

1 Title:

2 Health and disease imprinted in the time variability  
3 of the human microbiome

4 Running title:

5 Microbiota, are you sick?

6 Jose Manuel Martí<sup>1,2,\*</sup>, Daniel Martínez-Martínez<sup>1,2,3,\*</sup>, Manuel Peña<sup>2</sup>, César Gracia<sup>1,2</sup>,  
7 Amparo Latorre<sup>1,3,4,5</sup>, Andrés Moya<sup>1,3,4,5</sup> & Carlos P. Garay<sup>1,2,#</sup>

8 <sup>1</sup>Institute for Integrative Systems Biology (I2SysBio), 46980, Spain.

9 <sup>2</sup>Instituto de Física Corpuscular, CSIC-UVEG, P.O. 22085, 46071, Valencia, Spain.

10 <sup>3</sup>FISABIO, Avda de Catalunya, 21, 46020, Valencia, Spain.

11 <sup>4</sup>Cavanilles Institute of Biodiversity and Evolutionary Biology, UVEG, 46980, Spain.

12 <sup>5</sup>CIBER en Epidemiología y Salud Pública (CIBEResp), Madrid, Spain

13 Words count for the Abstract section: 154 of 250 max

14 Words count for the Importance section: 105 of 150 max

15 Words count for the rest of text: 3908 of 5000 max

16  

---

\* Equally contributed

# Corresponding author: penagaray@gmail.com

## Abstract

Animal microbiota (human included) plays an important role keeping healthy the physiological status of the host. Increasing research activity is dedicated to understand how changes in composition and function of the microbiota are associated to disease or not. We analyze 16S rRNA and whole genome sequencing (WGS) published data from the gut microbiota of 97 individuals monitored in time. Temporal fluctuations in the microbial composition reveal significant differences due to factors such as dietary changes, antibiotic intake, age or disease. Here we show that a fluctuation scaling law describes the temporal changes in the gut microbiota. This law allows to estimate the temporal variability of the microbial population and quantitatively characterizes the path toward disease by a noise-induced phase transition. The estimation of the systemic parameters for follow-up studies may have clinical use and, more generally, applications in other fields where it is important to know if a given community is stable or not.

## Importance

Human microbiota is tightly associated to the health status of a person. Here we analyse the microbial composition of several subjects under different conditions, over a time span that ranges from days to months. Using the Langevin equation as the basis of our mathematical framework in order to evaluate microbial temporal stability, we prove that we are capable to distinguish stable from unstable microbiotas. This first step will help us to determine how microbiota temporal stability is related to the healthiness of the people, and it will allow the development of a more complete framework in order to deepen the knowledge of this complex system.

**Keywords**— microbiome, systems biology, ecological modeling, metagenomics, stability

## Introduction

The desire to understand the factors that influence human health and cause diseases has always been one of the major driving forces of biological research. As evidence of new concepts 'holobiont' and 'hologenome' is increasing each day (1, 2), research not only focus on the human physiology but also on the microbial population that surround ourselves. We are populated by a myriad of microorganisms that are interacting with us in several physiological processes such as metabolism of the bile acids (3), of the choline (4) or key-route metabolites as short-chain fatty acids (5, 6) which are also involved in immune system maturation (7, 8). Human microbiota has been suggested to be closely related to diseases like type 2 diabetes (9), cardiovascular disease (CVD) (10), irritable bowel syndrome (11), Crohn's disease (12), some affections as obesity (13, 14), malnutrition (15) among other multiple diseases (16). High throughput methods for microbial 16S ribosomal RNA gene and WGS have now begun to reveal the composition of archaeal, bacterial, fungal and viral communities located both, in and on the human body. Modern high-throughput sequencing and bioinformatics tools provide a powerful means of understanding how the human microbiome contributes to health and its potential as a target for therapeutic interventions (17). To define normal microbiota and how it's compositional changes can origin some diseases are important issues still in need for scientific answers (18, 19).

Biology has recently acquired new technological and conceptual tools to investigate, model and understand living organisms at the system level, thanks to the spectacular progress in quantitative techniques, large-scale measurement methods and the integration of experimental and computational approaches. In particular, Systems Biology has placed a great effort to unveil the general laws governing the complex behaviour of microbial communities (20–22). Microbiota can be approached under the light of ecological theory where we can find, for instance, general principles as the Taylor's law (23), which relates spatial or temporal variability of the population with its mean. This law, also known as fluctuation scale law, is

ubiquitous in the natural world and can be found in several systems as random walks (24), stock markets (25, 26), animal populations (23, 27, 28), gene expression (29), or in the human genome (30). Taylor's law has been applied to microbiota in a spatial way in the work of Zhang *et al.*, (2014) (31), where they show that this population tend to be in an aggregated way rather than in a random distribution. Despite its ubiquity, it has never been applied in follow-up studies from microbiota even that a great effort has been made to infer the community structure from a dynamical point of view (32–34)

Here we present the imprints of health status (healthy or disease) in macroscopic properties of microbiota, by studying its temporal variability. We have analyzed more than 35000 time series of taxa from the gut microbiome of 97 individuals obtained from publicly available high throughput sequencing data on different conditions: diseases, diets, obese status, antibiotic therapy and healthy individuals. Having seen that all cases follows Taylor's law, we use this empirical fact to model how the relative abundances of taxa evolves toward time thanks to the Langevin equation, in a similar way as it was applied recently by Blumm *et al.* (35). We use this mathematical framework to explore the temporal stability of the microbiota in different conditions in order to understand how this affects the healthy status of the subjects.

