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**[1,2,*]**, Daniel Martínez-Martínez**[1,2,3,*]**, Manuel Peña**[2]**, César Gracia**[1,2]**,**

7 **Amparo Latorre**[1,3,4,5]**, Andrés Moya**[1,3,4,5,#] **& Carlos P. Garay**[1,2,#]

8 [1]Institute for Integrative Systems Biology (I2SysBio), 46980, Spain.

9 [2]Instituto de Fisica Corpuscular, CSIC-UVEG, P.O. 22085, 46071, Valencia, Spain.

10 [3]FISABIO, Avda de Catalunya, 21, 46020, Valencia, Spain.

11 [4]Cavanilles Institute of Biodiversity and Evolutionary Biology, UVEG, 46980, Spain.

12 [5]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: 5437 of 5000 max

16

[*] Equally contributed
[#] Corresponding authors: andres.moya@uv.es, 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 shotgun metagenomic sequencing (SMS) 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 us 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.

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. However, these concepts are still in debate (*3*). We are populated by a myriad of microorganisms that are interacting with us in several physiological processes such as metabolism of the bile acids (*4*), of the choline (*5*) or key-route metabolites as short-chain fatty acids (*6, 7*) which are also involved in immune system maturation (*8, 9*). Human microbiota has been suggested to be closely related to diseases like type 2 diabetes (*10*), cardiovascular disease (CVD) (*11*), irritable bowel syndrome (*12*), Crohn's disease (*13*), some affections as obesity (*14, 15*), malnutrition (*16*) among other multiple diseases (*17*). Recently, even a mystifying and elusive diagnosis condition as chronic fatigue syndrome, which has often been suggested to be a psychosomatic disease, has been closely related to reduced diversity and altered composition of the gut microbiome (*18*). 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 (*19*). To define normal microbiota and how it's compositional changes can origin some diseases are important issues still in need for scientific answers (*20, 21*).

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 (*22–24*),

even proposing that they have universal dynamics (*25*). Microbiota can be approached under the light of ecological theory where we can find, for instance, general principles as the Taylor's law (*26*), 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 (*27*), stock markets (*28, 29*), animal populations (*26, 30, 31*), gene expression (*32*), or in the human genome (*33*). Taylor's law has been applied to microbiota in a spatial way in the work of Zhang *et al.*, (2014) (*34*), where they show that this population tend to be in an aggregated way rather than in a random distribution. Despite its ubiquity, it has been studied only in experimental settings (*35, 36*) but 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 (*37–39*)

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.* (*40*). 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 analyzed 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,

89 changes in relative abundances of taxa follow a ubiquitous pattern known as the fluctua-

90 tion scaling law (*51*) or Taylor's power law (*26*), i.e., microbiota of all detected taxa follows

91 $\sigma_i = V \cdot x_i^\beta$, a power law dependence between mean relative abundance x_i and dispersion

92 σ_i. The law seem to be ubiquitous, spanning even to six orders of magnitude in the observed

93 relative abundances. As can be seen in Figure 1, the most abundant species are less volatile

94 in relative terms than the less abundant. The fitting to the power law is always robust ($R^2 >$

95 0.88) and does not depend on the microbiome condition. The power law (or scaling) index β

96 and the variability V (hereafter Taylor parameters) appear to be correlated with the stability

97 of the community and related with the health status of the host, which we consider the main

98 finding exposed in this article.

99 Taylor parameters describing the temporal variability of the gut microbiome in our sampled

100 individuals are shown in Tables S1 to S6. Our results hint at an ubiquitous behavior. On

101 the first hand, the variability (which corresponds to the maximum amplitude of fluctuations)

102 is large, which suggests resilient capacity of the microbiota. On the other hand, the scaling

103 index is always smaller than one, which means that more abundant taxa are less volatile than

104 less abundant ones. In addition, Taylor parameters for the microbiome of healthy individuals

105 in different studies are compatible within estimated errors. This enables us to define an area

106 in the Taylor parameter space that we called the *healthy zone*.

