Response to Reviewers (mSystems00144-16R1)

We thank the Reviewers not only for the thorough reading of our work, but also for all their comments, which have helped us to improve the manuscript significantly.

**Reviewer #1**

*Comments for the Author:*

This is a very interesting paper exploring temporal variability of the human gut microbiota in the content 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.

* The manuscript has been revised by a native English translator and changed accordingly.

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

* Now, Figure 1 axis labels indicate that they correspond to relative abundances. Figure caption specify which type of statistical error the error bars are indicating (it is the SEM).

Line 57: 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.

* Although SMS acronym is not very often used in the main text, it also appears in some figures and we still think it is worth to use it so that we have finally decided to keep it and to expand the SMS acronym in the Introduction.

Line 100: 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?

* We have added a physical description of the Taylor’s parameters in the suggested location in the Results section and we have related these parameters with the microbial communities as well as other systems (the latter as suggested by Reviewer #2).

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.)

* We have clarified in the Figure caption that we are just representing the microbiota of compromised subjects.

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.

* We agree on the relevance of testing the robustness of our conclusions. In our type of analysis, we can test the robustness of each independent study by randomly eliminating one of the healthy individuals and then, redefining the region of health and determining the position of unhealthy samples respect to the new region of health. When doing that, “unhealthy” samples are always outside of the new healthy region. On the other hand, a “leave-one-out” analyses does not change our plot at all. The standardization takes care of this. Due to the different systematics in each study, we define a health region for each study, standardize to mean zero and variance one and compute mean and variance of unhealthy with this standardization. Therefore, different studies are isolated as individuals from a given study do not affect the results on the unhealthy individuals of the other studies. We think this statistical approach is safer, as we avoid to combine data with very different systematic errors. We added a clarification in the text in Results, when the standardization is first mentioned.

Line 121: 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 intentionally modified. 

* We have changed “altered” for “compromised” as the microbiota has not been 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 faciliate 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).

* We have deeply changed the rank&RSI plots according to the suggestions in these comments. Now, both the colormap of the rank matrix and the colormap of the RSI percentage column not only are perceptually uniform colormaps, but also avoid color blindness (avoiding sets with both red and green). When the values are over a threshold, the font color is changed to keep the number readable on a dark background. Moreover, we have duplicated the number of those plots to show more cases by adding Supplementary Figures S1 and S2. For this, we have finally chosen between the studies included in this work the one with the smallest grid (daily) largest sampling published so far (suggested by the Reviewer #2). Finally, the rank&RSI figures now contain related variability plots that help with the analysis of the rank vectors over time (more details in “Rank stability index and variability” subsection of Material and Methods).

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.

* The x-axis of Figures 4 and 5 was poorly labelled. According to the suggestion, the labelling has been improved for these Figures so that it is clear now that the x-axis represents time. The new Supplementary Figures S1 and S2 are also labelled in this manner.

Line 195: "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?

* Yes, that is right, it was in reference to Figure 6. We have fixed this part in the text and now it refers only to both samples in Figure 6. Besides, the text writing has been improved to avoid further confusion.

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.

* The font size of both the key and the axis labels has been fixed for these Figures. They have been moved to supplementary files as suggested, so that  
  they are now Supplementary Figure S3 and S4, respectively.

Line 236: "flips or jumps" should be replaced with a more specific description.

* We have changed this part for a more specific description of the process.

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

* Totally right. We are sorry for the confusion. We have changed “sensible” for “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.

* We have deleted that part for it is confusing, as suggested.

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

* We have dropped that unnecessary clarification.

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).

* We have added a link to a GitHub repository where these scripts can be found. The reference link is in the related Material and Methods section.

**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.

* The manuscript has been revised by a native English translator and changed accordingly.

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: var(x) = alpha \* 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 var(x)).