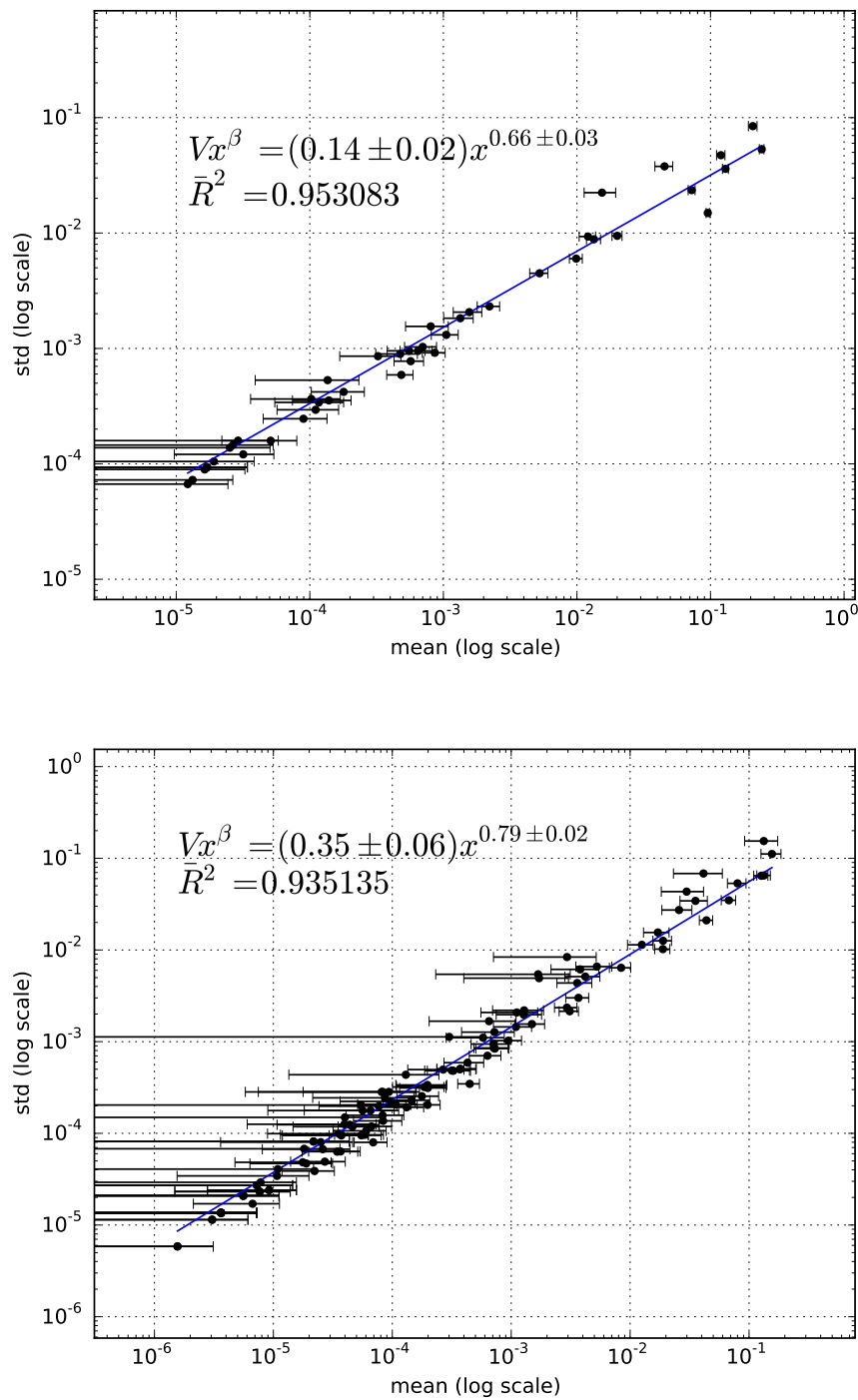
## Results

We have analysed the microbiome temporal variability to extract global properties of the system. As fluctuations in total counts are plagued by systematic errors we worked on temporal variability of relative abundances for each taxon. Our first finding was that, in all cases, changes in relative abundances of taxa follow a ubiquitous pattern known as the fluctuation scaling law (44) or Taylor's power law (23), i.e., microbiota of all detected taxa follows  $\sigma_i = V \cdot x_i^\beta$ , a power law dependence between mean relative abundance  $x_i$  and dispersion  $\sigma_i$ . The law seem to be ubiquitous, spanning even to six orders of magnitude in the observed relative abundances (see Figure 1).

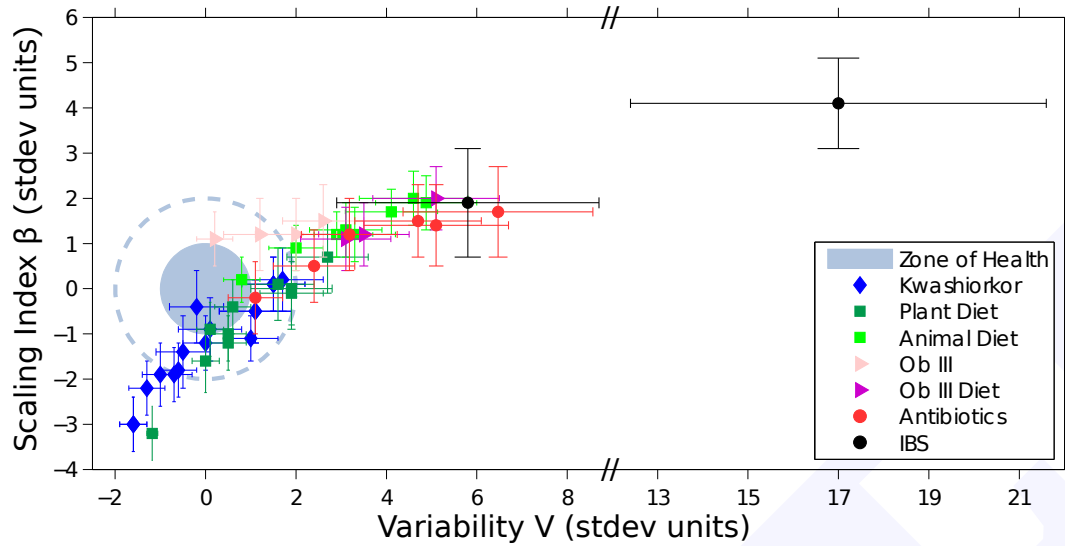
The power law (or scaling) index  $\beta$  and the variability  $V$  (hereafter Taylor parameters) appear to be correlated with the stability of the community and related with the health status of the host, which we consider the main finding exposed in this article (see Figure 2).

Taylor parameters describing the temporal variability of the gut microbiome in our sampled individuals are shown in Tables S1 to S6. Our results hint at an ubiquitous behaviour. On the first hand, the variability (which corresponds to the maximum amplitude of fluctuations) is large, which suggests resilient capacity of the microbiota. On the other hand, the scaling index is always smaller than one, which means that more abundant taxa are less volatile than less abundant ones. In addition, Taylor parameters for the microbiome of healthy individuals in different studies are compatible within estimated errors. This enables us to define an area in the Taylor parameter space that we called the *healthy zone*.

In order to jointly visualize and compare the results of individuals from different studies, their Taylor parameters have been standardized, where standardization means that each parameter is subtracted by the mean value and divided by the standard deviation of the group of healthy individuals for each study (for details of the procedure, please see Standardization subsection in Material and Methods). The healthy zone and the standardized Taylor param-



**Figure 1.** X-weighted power-law fits of the standard deviations versus the mean values for each bacterial genus monitored in time. We show the fit for samples from a healthy subject (top) and from a subject diagnosed with irritable bowel syndrome (bottom), studied in our lab (11). Taylor's power law seems to be ubiquitous, spanning to six orders of magnitude.



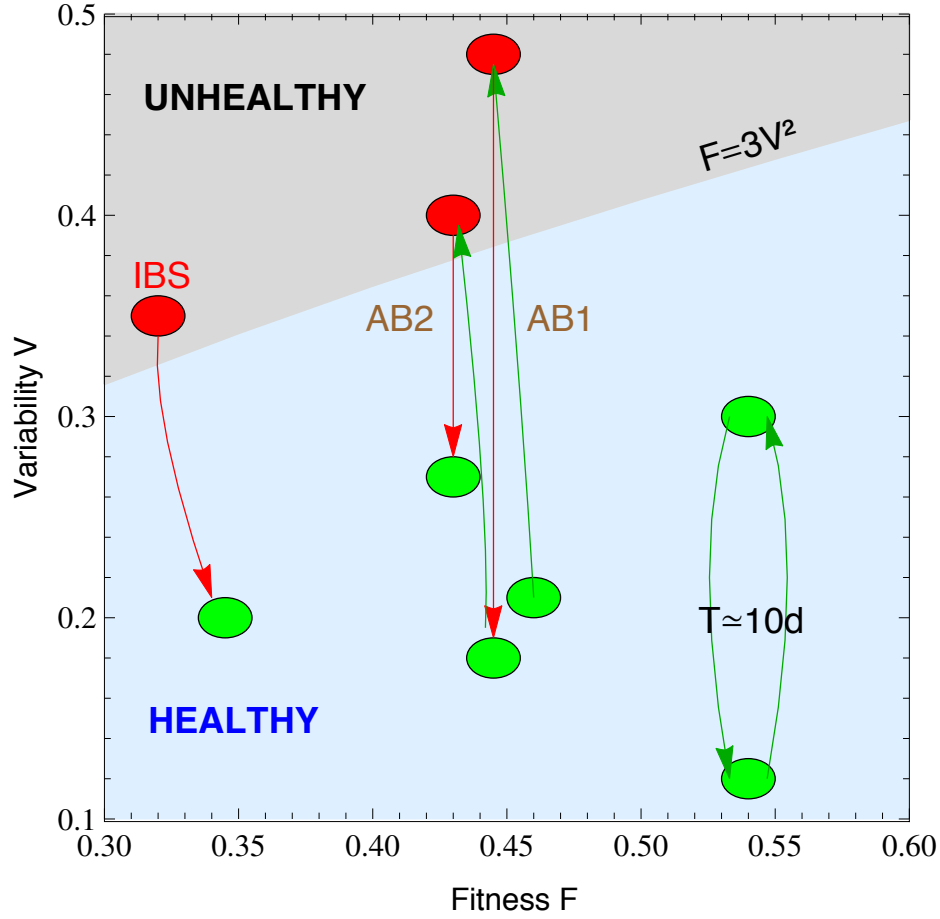
**Figure 2.** Taylor's law parameter space. We have compiled here all the data studied in this work. The coloured circle corresponds to 68% confidence level (CL) region of healthy individuals in the Taylor parameter space, while dashed line delimites the 98% CL region. Points with errors place each individual gut microbiome in the Taylor space. Note that the parameters have been standardized (stdev units) to the healthy group in each study for demonstrative and comparative purposes.

eters for individuals whose gut microbiota is threatened (i.e., suffering from kwashiorkor, altered diet, antibiotics or IBS) is shown in Figure 2. Children developing kwashiorkor show smaller variability than their healthy twins. A meat/fish-based diet increases the variability significantly when compared to a plant-based diet. All other cases presented increased variability, which is particularly severe, and statistically significant at more than 95% CL, for obese patients grade III on a diet, individuals taking antibiotics or IBS–diagnosed patients. A global property emerges from all worldwide data collected: Taylor parameters characterize the statistical behaviour of microbiome changes. Furthermore, we have verified that our conclusions are robust to systematic errors due to taxonomic assignment.