107 In order to jointly visualize and compare the results of individuals from different studies, their

108 Taylor parameters have been standardized, where standardization means that each param-

109 eter is subtracted by the mean value and divided by the standard deviation of the group of

110 healthy individuals for each study (for details of the procedure, please see Standardization

111 subsection in Material and Methods). The healthy zone and the standardized Taylor parame-

112 ters for individuals whose gut microbiota is altered (i.e., suffering from kwashiorkor, altered

113 diet, antibiotics or IBS) is shown in Figure 2. Children developing kwashiorkor show smaller

114 variability than their healthy twins. A meat/fish-based diet increases the variability signifi-
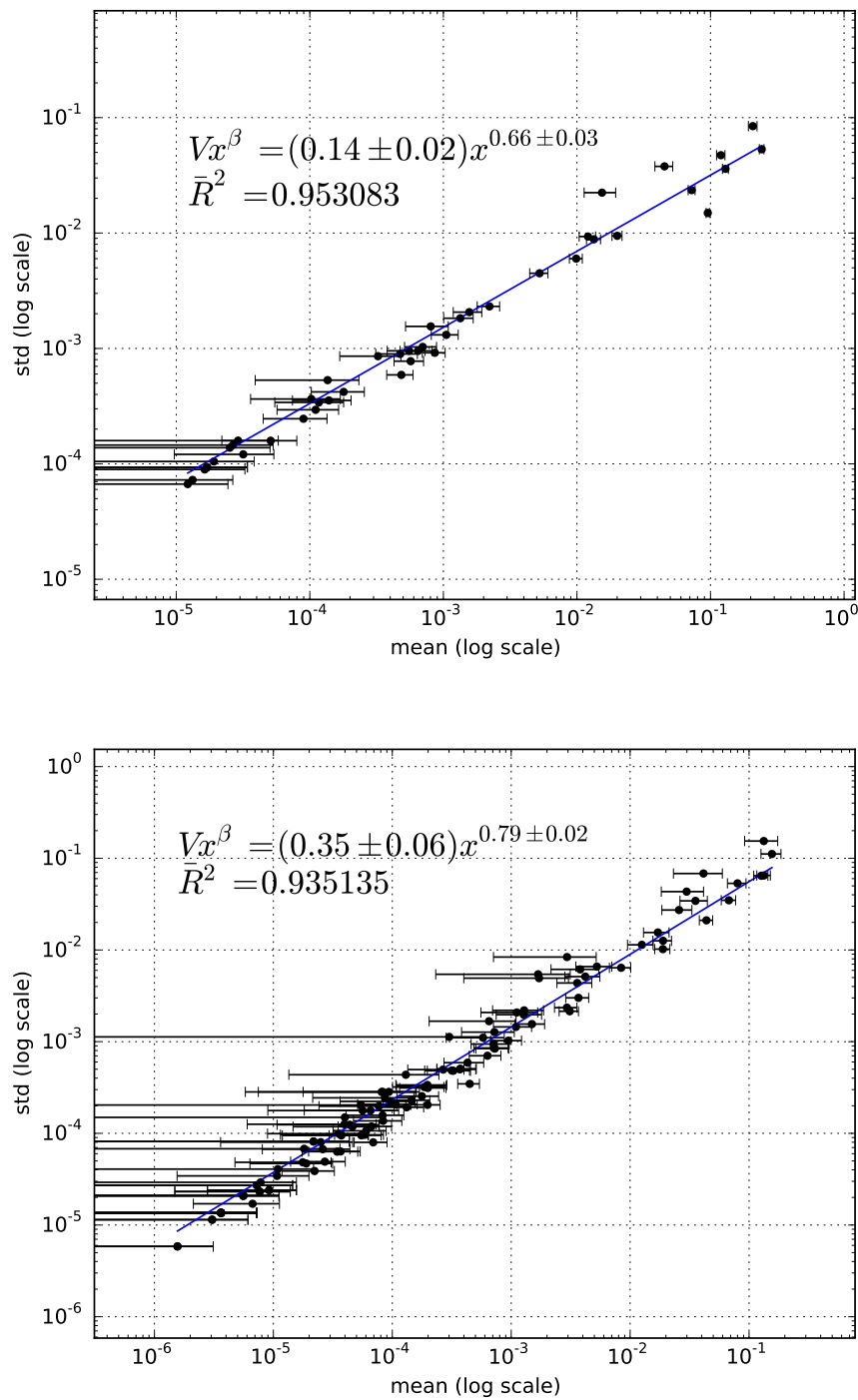
Figure 1. X-weighted power-law fits of the standard deviations versus the mean values for each bacterial genus monitored in time. We show the fit for samples from a healthy subject (top) and from a subject diagnosed with irritable bowel syndrome (bottom), studied in our lab (*12*). Taylor's power law seems to be ubiquitous, spanning to six orders of magnitude.
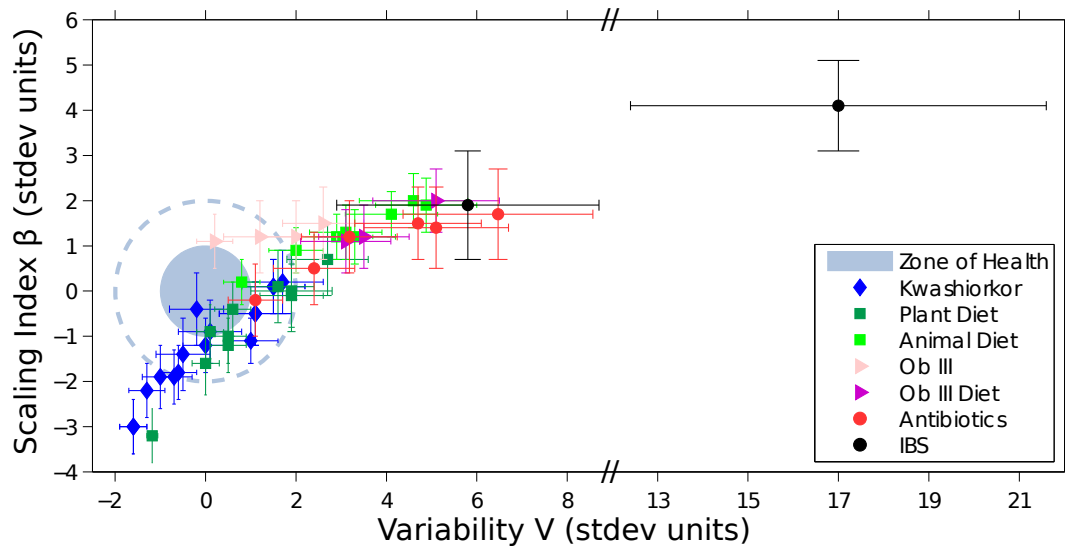
Figure 2. Taylor's law parameter space. We have compiled here all the data studied in this work. The coloured circle corresponds to 68% confidence level (CL) region of healthy individuals in the Taylor parameter space, while dashed line delimites the 98% CL region. Points with errors place each individual 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.

