We thank the referees for the thorough reading of our manuscript and their comments, which have helped us to significantly improve it.

**Reviewer #1**   
  
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.

* Figure 1 axis now indicate that they correspond to relative abundances. Figure caption clarify which type of error bars we are using (SEM error bars)

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.

* We have explained 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 brief physical description of the Taylor parameters in the Results section.

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 only represent 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.

* Since the leave-one-out analysis is a powerful test to check the robustness of a method like the one we have proposed, it will not have any effect in the plot. The standardization method that we use here takes each study per separate and places all the “perturbed” microbiotas in reference to the “healthy” group of samples from that particular study, in std deviation units. The main purpose of this plot is to be able to represent all the datasets into one figure, for there is an intrinsic variability in each study due to the use of different laboratory techniques or even different bioinformatic analysis of the sequences, and in each study we got different ranges of *V* and *β* values that are not directly comparable.

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” for it fits better the idea we want to express.

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)? 

* Write something

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).   
  
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.

* Write something

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 correct. We have fixed this part in the text and now it refers only to both samples in Figure 6. Besides, it is now correctly structured 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 size of the axis text has been fixed for both figures.

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"?

* 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.

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

* We have deleted that part.

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 you can find these scripts. The link is in the page 17, in the Material and Methods section.

**Reviewer #2**

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 referee 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.  We also cite the original H. F. Smith, J. Agric. Sci. 28, 1 (1938), when referring 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 referee's advise, 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 largests 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 X and X

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 XXX. beta is the power law index and characterizes how the variability grows with abundances. If beta is 1/2, the system behaves like a Poisson distribution. If beta is 1, the system behaves as an exponential distribution (main changes occur at one time). Generally, metagenomes vary with time with beta 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, in the ranking is stable, as for example the number of diagnoses of a particular disease recorded in Medicare during a month. If V is large as in metagenomic samples, the ranking might be unstable like in the number of hourly page views of articles in Wikipedia.

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 units of time^-1 and corresponds, approximately, to the half-life of the system when decaying to the stable state. In fact, it is exactly the half-life if alpha is one and V is negligible. We have explained the meaning of F in XXX.

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 referee 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 referee’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 timeseries. We have added Teresa Rubio in the author list, who was in charge of this piece of work.

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

* No. There is no dynamical model in it. This figure is now Supplementary figure S4, as suggested by referee 1. There, we just plot the Taylor parameters for different time series on the left panels. We show the same data standardized on the right panels.