Taylor’s power law has been explained in terms of various effects, all without general consensus. It can be shown to have its origin in a mathematical convergence similar to the central limit theorem, so virtually any statistical model designed to produce a Taylor law converge to a Tweedie distribution (45), providing a mechanistic explanation based on the statistical theory of errors (46–48). To unveil the generic mechanisms that drive different scenarios in the  $\beta$ – $V$  space, we model the system by assuming that taxon relative abundance follows a Langevin equation with, on the one hand, a deterministic term that captures the fitness of each taxon and, on the other hand, a randomness term associated with Gaussian random noise (35). Both terms are modeled by power laws, with coefficients that can be interpreted as the taxon fitness  $F_i$  and the variability  $V$  (see Model under Material and Methods). In this model, when  $V$  is sufficiently low, abundances are stable in time. Differences in variability  $V$  can induce a noise-induced phase transition in relative abundances of taxa. The temporal evolution of the probability of a taxon having abundance  $x_i$  given its fitness is governed by the Fokker–Planck equation. The results of solving this equation show that the stability is best captured by a phase space determined by fitness  $F$  and amplitude of fluctuations  $V$  (see Figure 3).

The model predicts two phases for the gut microbiome: a stable phase with large variability





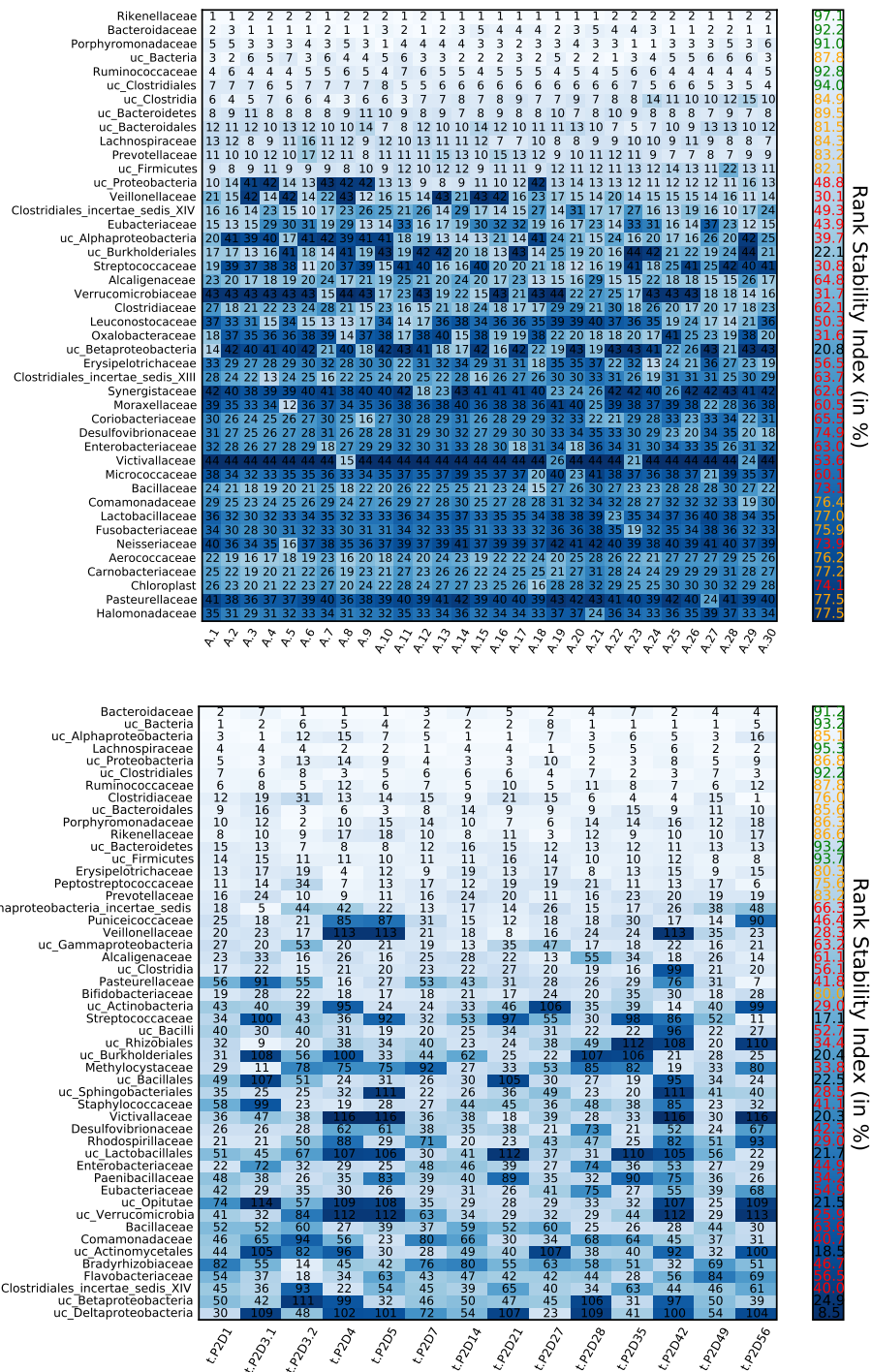
**Figure 3.** Microbiota states can be placed in the phase space  $F$ – $V$ . The light blue shaded region corresponds to the stable phase, while the grey shaded region is the unstable phase (the phase transition line is calculated for  $\alpha = \beta = 0.75$ ). We place healthy individuals (green) and individuals whose gut microbiota is threatened (antibiotics, IBS) in the phase space fitness–variability. Gut microbiota of healthy individuals over a long term span show a quasi-periodical variability (central period is ten days). We show that taking antibiotics (AB1 and AB2 correspond to first and second treatment respectively) induces a phase transition in the gut microbiota, which impacts its future changes. We also show an IBS–diagnosed patient transiting from the unstable to the stable phase.

that permits some changes in the relative abundances of taxa and an unstable phase with larger variability, above the phase transition, where the order of abundant taxa varies significantly with time. The microbiome of all healthy individuals was found to be in the stable phase, while the microbiome of several other individuals was shown to be in the unstable phase. In particular, individuals taking antibiotics and IBS–diagnosed patient P2 had the most severe symptoms. In this phase diagram, each microbiota state is represented by a point at its measured variability  $V$  and inferred fitness  $F$ . The model predicts high average fitness for all taxa, i.e., taxa are narrowly distributed in  $F$ . The fitness parameter has been chosen with different values for demonstrative purposes. Fitness is larger for the healthiest subjects and smaller for the IBS–diagnosed patients.

## Rank stability of the taxa

The rank dynamics and stability plot in Figure 4 shows the variation in the rank with time for the most dominant taxa and their calculated Rank Stability Index (RSI, as discussed in Material and Methods) for the taxa of a healthy subject (individual A, top) and from a subject diagnosed with IBS (patient P2, bottom) of the IBS study (11). The taxa are listed ordered by the accumulated frequency along the time series, so y-axis is an overall dominance axis for each sample set. Generally speaking, we observe that the most dominant taxa are the most rank stable.

Nevertheless, in the particular case of the healthy individual, *Burkholderiales* and *Betaproteobacteria* (taxa ordered as 18th and 25th in the dominance axis) show comparatively very low rank stability regarding similar dominant taxa while, on the other hand, *Comamonadaceae*, *Lactobacillaceae*, *Fusobacteriaceae*, *Aerococcaceae* and *Carnobacteriaceae* show higher stability than other more dominant taxa, forming a kind of *rank stability island* for medium-ranked taxa around position 40 in the dominance axis, and thus colored in orange following



**Figure 4.** Matrixes showing the rank variation throughout time for the most dominant elements (taxa) and their calculated Rank Stability Index (as discussed in Material and Methods). We show the matrix for samples from a healthy subject (top) and from a subject diagnosed with irritable bowel syndrome (bottom), studied in our lab (11).