115 cantly when compared to a plant-based diet. All other cases presented increased variability,

116 which is particularly severe, and statistically significant at more than 95% CL, for obese pa-

117 tients grade III on a diet, individuals taking antibiotics or IBS–diagnosed patients. A global

118 property emerges from all worldwide data collected: Taylor parameters characterize the sta-

119 tistical behavior of microbiome changes. Furthermore, we have verified that our conclusions

120 are robust to systematic errors due to taxonomic assignment (see Figure Sx in Supplemental

121 Material).

122 Taylor's power law has been explained in terms of various effects, all without general consen-

123 sus. It can be shown to have its origin in a mathematical convergence similar to the central

124 limit theorem, so virtually any statistical model designed to produce a Taylor law converge

to a Tweedie distribution (*52*), providing a mechanistic explanation based on the statistical theory of errors (*53–55*). To unveil the generic mechanisms that drive different scenarios in the β–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 (*40*). 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 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.
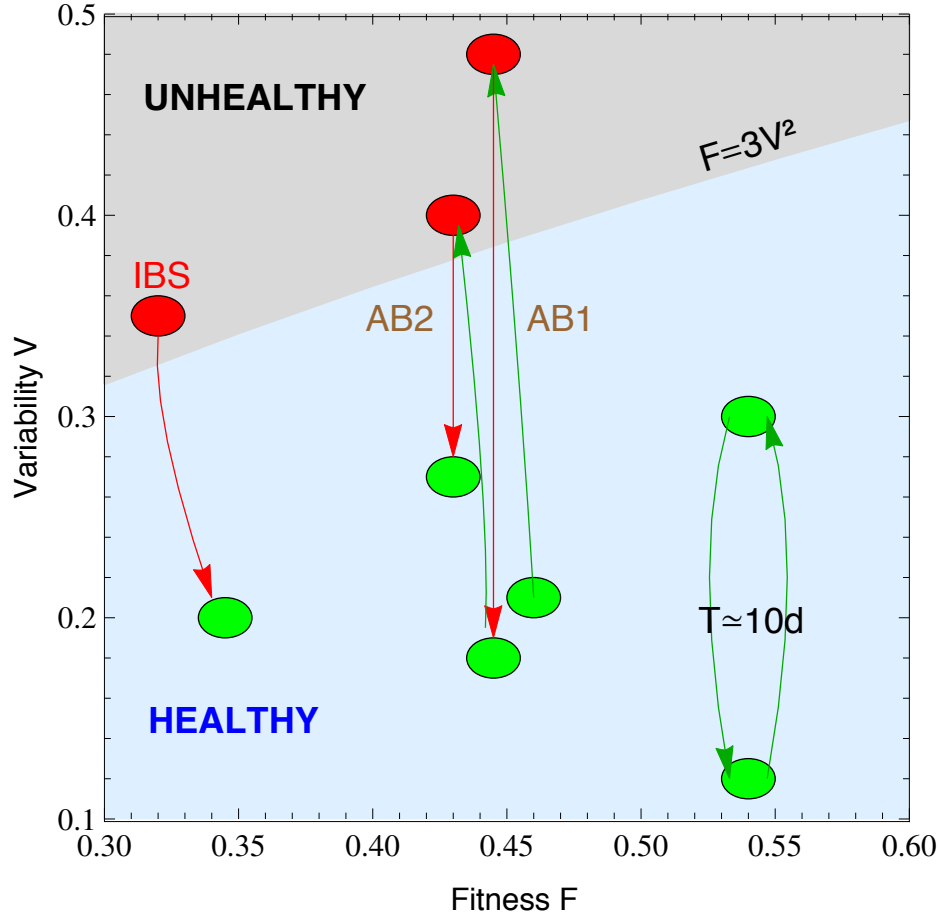
Figure 3. Microbiota states can be placed in the phase space F–V. The light blue shaded region corresponds to the stable phase, while the grey shaded region is the unstable phase (the phase transition line is calculated for $\alpha = \beta = 0.75$). We place healthy individuals (green) and individuals 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.

Rank stability of the taxa

The rank dynamics and stability plots in Figure 4 and 5 show 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*, Figure 4) and from a subject diagnosed with IBS (patient *P2*, Figure 5) of the IBS study (*12*). 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.
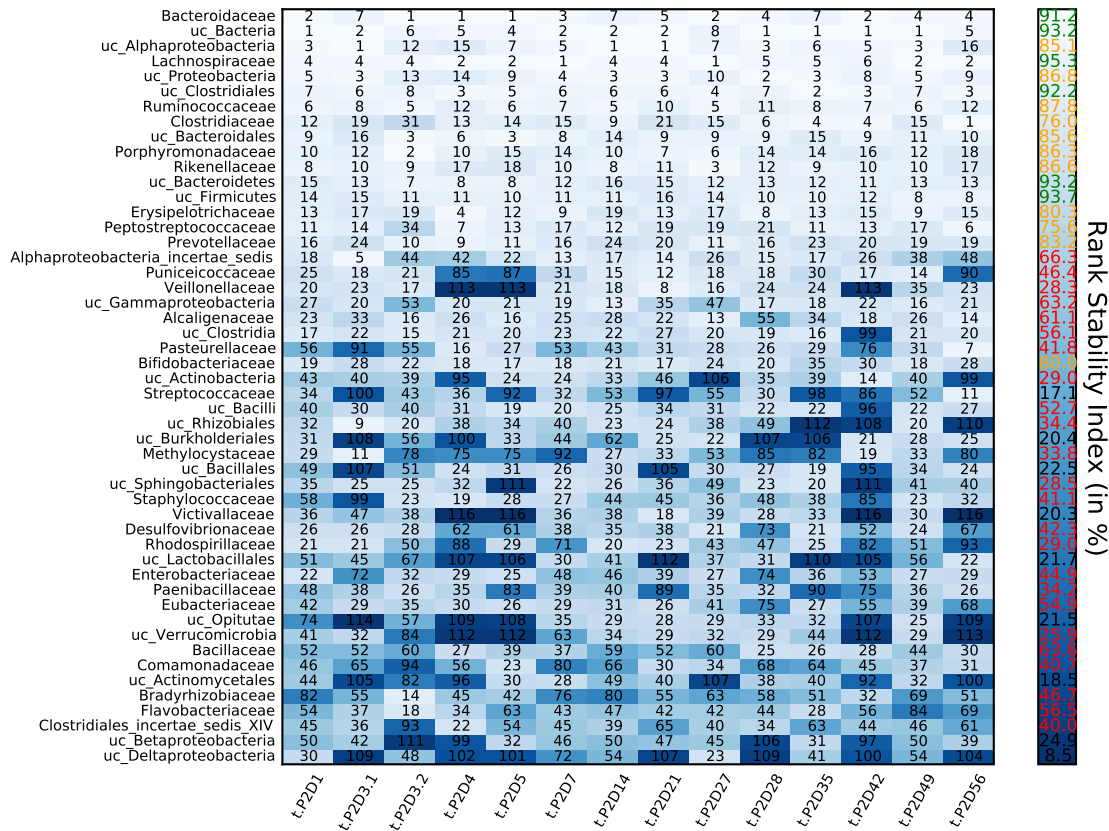
Nevertheless, in the particular case of the healthy individual in Figure 4, *Burkholderiales* and *Betaproteobacteria* (taxa ordered as 18th and 25th in the dominance axis) show comparatively very low rank stability regarding similar dominant taxa while, on the other hand, *Comamonadaceae, Lactobacillaceae, 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 the color criteria shown in the table included in Figure 4, since they show a moderately stable RSI.