* The Reviewer is right on pointing out the difficulties of a deep understanding of power laws in nature. Certainly, this is not the focus of our work. We highlighted that this ubiquitous behavior is also generic in metagenomic samples and characterized for the first time significantly different power law indexes and variability between healthy and unhealthy samples. Variability V is a global index of the system, which sets the potential of other studies like the rank instability. It is true that, originally, Taylor’s law was expressed as Variance versus mean value. We preferred to use the root square of variance as in the compilation given by Eisler, Bartos and Kertesz (Eisler *et al.* Europhys Lett 69:664–670, 2005). Finally, we have included in the manuscript the citation to the original article (H. F. Smith. J. Agric. Sci., 28:1–23, 1938) in the first reference to Taylor’s law.

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).

* The system of metagenomes is stochastic. Depending on the question addressed, the system can be characterized by a set of dominant variables, while the others can be treated as noise. The Langevin equation introduces (and assumes) dynamics in the data analysis. In the case we have considered, the differential equation has an analytic solution. Introducing a dynamic opens questions to explore and understand, as for example, which is the route from a stable (healthy) state to a noisy (unhealthy) one and back to a new stable state. These studies require new experiments and analysis which better monitor the main genomes that are governing these paths in the phase space of stable states.

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.

* In systems with large Variability V, like gut microbiome, more details of the system are within reach, in particular the study of rank-stability. Following Reviewer's advice, we have stressed the importance of the rank instability studies and we have correlated with other published studies. In particular, we have used the new dataset included in the manuscript, because it contains the largest daily time series published so far, which permitted us to more clearly illustrate the potential of studying the rank-instability. A full discussion is now added in results and discussion, and illustrated in Figures 4 and 5, and Supplementary Figures S1 and S2

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.

* We have made a more detailed discussion on the Taylor law parameters and have compared them with other systems in nature in page 5. β is the power law index and characterizes how the variability grows with abundances. If β is 1/2, the system behaves like a Poisson distribution. If β is 1, the system behaves as an exponential distribution (main changes occur at one time). Generally, metagenomes vary with time with β between these two universal classes. V represents the maximum variability attainable by a hypothetical dominant genus (with relative abundance close to 1). It is an important parameter that characterizes the type of system. If V is small, the ranking is stable such as, for example, the number of diagnoses of a particular disease recorded in Medicare during a month. If V is large, as it is the case of metagenomic samples, the ranking might be unstable, like the number of hourly page views of articles in Wikipedia (N. Blumm *et al.* Phys Rev Lett 109:128701, 2012) (Z. Eisler *et al.* Adv Phys 57:89–142, 2008).

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'?

* In our model, Fitness F captures the time scale the system needs to reach equilibrium (the size of variability V may or may not allow to reach it). F has dimensions of 1/time and corresponds, roughly, to the half-life of the system in the decay process to the stable state. In fact, it is exactly the half-life if β equals 1 and V is negligible. Accordingly, we have explained the meaning of F in Results.

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

* Yes, we observed this sub-linear scaling. In fact, in most cases, beta is smaller than one, which implies that relative changes are smaller for most abundant species/genera, i.e., their volatility is lower than that of least abundant ones. Such sub-linear behavior contributes to the stability of the high-ranked, as we do detect in a correlation between beta and the observed ranking stability. We need more and better data to significantly characterize this correlation. We have further clarified this point in the text.

Why didn't you include the time series from this paper? https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-7-r89

* We thank the Reviewer for pointing this reference out. The paper presents the smallest grid (daily) largest sampling published so far. In fact, we were already working on these datasets to search for modes (periods) in the samples. Following this Reviewer’s suggestion, we have added the data of this paper in our studies. Moreover, we have used it as a guide for the rank instability studies, because it leads to more precise results, driven by the larger time series. We have added Teresa Rubio in the author list, who has contributed to this piece of work.

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

* In this case, the answer is negative as there is no dynamical model in it. On the left panels, that figure just shows the Taylor’s parameters for different time series. On the right panels, the same data is plotted, but after the standardization. Finally, we should point out that such figure is now the Supplementary Figure S4, moved to supplementary material as suggested by Reviewer #1.