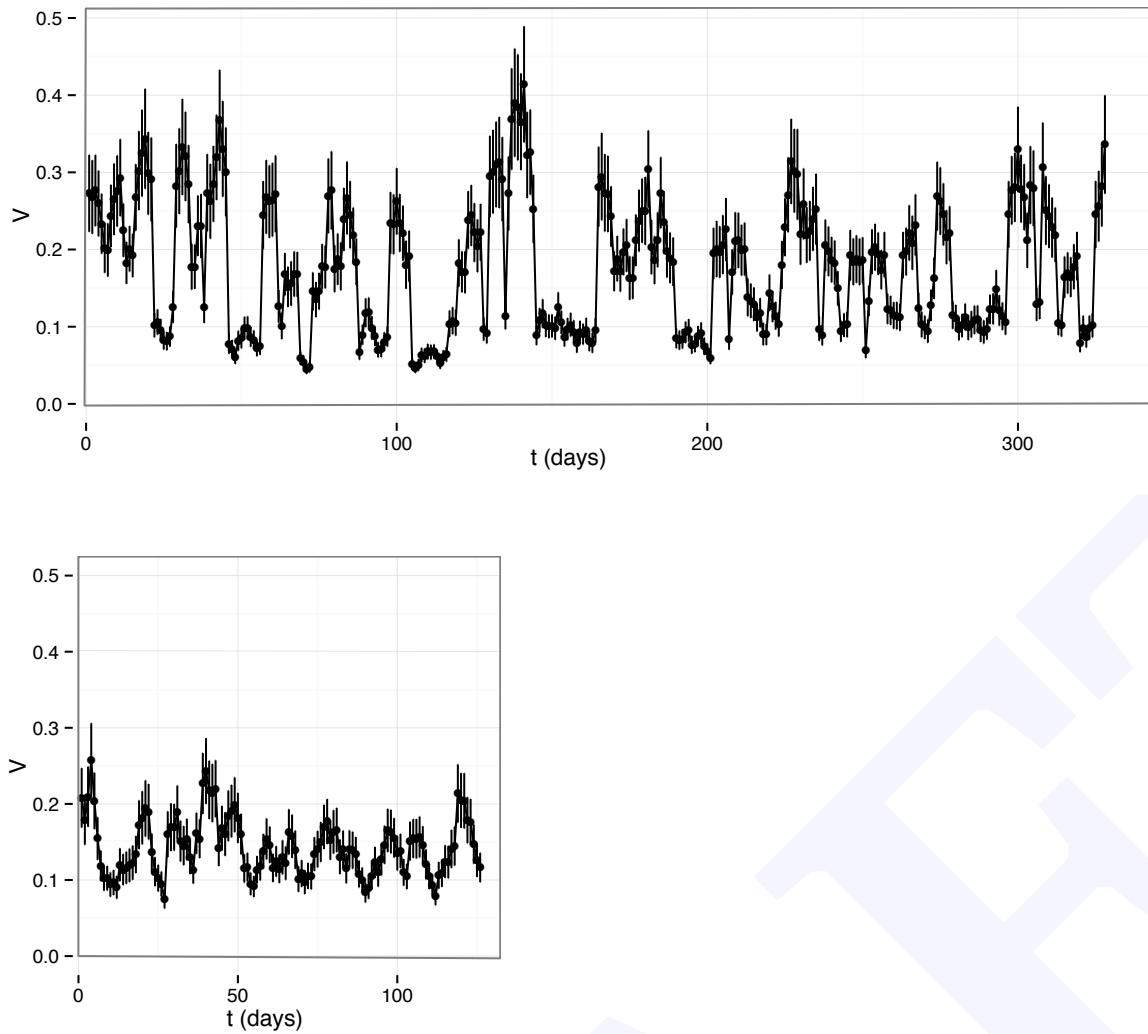
156 Table 1 criteria, since they show a moderately stable RSI.

157 In the IBS diagnosed patient, beyond the differences in dominance for the particular taxa, we  
 158 still observe that the most dominant are the most rank stable. However, as opposed to the  
 159 healthy individual results, far from presenting a *rank stability island*, the medium-ranked taxa  
 160 are very rank unstable, mostly due to transient (often one or two consecutive samples) but  
 161 deep drops in their relative abundance, which are usually happening more than twice along  
 162 the time series. That is, for instance, the case of *Sphingobacteriales* with two non-consecutive  
 163 samples dropping to 111th rank position. In other cases, the high rank instability comes from  
 164 a rank fluctuation over all the time series, as for *Streptococcaceae* and *Burkholderiales*, which  
 165 are ranking 26th and 29th respectively in the overall dominance axis but show very low RSI,  
 166 and thus colored in black attending to Table 1.

167 We found the presence of such of *rank stability island* for medium-ranked taxa in the other  
 168 healthy subjects (*B* and *C*) of the IBS study (11) together with its total absence for the  
 169 other IBS diagnosed patient (patient *P1*), which also presents very high rank instability in  
 170 its medium-ranked taxa.

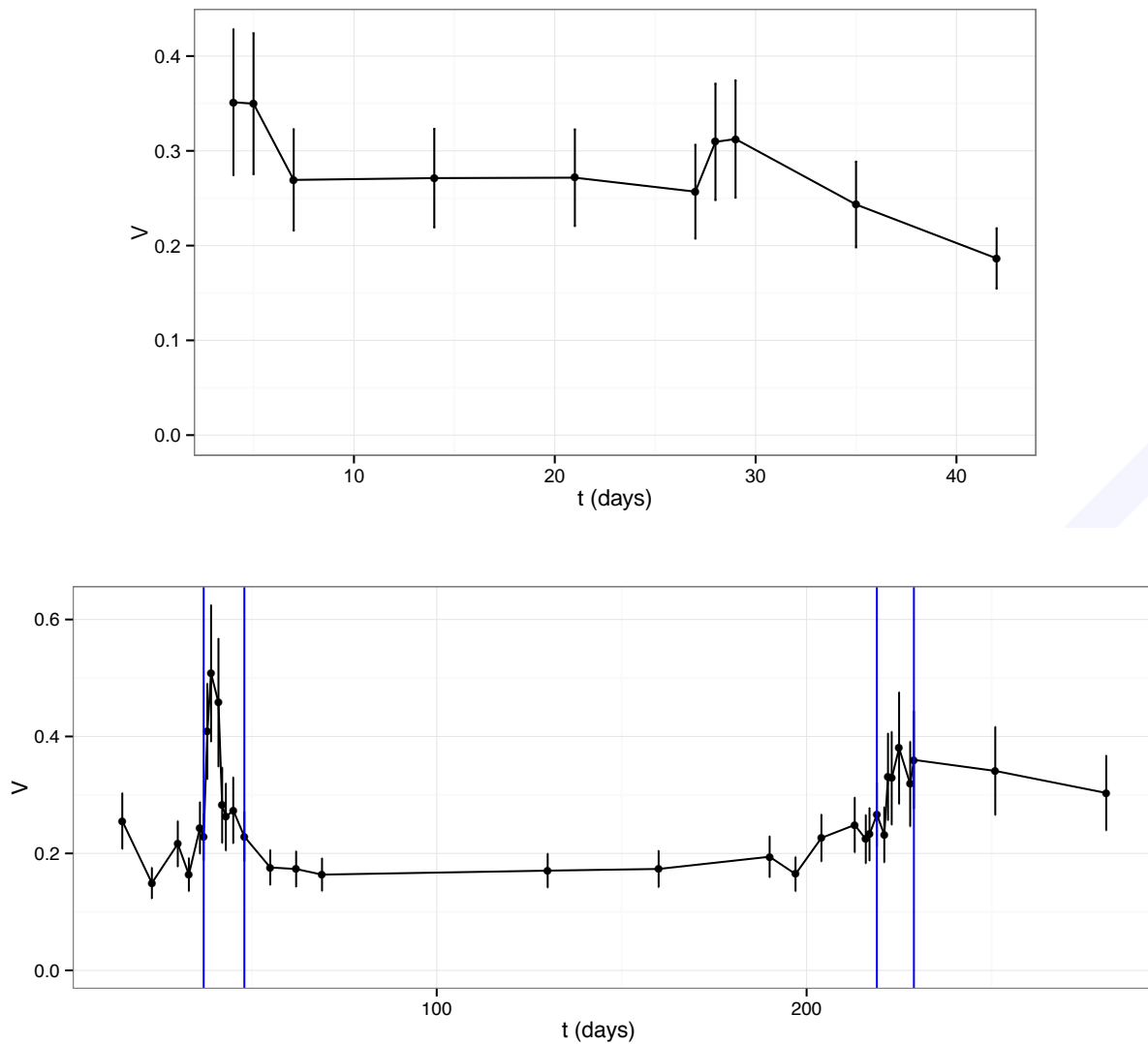
## 171 Time dependence of model parameters

172 Finally, we have studied the time dependence of the variability  $V$  and power law index  $\beta$  (see  
 173 Model under Material and Methods) by using a sliding window approach. The total number  
 174 of time points are divided in subsets of five points, where next subset is defined by adding  
 175 next time sampling and by eliminating the earliest one. Both parameters were calculated for  
 176 each subset against the average time lapse. Figure 5 shows the variability  $V$  as a function  
 177 of time for the largest sampling: two individuals in the Caporaso's study (36) corresponding  
 178 to the gut microbiota of a male (upper plot) and a female (lower plot). Figure 6 shows the  
 179 time evolution of  $V$  for patient P2 of the IBS study (11) (upper plot) and patient D in the



**Figure 5.**  $V$  as a function of time for the two individuals in the Caporaso's study (36): samples of gut microbiome of a male (upper plot) and a female (lower plot). Both samples show changes in the variability  $V$  with quasi-periodic behavior peaked at about 10 days. Variability grows more for the gut microbiota of the male and share a minimal value around 0.1 with the gut microbiota of the female.