In the IBS diagnosed patient of Figure 5, beyond the differences in dominance for the particular taxa, we still observe that the most dominant are the most rank stable. However, as opposed to the healthy individual results, far from presenting a *rank stability island*, the medium-ranked taxa are very rank unstable, mostly due to transient (often one or two consecutive samples) but deep drops in their relative abundance, which are usually happening more than twice along the time series. That is, for instance, the case of *Sphingobacteriales* with two non-consecutive samples dropping to 111th rank position. In other cases, the high rank instability comes from a rank fluctuation over all the time series, as for *Streptococcaceae* and *Burkholderiales*, which are ranking 26th and 29th respectively in the overall dominance

Figure 4. Rank variation throughout time for the most dominant elements (taxa) and their calculated Rank Stability Index (as shown in Material and Methods) for samples from a healthy subject studied in our lab (*12*).

	t.P2D1	t.P2D3.1	t.P2D3.2	t.P2D4	t.P2D5	t.P2D7	t.P2D14	t.P2D21	t.P2D27	t.P2D28	t.P2D35	t.P2D42	t.P2D49	t.P2D56	RSI %
Bacteroidaceae	2	7	1	1	1	3	7	5	2	4	7	2	4	4	91.2
uc_Bacteria	1	2	6	5	4	2	2	2	8	1	1	1	1	5	93.2
uc_Alphaproteobacteria	3	1	12	15	7	5	1	1	7	3	6	5	3	16	85.1
Lachnospiraceae	4	4	4	2	2	1	4	4	1	5	5	6	2	2	95.3
uc_Proteobacteria	5	3	13	14	9	4	3	3	10	2	3	8	5	9	86.8
uc_Clostridiales	7	6	8	3	5	6	6	6	4	7	2	3	7	3	92.2
Ruminococcaceae	6	8	5	12	6	7	5	10	5	11	8	7	6	12	87.8
Clostridiaceae	12	19	31	13	14	15	9	21	15	6	4	4	15	1	76.0
uc_Bacteroidales	9	16	3	6	3	8	14	9	9	9	15	9	11	10	85.6
Porphyromonadaceae	10	12	2	10	15	14	10	7	6	14	14	16	12	18	86.6
Rikenellaceae	8	10	9	17	18	10	8	11	3	12	9	10	10	17	86.6
uc_Bacteroidetes	15	13	7	8	8	12	16	15	12	13	12	11	13	13	93.2
uc_Firmicutes	14	15	11	11	10	11	11	16	14	10	10	12	8	8	93.7
Erysipelotrichaceae	13	17	19	4	12	9	19	13	17	8	13	15	9	15	80.5
Peptostreptococcaceae	11	14	34	7	13	17	12	19	19	21	11	13	17	6	75.6
Prevotellaceae	16	24	10	9	11	16	24	20	11	16	23	20	19	19	83.2
Alphaproteobacteria_incertae_sedis	18	5	44	42	22	13	17	14	26	15	17	26	38	48	66.3
Puniceicoccaceae	25	18	21	85	87	31	15	12	18	18	30	17	14	90	46.4
Veillonellaceae	20	23	17	113	113	21	18	8	16	24	24	113	35	23	28.3
uc_Gammaproteobacteria	27	20	53	20	21	19	13	35	47	17	18	22	16	21	63.2
Alcaligenaceae	23	33	16	26	16	25	28	22	13	55	34	18	26	14	61.1
uc_Clostridia	17	22	15	21	20	23	22	27	20	19	16	99	21	20	56.1
Pasteurellaceae	56	91	55	16	27	53	43	31	28	26	29	76	31	7	41.8
