacteria follow Taylor's Law (i.e. power law scaling between mean and variance). They fit a stochastic differential equation that includes power-law terms to many gut time series obtained from the literature. The overall conclusions are that power laws can be fit to microbial time series and community rank instability is higher in disease.

First, the manuscript is poorly written and needs English language editing.

Second, I don't think the Taylor's Law results are terribly useful. There is no clear interpretation for these power law phenomena. Prior work has shown that these patterns can be produced by sampling processes from a skewed distribution (e.g. OTU abundance profiles). If statistical artifacts are driving these patterns, I'm not sure how biologically useful they are. Also, I believe Taylor's Law is usually in the form: $\text{var}(x) = \alpha * \text{mean}(x)^\beta$. However, you are relating the standard deviation and the mean, not the variance in the mean. This is why your beta term is less than one (beta is usually reported as being between 1 and 2 in most real-world data, with beta = 1.0 being a Poisson process - but this is only true if you are looking at $\text{var}(x)$).

Third, I'm not sure how useful the Langevin model fitting is. I know you're using it to get your 'fitness' (F) and 'amplitude of fluctuations' (V). However, you can probably obtain the same rolling window V plot by looking at the overall community variance along similar rolling windows (i.e. with a very simple calculation, rather than going to the trouble of fitting a stochastic model).

Fourth, I found the rank-instability stuff interesting. To me, this was the main biological result of the paper. I think the manuscript would be greatly improved if the focus was shifted to this rank instability and the presence of these rank stability 'islands'. You could dig more into the taxonomic identities of these organisms, and try to connect their dynamics to what we already know in the literature.

Overall, you have gathered a large number of time series (97). There is a lot of potential in these data, but I don't think the current approach with Taylor's Law and model fitting is the best way to go.

Specific Comments:

If you're going to fit Taylor's Law, spend more time discussing what the parameters mean biologically and how your data compare to data from other systems.

You don't properly define the F parameter, that you call 'fitness', in your stochastic differential equation. What is the biological intuition behind this parameter? What do you mean by 'fitness'?

Have you taken compositional effects into account in your Taylor's Law scaling? For example, do you expect sub-linear scaling to be due to damping of variance in abundant

Why didn't you include the time series from this paper?

<https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-7-r89>

In Fig. 8, should the y-axes be labeled 'F' instead of 'beta'?