180 antibiotics study (40) (lower plot).



**Figure 6.**  $V$  as a function of time for patient P2 of the IBS study (11) (upper plot) and patient D in the antibiotics study (40) (lower plot). The variability of the gut microbiota of P2 decreases from above 0.3 to below 0.2, showing a slow tendency to increase the order of the system. Antibiotic intake leads to a quick increase of variability which lasts for a few days to recover ordering. The second antibiotic treatment shows some memory (lower increase of variability) with a slower recovery. NOTE: The blue vertical lines in the lower plot are showing the periods of antibiotic treatment.

## Discussion

We have quantitatively characterized whether the microbiota belongs to a healthy individual or a subject corresponding to an altered or pathological state (i.e., altered diet, antibiotic treatment, early gut development, diagnosed IBS). Deciphering the mechanisms of disease requires in depth knowledge of the underlying biological mechanisms. We describe here the macroscopic behavior of disease by a noise-induced phase transition with a control parameter that can be measured by the temporal variability of the microbiome. The microbiota of healthy individuals and of individuals with pathologies represent different phases separated by this noise-induced phase transition. Improved high-throughput sequencing of samples from individuals monitored over time and taxonomic assigning methods will provide a better distinction among pathologies or altered states of the microbiota.

Specifically, the analysis of the rank stability of the samples of healthy and IBS diagnosed patients studied in our lab (11), suggests that the presence of a *rank stability island* among medium-ranked taxa could be an indicator of a healthy microbiota.

Final paragraph to talk about limitations and future perspectives: can we model stability in the functional landscape? Community assembly for itself doesn't explain everything, we need to move forward and think about what can be happening in complex ecosystems as the human microbiota.

## Materials and Methods

### Model

We model the microbial abundances across time along the lines of Blumm *et al.* (35). The dynamics of taxon relative abundances is described by the Langevin equation:

$$\dot{x}_i = F_i \cdot x_i^\alpha + V \cdot x_i^\beta \xi_i(t) - \phi(t) \cdot x_i, \quad (1)$$

where  $F_i$  captures the fitness of the taxon  $i$ ,  $V$  corresponds to the noise amplitude and  $\xi_i(t)$  is a Gaussian random noise with zero mean  $\langle \xi_i(t) \rangle = 0$  and variance uncorrelated in time,  $\langle \xi_i(t) \xi_i(t') \rangle = \delta(t' - t)$ . The function  $\phi(t)$  ensures the normalization at all times,  $\sum x_i(t) = 1$ , and corresponds to  $\phi(t) = \sum F_i x_i^\alpha + \sum V x_i^\beta \xi_i(t)$ . The temporal evolution of the probability that a taxon  $i$  has a relative abundance  $x_i(t)$ ,  $P(x_i, t)$ , is determined by the Fokker-Planck equation:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x_i} [(F_i \cdot x_i^\alpha - \phi(t) \cdot x_i) \cdot P] + \frac{1}{2} \frac{\partial^2}{\partial x_i^2} (V^2 \cdot x_i^{2\beta} \cdot P). \quad (2)$$

The microbiota evolves towards a steady-state with a time-independent probability depending on the values of  $\alpha$ ,  $\beta$ ,  $F_i$  and  $V$ . For  $\alpha < 1$  (otherwise, systems are always unstable), the steady-state probability may be localized in a region around a preferred value or broadly distributed over a wide range, depending on whether the fitness  $F_i$  dominates or is overwhelmed by the noise amplitude  $V$ . The steady-state solution of the Fokker-Planck equation is given by:

$$P_0(x_i) = C_{ne}(\alpha, \beta, F_i, V) \cdot x_i^{-2\beta} \cdot \exp\left[\frac{2F_i}{V^2} \frac{x_i^{1+\alpha-2\beta}}{1+\alpha-2\beta} - \frac{\phi_0}{V^2} \frac{x_i^{2-2\beta}}{1-\beta}\right] \quad \text{if } 2\beta \neq 1+\alpha,$$

$$P_0(x_i) = C_e(\alpha, \beta, F_i, V) \cdot x_i^{\frac{2F_i}{V^2}-2\beta} \cdot \exp\left[\frac{\phi_0}{V^2} \frac{x_i^{2-2\beta}}{1-\beta}\right] \quad \text{if } 2\beta = 1+\alpha,$$



218 where  $\phi_0 = (\sum_i F_i^{1/(1-\alpha)})^{1-\alpha}$  and  $C_{ne}$  and  $C_e$  are integrals that should be solved numerically  
 219 for the parameters of interest. The ordered phase happens when the solution has a maximum  
 220 in the physical interval ( $0 < x_i < 1$ ). For larger  $V$ , the transition to a disordered phase happens  
 221 when the maximum shifts to the unphysical region  $x_i < 0$ , which sets the phase transition  
 222 region  $V(\alpha, \beta, F_i)$ . The phase transition region can be calculated analytically in particular  
 223 cases:

$$\begin{aligned} 224 \quad F_i^2 &= 4\beta\phi_0V^2 \quad \text{if } \beta = \alpha \neq 1, \\ 225 \quad F_i &= \beta V^2 \quad \text{if } 2\beta = 1 + \alpha, \end{aligned}$$

226 where the first case, simplifies to  $F = 3V^2$  if  $\beta = 0.75$  and the fitness of this taxon dominates  
 227 in  $\phi_0$ . In many physical systems (Brownian motion is the classical example), the two terms  
 228 of the Langevin equation are related. The *fluctuation–dissipation theorem* states a general  
 229 relationship between the response to an external disturbance and the internal fluctuations  
 230 of the system (49). The theorem can be used as the basic formula to derive the fitness from  
 231 the analysis of fluctuations of the microbiota, assuming that it is in equilibrium (the ordered  
 232 phase).

233 Explain better the fluctuation-dissipation theorem

## 234 Selection and Methods

235 The bacteria and archaea taxonomic assignments were obtained by analysing 16S rRNA se-  
 236 quences, which were clustered into operational taxonomic units (OTUs) sharing 97 % se-  
 237 quence identity using QIIME (41). WGS data (38) were analysed and assigned at strain level  
 238 by the Livermore Metagenomic Analysis Toolkit (LMAT) (42), according to their default qual-  
 239 ity threshold. Genus, with best balance between error assignment and number of taxa, was  
 240 chosen as our reference taxonomic level. We have verified that our conclusions are not sig-

nificantly affected by selecting family or species as the reference taxonomic level (see Figure 7).

Specify, in each study treated, the nature of the samples (conditions, timespan between timepoints, subjects). Specify, and it is very important, what we consider healthy in each study (for example: pre-antibiotics is healthy)

## Sample selection

We have chosen studies about relevant pathologies containing metagenomic sequencing time data series of bacterial populations from humans in different healthy and non-healthy states. We have selected only those individuals who had three or more time points of data available in databases. Metadata of each study is provided in Tables ?? to ??. All used 16S rRNA gene sequencing except for the study of the discordant kwashiorkor twins (38) (see Tables ?? and ??) where shotgun metagenomic sequencing (SMS) and 16S rRNA were used. In the latter case we selected to work with SMS data to show that our method is valid regardless of the source of taxonomic information. Each one of the datasets was treated as follows:

## 16rRNA sequences processing

Reads from the selected studies were first quality filtered using the FastX toolkit (50), allowing only those reads which had more than 25 of quality along the 75% of the complete sequence. 16S rRNA reads were then clustered at 97% nucleotide sequence identity (97% ID) into operational taxonomic units (OTUs) using QIIME package software (41) (version 1.8) We followed open reference OTU picking workflow in all cases. The clustering method used was uclust, and the OTUs were matched against Silva database (51) (version 111, July 2012) and were assigned to taxonomy with an uclust-based consensus taxonomy assigner. The parameters used in this step were: similarity 0.97, prefilter percent id 0.6, max accepts 20,