Bifidobacteriaceae	19	28	22	18	17	18	21	17	24	20	35	30	18	28	53.0
uc_Actinobacteria	43	40	39	95	24	24	33	46	106	35	39	14	40	99	29.0
Streptococcaceae	34	100	43	36	92	32	53	97	55	30	98	86	52	11	17.1
uc_Bacilli	40	30	40	31	19	20	25	34	31	22	22	96	22	27	52.7
uc_Rhizobiales	32	9	20	38	34	40	23	24	38	49	112	108	20	110	34.4
uc_Burkholderiales	31	108	56	100	33	44	62	25	22	107	106	21	28	25	20.4
Methylocystaceae	29	11	78	75	75	92	27	33	53	85	82	19	33	80	33.8
uc_Bacillales	49	107	51	24	31	26	30	105	30	27	19	95	34	24	22.5
uc_Sphingobacteriales	35	25	25	32	111	22	26	36	49	23	20	111	41	40	28.5
Staphylococcaceae	58	99	23	19	28	27	44	45	36	48	38	85	23	32	41.1
Victivallaceae	36	47	38	116	116	36	38	18	39	28	33	116	30	116	20.3
Desulfovibrionaceae	26	26	28	62	61	38	35	38	21	73	21	52	24	67	42.5
Rhodospirillaceae	21	21	50	88	29	71	20	23	43	47	25	82	51	93	29.0
uc_Lactobacillales	51	45	67	107	106	30	41	112	37	31	110	105	56	22	21.7
Enterobacteriaceae	22	72	32	29	25	48	46	39	27	74	36	53	27	29	41.9
Paenibacillaceae	48	38	26	35	83	39	40	89	35	32	90	75	36	26	34.2
Eubacteriaceae	42	29	35	30	26	29	31	26	41	75	27	55	39	68	34.0
uc_Opitutae	74	114	57	109	108	35	29	28	29	33	32	107	25	109	21.5
uc_Verrucomicrobia	41	32	84	112	112	63	34	29	32	29	44	112	29	113	22.9
Bacillaceae	52	52	60	27	39	37	59	52	60	25	26	28	44	30	43.0
Comamonadaceae	46	65	94	56	23	80	66	30	34	68	64	45	37	31	40.7
uc_Actinomycetales	44	105	82	96	30	28	49	40	107	38	40	92	32	100	18.5
Bradyrhizobiaceae	82	55	14	45	42	76	80	55	63	58	51	32	69	51	46.7
Flavobacteriaceae	54	37	18	34	63	43	47	42	42	44	28	56	84	69	58.5
Clostridiales_incertae_sedis_XIV	45	36	93	22	54	45	39	65	40	34	63	44	46	61	40.0
uc_Betaproteobacteria	50	42	111	99	32	46	50	47	45	106	31	97	50	39	24.9
uc_Deltaproteobacteria	30	109	48	102	101	72	54	107	23	109	41	100	54	104	8.5

Rank Stability Index (in %)

Colour code for the RSI percentage column

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

Figure 5. Rank variation throughout time for the most dominant elements (taxa) and their calculated Rank Stability Index for samples from a subject diagnosed with irritable bowel syndrome studied in our lab (*12*).

axis but show very low RSI, and thus colored in black attending to the color criteria shown in the table included in Figure 5.

We found the presence of such of *rank stability island* for medium-ranked taxa in the other healthy subjects (*B* and *C*) of the IBS study (*12*) together with its total absence for the other IBS diagnosed patient (patient *P1*), which also presents very high rank instability in its medium-ranked taxa.