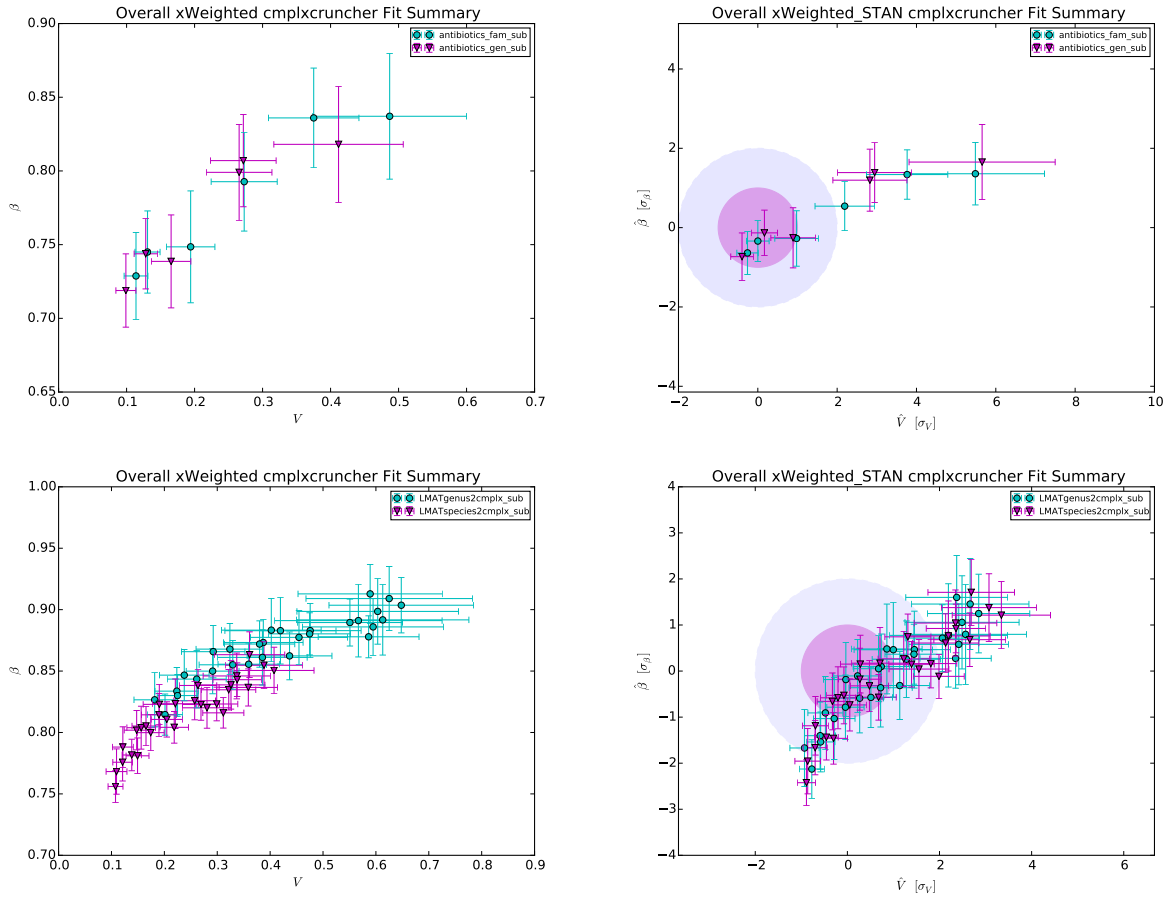
264 max rejects 500.

## 265 **Metagenomic sequences processing**

266 Metagenomic shotgun (and 16S too) sequences were analyzed with LMAT (Livermore Metage-  
 267 nomics Analysis Toolkit) software package (42) (version 1.2.4, with Feb'15 release of data  
 268 base *LMAT-Grand*). LMAT was run using a Bull shared-memory node belonging to the team's  
 269 HPC (high performance computing) cluster. It is equipped with 32 cores (64 threads available  
 270 using Intel Hyper-threading technology) as it has 2 Haswell-based Xeons, the E5-2698v3@2.3  
 271 GHz, sharing half a tebibyte (0.5 TiB, that is, 512 gibibytes) of DRAM memory. This node is  
 272 also provided with a card PCIe SSD as NVRAM, the P420m HHHL, with 1.4 TB, and 750000  
 273 reading IOPS, 4 KB, achieving 3.3 GB/s, which Micron kindly issued free of charge, as a  
 274 sample for testing purposes. The computing node was supplied with a RAID-0 (striping)  
 275 scratch disk area. We used the "Grand" database (43), release Feb'15, provided by the LMAT  
 276 team, where "Grand" refers to a huge database that contains k-mers from all viral, prokary-  
 277 ote, fungal and protist genomes present in the NCBI database, plus Human reference genome  
 278 (hg19), plus GenBank Human, plus the 1000 Human Genomes Project (HGP) (this represent  
 279 about 31.75 billion k-mers occupying 457.62 GB) (43). Previously to any calculation, the full  
 280 database was loaded in the NVRAM. With this configuration the observed LMAT sustained se-  
 281 quence classification rate was 20 kpb/s/core. Finally, it is worth mentioning that a complete  
 282 set of Python scripts have been developed as back-end and front-end of the LMAT pipeline in  
 283 order to manage the added complexity of time series analysis.

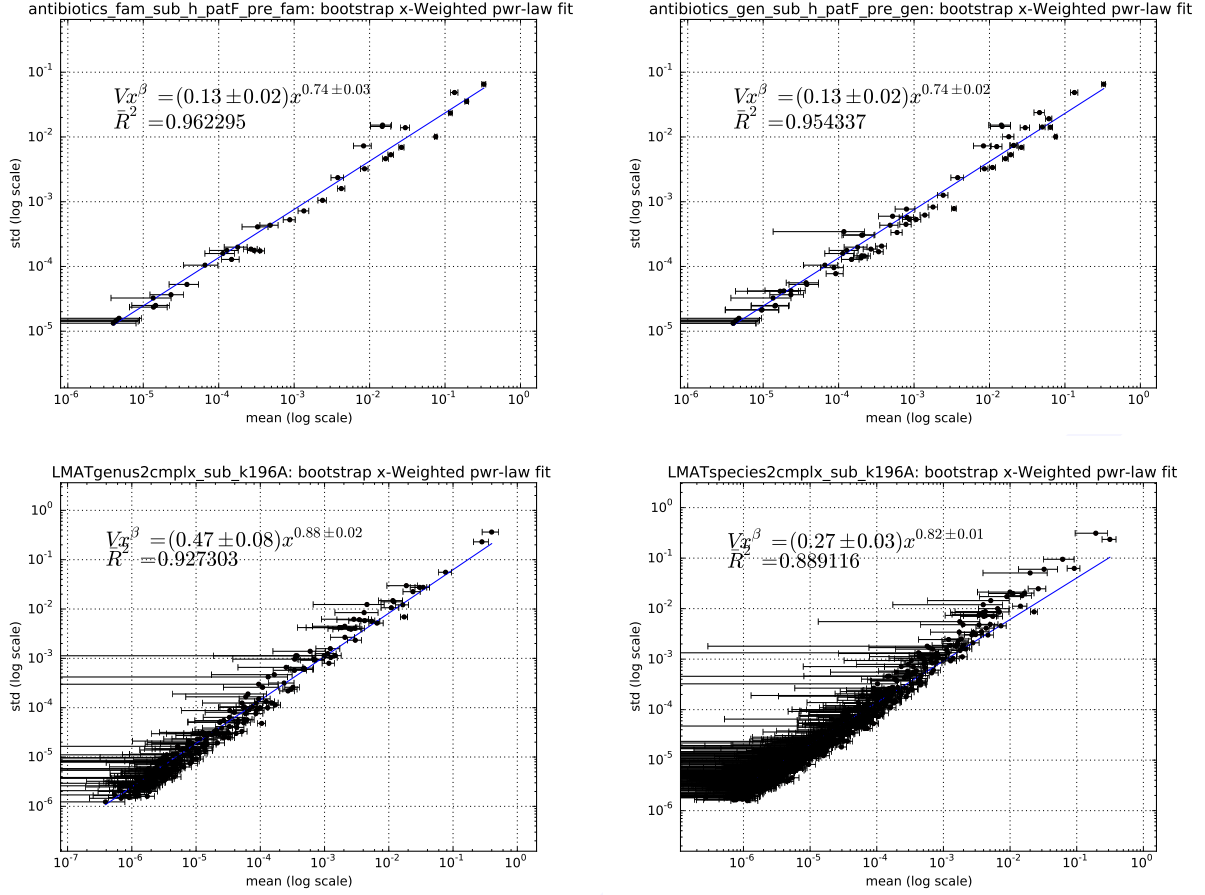
## 284 **Taxa level selection**

285 We selected genus as taxonomic level for the subsequent steps of our work. In order to ensure  
 286 that, between adjacent taxonomic levels, there were not crucial differences which could still



**Figure 7.** Overview of comparison of different approaches based on adjacent taxonomic levels using plots in the Taylor-parameters space. For 16S (former row of subfigures), the levels are family vs. genus, whereas for SMS (latter row of subfigures) levels are genus vs. species. The left column shows the raw results and the right column plots the standardized results (see Section )

be of relevance after standardization (see last subsection of Material and Methods), we tested two different data sets. In the former, the antibiotics study (40) with 16S data, we tested the differences between genus and family levels. The latter dataset tested was the kwashiorkor discordant twins study (38) for both genus and species taxonomic levels. The Figures 7 (overview) and 8 (detail) plot the comparison between studies (and so, 16S and SMS) and between adjacent taxonomic levels.



**Figure 8.** Detail of comparison of different approaches based on adjacent taxonomic levels using plots of X-weighted power-law fits (see Material and Methods). The former row of subfigures shows examples for 16S, whereas the latter row of subfigures plots examples for SMS. The left column shows results for the superior taxonomic level (family for 16S, genus for SMS), while the right column shows results for the inferior level (genus for 16S, specie for SMS).

## 293 X-weighted power-law fit

294 When fitting the power-law of std vs. mean, we can take into account that every mean has  
 295 uncertainty and estimate it for a sample size  $n$  by the SEM (*Standard Error of the Mean*). Here,  
 296 the uncertainties affect the independent variable, so the fit is not so trivial as a Y-weighted  
 297 fit, where the uncertainties affect the dependent variable. A standard approach to do this  
 298 fit is: a) invert your variables before applying the weights, b) then perform the weighted  
 299 fit, and finally, c) revert the inversion. This method is deterministic, but the approximate  
 300 solution worsens with smaller coefficients of determination. To overcome this limitation, we  
 301 developed a stochastic method by using a bootstrapping-like strategy that avoids the inversion  
 302 and is applicable regardless of the coefficient of determination.

303 The basic idea of bootstrapping is that inference about a population from sample data (sample  
 304  $\rightarrow$  population) can be modeled by resampling the sample data and performing inference on  
 305 (resample  $\rightarrow$  sample). To adapt this general idea to our problem, we resample the x-data  
 306 array using its errors array. That is, for each replicate, a new x-data array is computed based  
 307 on:

$$308 \quad x_i^* = x_i + v_i$$

309 where  $v_i$  is a Gaussian random variable with mean  $\mu_i = 0$  and standard deviation  $\sigma_i = \text{SEM}_i$ ,  
 310 as defined previously. For each replicate a complete un-weighted power-law fit is performed,  
 311 where to choose between fitting power laws ( $y = Vx^\beta$ ) using linear regression on log-  
 312 transformed (LLR) data versus non-linear regression (NLR) we mainly follow *General Guide-*  
 313 *lines for the Analysis of Biological Power Laws* (52). The parameters of the X-weighted fit are  
 314 then estimated by averaging through all the replicate fits performed, and their errors are es-  
 315 timated by computing the standard deviation also for all the fits. At the end of each step, the  
 316 relative error is calculated by comparing the fit parameters estimation in the last step with  
 317 the previous one. Finally, both the coefficient of determination of the fit and the coefficient

Case	Condition	Colour	Description
1	$1 \geq \text{RSI} > 0.99$	blue	constant rank
2	$\text{RSI} > 0.90$	green	highly stable rank
3	$\text{RSI} > 0.75$	orange	moderately stable rank
4	$\text{RSI} > 0.25$	red	unstable rank
5	$0.25 \geq \text{RSI} \geq 0$	black	very unstable rank

**Table 1.** Colour code of the RSI percentage text shown in Figure 4, following the first condition satisfied.

of correlation between the fit parameters are estimated by averaging.

## Rank Stability Index

The Rank Stability Index (RSI) is shown as a percentage in a separate bar on the right of the rank matrix plot shown in Figure 4. The RSI is strictly 1 for an element whose range never changes over time, and is strictly 0 for an element whose rank oscillates between the extremes from time to time. So, RSI is calculated, per element, as 1 less the quotient of the number of true rank hops taken between the number of maximum possible rank hops, all powered to  $p$ :

$$\text{RSI} = \left(1 - \frac{\text{true rank hops}}{\text{possible rank hops}}\right)^p = \left(1 - \frac{D}{(N-1)(t-1)}\right)^p$$

where  $D$  is the total of rank hops taken by the studied element,  $N$  is the number of elements that have been ranked, and  $t$  is the number of time samples. The power index  $p = 4$  is arbitrarily chosen to increase the resolution in the stable region.

The colour code of the RSI percentage text in the rank plot shown in Figure 4 is chosen following the first condition satisfied from those shown in Table 1.

### 331 Standardization

332 In order to properly show all the studies under common axes, we decided to standardize the  
 333 Taylor parameters using the group of healthy individuals for each study. With this approach,  
 334 all the studies can be visualized in a shared plot with units of Taylor-parameters standard-  
 335 deviation on their axes.

336 For a Taylor parameter, e.g.  $V$ , the estimate of the mean ( $\hat{V}$ ) for the healthy subpopulation,  
 337 composed of  $h$  individuals, is:

$$338 \quad \hat{V} = \frac{1}{W_1} \sum_{i=1}^h V_i \omega_i = \sum_{i=1}^h V_i \omega_i$$

339 as  $W_1 = \sum_{i=1}^h \omega_i = 1$ , since  $\omega_i$  are normalized weights calculated as:

$$340 \quad \omega_i = \frac{\frac{1}{\sigma_{V_i}^2}}{\sum_{i=1}^h \frac{1}{\sigma_{V_i}^2}}$$

341 being  $\sigma_{V_i}$  the estimation of the uncertainty in  $V_i$  obtained together with  $V_i$  from the X-weighted  
 342 power-law fit described in Section , for healthy individuals.

343 Likewise, the estimation of the standard deviation for the healthy population ( $\hat{\sigma}_V$ ) is:

$$344 \quad \hat{\sigma}_V = \sqrt{\frac{1}{W_1 - \frac{W_2}{W_1}} \sum_{i=1}^h [\omega_i (V_i - \hat{V})^2]}$$

345 being  $W_2 = \sum_{i=1}^h \omega_i^2$ , which finally yields to:

$$346 \quad \hat{\sigma}_V = \sqrt{\frac{1}{1 - \sum_{i=1}^h \omega_i^2} \sum_{i=1}^h [\omega_i (V_i - \hat{V})^2]}$$



## 347 Acknowledgments

348 Authors declare that there are no competing financial interests in relation to the work de-  
349 scribed here.

## 350 Funding Information

351 This work is supported by Generalitat Valenciana Prometeo Grants II/2014/050, II/2014/065,  
352 by the Spanish Grants FPA2011-29678, BFU2012-39816-C02-01 of MINECO and by PITN-GA-  
353 2011-289442-INVISIBLES. JMM & DMM acknowledge FPI and FISABIO fellowships. **Modificar**  
354 **becas de JMM y DMM ¿poner algún grant más?**

## References

1. **Rosenberg E, Zilber-Rosenberg I.** 2016. Microbes Drive Evolution of Animals and Plants: the Hologenome Concept. *MBio* **7**:e01395–15–.
2. **Bordenstein SR, Theis KR.** 2015. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLOS Biol* **13**:e1002226.
3. **Swann JR, Want EJ, Geier FM, Spagou K, Wilson ID, Sidaway JE, Nicholson JK, Holmes E.** 2011. Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc Natl Acad Sci* **108**:4523–4530.
4. **nSpencer MD, Hamp TJ, Reid RW, Fischer LM, Zeisel SH, Fodor AA.** 2011. Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* **140**:976–986.
5. **Samuel BS, Shaito A, Motoike T, Rey FE, Backhed F, Manchester JK, Hammer RE, Williams SC, Crowley J, Yanagisawa M, Gordon JI.** 2008. Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc Natl Acad Sci* **105**:16767–16772.
6. **Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, Glickman JN, Garrett WS.** 2013. The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science (80- )* **341**:569–573.
7. **Kimura I, Ozawa K, Inoue D, Imamura T, Kimura K, Maeda T, Terasawa K, Kashiwara D, Hirano K, Tani T, Takahashi T, Miyauchi S, Shioi G, Inoue H, Tsujimoto G.** 2013. The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. *Nat Commun* **4**:1829.
8. **Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Di Yu, Schilter HC, Rolph MS, Mackay F, Artis D, Xavier RJ, Teixeira MM, Mackay CR.** 2009. Regulation of in-

inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**:1282–1286.

9. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto J-M, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**:55–60.
10. Brown JM, Hazen SL. 2015. The Gut Microbial Endocrine Organ: Bacterially Derived Signals Driving Cardiometabolic Diseases. *Annu Rev Med* **66**:343–359.
11. Durbán A, Abellán JJ, Jiménez-Hernández N, Artacho A, Garrigues V, Ortiz V, Ponce J, Latorre A, Moya A. 2013. Instability of the faecal microbiota in diarrhoea-predominant irritable bowel syndrome. *FEMS Microbiol Ecol* **86**:581–589.
12. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, González A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ. 2014. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**:382–392.
13. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau L, Griffi NW, Lombard V, Henrissat B, Bain JR, Michael J, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle J, Martini MC, Ursell LK, Clemente JC, Treuren W Van, William A, Knight R, Newgard CB, Heath AC, Gordon JI, Kau AL, Griffin NW, Muehlbauer MJ. 2013. Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice Gut

Microbiota from Twins Metabolism in Mice. *Science* **341**:1241214.

14. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. 2009. LETTERS A core gut microbiome in obese and lean twins. *Nature* **457**:480–484.
15. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A, DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q, Province MA, Petri WA, Ahmed T, Gordon JI. 2014. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510**:417–21.
16. Moya A, Ferrer M. 2016. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends Microbiol* **24**:402–413.
17. Marchesi JR, Adams DH, Fava F, Hermes GD a, Hirschfield GM, Hold G, Quraishi MN, Kinross J, Smidt H, Tuohy KM, Thomas L V, Zoetendal EG, Hart A. 2015. The gut microbiota and host health: a new clinical frontier. *Gut* 1–10.
18. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, Dhoe K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaut L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. 2016. Population-level analysis of gut microbiome variation. *Science* (80- ) **352**:560–564.
19. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, Wang J, Imhann F, Brandsma E, Jankipersadsing SA, Joossens M, Cenit MC, Deelen P, Swertz MA, Weersma RK, Feskens EJM, Netea MG, Gevers D, Jonkers D, Franke L, Aulchenko YS, Huttenhower C, Raes J, Hofker MH, Xavier RJ, Wijmenga C, Fu J. 2016. Population-based

metagenomics analysis reveals markers for gut microbiome composition and diversity.  
Science (80- ) **352**:565–569.

20. **Wu H, Tremaroli V, Bäckhed F** 2015. Linking Microbiota to Human Diseases: A Systems Biology Perspective. Trends Endocrinol Metab **26**:758–770.

21. **Noecker C, Eng A, Srinivasan S, Theriot CM, Young VB, Jansson JK, Fredricks DN, Borenstein E** 2016. Metabolic Model-Based Integration of Microbiome Taxonomic and Metabolomic Profiles Elucidates Mechanistic Links between Ecological and Metabolic Variation. mSystems **1**:e00013–15.

22. **Greenblum S, Turnbaugh PJ, Borenstein E** 2012. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc Natl Acad Sci **109**:594–599.

23. **Taylor, L.R.** 1961. Aggregation, Variance and the mean. Nature **189**, 732-35.

24. **de Menezes MA, Barabási A-L.** 2004. Fluctuations in network dynamics. Phys Rev Lett **92**:1–4.

25. **Mantegna RN, Stanley HE.** 1995. Scaling behaviour in the dynamics of an economic index. Nature **376**:46–49.

26. **Eisler Z, Kertesz J, Yook SH, Barabasi AL.** 2005. Multiscaling and non-universality in fluctuations of driven complex systems. Europhys Lett **69**:664–670.

27. **Reed DH, Hobbs GR.** 2004. The relationship between population size and temporal variability in population size. Anim Conserv **7**:1–8.

28. **Anderson RM, Gordon DM, Crawley MJ, Hassell MP** 1982. Variability in the abundance of animal and plant species. Nature **18**: 245–248

29. **Živković J, Tadić B, Wick N, Thurner S.** 2006. Statistical indicators of collective behavior and functional clusters in gene networks of yeast. Eur Phys J B **50**:255–258.

- 453 30. **Kendal WS.** 2003. An Exponential Dispersion Model for the Distribution of Human  
454 Single Nucleotide Polymorphisms. *Mol Biol Evol* **20**:579–590.
- 455 31. **Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, Ma ZS, Shi P** 2014. Spatial het-  
456 erogeneity and co-occurrence patterns of human mucosal-associated intestinal micro-  
457 biota. *ISME J* **8**:881–93.
- 458 32. **Pérez-Cobas AE, Artacho A, Ott SJ, Moya A, Gosalbes MJ, Latorre A.** 2014. Struc-  
459 tural and functional changes in the gut microbiota associated to *Clostridium difficile*  
460 infection. *Front Microbiol* **5**:1–15.
- 461 33. **Ding T, Schloss PD.** 2014. Dynamics and associations of microbial community types  
462 across the human body. *Nature* **509**:357–360.
- 463 34. **Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, Koenig SSK, Fu L,**  
464 **Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J.** 2012. Temporal dynamics of the human  
465 vaginal microbiota. *Sci Transl Med* **4**:132ra52.
- 466 35. **Blumm N, Ghoshal G, Forró Z, Schich M, Bianconi G, Bouchaud J-P, Barabási A-L.**  
467 2012. Dynamics of Ranking Processes in Complex Systems. *Phys Rev Lett* **109**:128701.
- 468 36. Caporaso, J.G. et al. Moving pictures of the human microbiome. *Genome Biol.* **12**, R50  
469 (2011).
- 470 37. Faith, J.J. et al. The long-term stability of the human gut microbiota. *Science* **341**,  
471 1237439 (2013).
- 472 38. Smith M.I. et al. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor.  
473 *Science* **339**, 548-54 (2013).
- 474 39. David, L.A. et al. Diet rapidly and reproducibly alters the human gut microbiome.  
475 *Nature* **505**, 559-63 (2014).

40. Dethlefsen L., Relman D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Nat. Acad. Sci. USA* **108**, 4554-61 (2011).
41. Caporaso, J.G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335-6 (2010).
42. **Ames SK, Hysom DA, Gardner SN, Lloyd GS, Gokhale MB, Allen JE.** 2013. Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics* **29**:2253-2260.
43. **Ames SK, Gardner SN, Marti JM, Slezak TR, Gokhale MB, Allen JE.** 2015. Using populations of human and microbial genomes for organism detection in metagenomes. *Genome Res.* **25**:1056-67.
44. Eisler,Z., Bartos,I., Kertesz,J. Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 85 (2008).
45. Jorgensen,B., Martinez,J.R., Tsao,M. Asymptotic behaviour of the variance function. *Scand. J. Statist..* **21**, 223-243 (1994).
46. Fronczak,A., Fronczak,P. Origins of Taylor's power law for fluctuation scaling in complex systems. *Phys. Rev. E* **81**, 066112 (2010).
47. Kendal, W.S., Jorgensen,B. Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. *Phys. Rev. E* **83**, 066115 (2011).
48. Kendal, W.S., Jorgensen,B. Tweedie convergence: A mathematical basis for Taylor's power law. *Phys. Rev. E* **84**, 066120 (2011).
49. Weber, J. *et al.* Fluctuation dissipation theorem. *Phys. Rev.* **101**, 1620-6 (1956).
50. Gordon, A., Hannon, G.J. FASTX-Toolkit. FASTQ/A shortreads pre-processing tools (2010). [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/) (accessed 23 Feb 2015).

- 500 51. Quast C. *et al.* The SILVA ribosomal RNA gene database project: improved data pro-  
501 cessing and web-based tools (2013)
- 502 52. Xiao Xiao, Ethan P. White, Mevin B. Hooten, and Susan L. Durham. On the use of log-  
503 transformation vs. nonlinear regression for analyzing biological power laws. *Ecology*  
504 **92**, 10, 1887-1894 (2011).
- 505 53. Magee L.,  $R^2$  measures based on wald and likelihood ratio joint significance tests. *The*  
506 *American Statistician* **44**, 3, 250-253 (1990).
- 507 54. Nagelkerke N.J.D., A note on a general definition of the coefficient of determination.  
508 *Biometrika* **78**, 3, 691-692 (1991).
- 509 55. Wu, C.F.J. Jackknife, bootstrap and other resampling methods in regression analysis.  
510 (with discussions) *The Annals of Statistics* **14**: 1261-1350 (1986)
- 511 Eliminar *et al.* y poner la referencia completa como exige la guía de estilo  
512 de la revista...