Time dependence of model parameters

Finally, we have studied the time dependence of the variability V and power law index β (see Model under Material and Methods) by using a sliding window approach. The total number of time points are divided in subsets of five points, where the next subset is defined by adding the next time sampling and by eliminating the earliest one. Both parameters were calculated for each subset against the average time lapse. Figure 6 shows the variability V as a function of time for the largest sampling: two individuals in the Caporaso's study (*41*) corresponding to the gut microbiota of a male (upper plot) and a female (lower plot). Figure 7 shows the time evolution of V for patient P2 of the IBS study (*12*) (upper plot) and patient D in the antibiotics study (*45*) (lower plot). Both samples shows 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. 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 leaks 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.
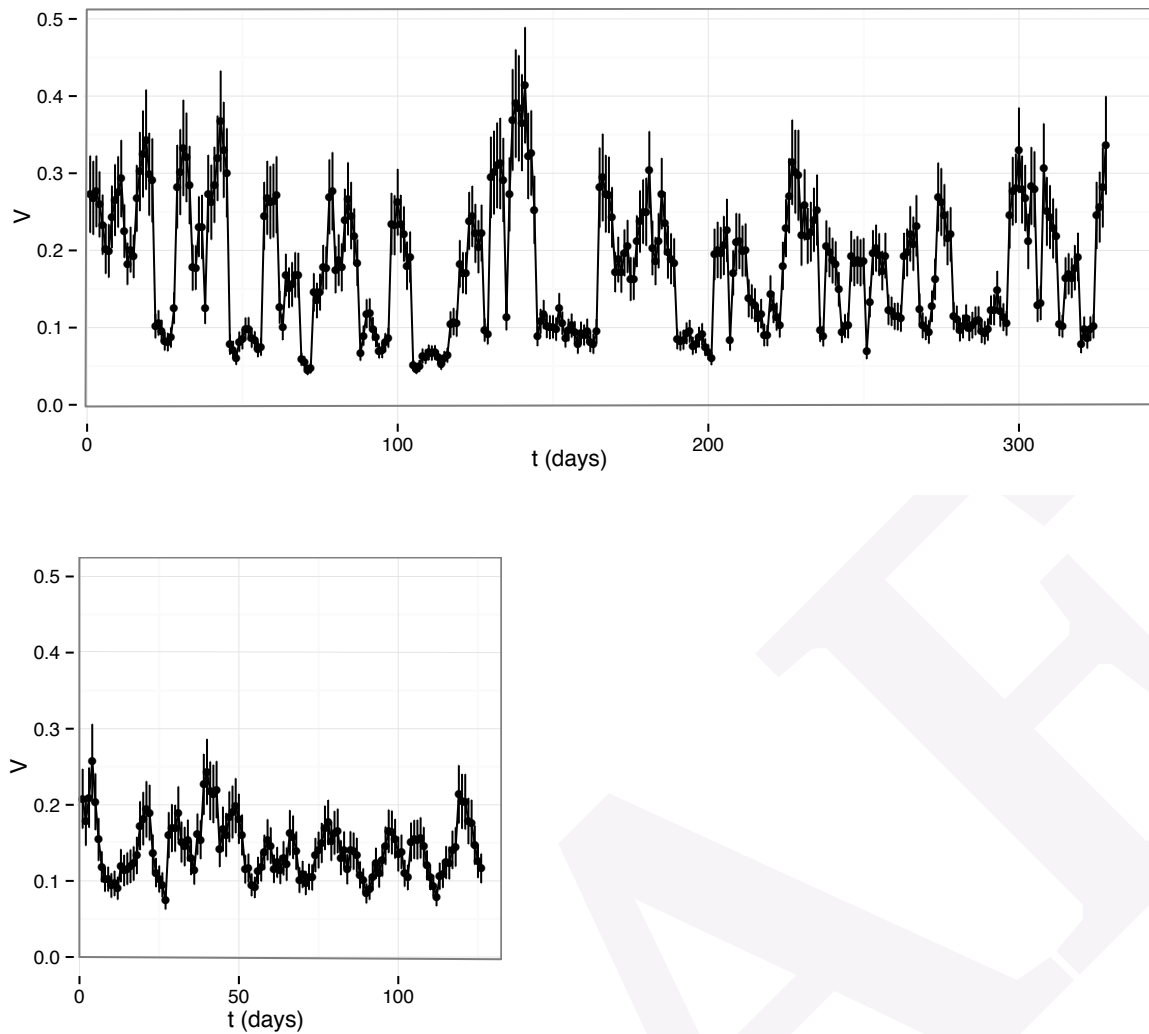
Figure 6. *V* as a function of time for the two individuals in the Caporaso's study (*41*): samples

of gut microbiome of a male (upper plot) and a female (lower plot).
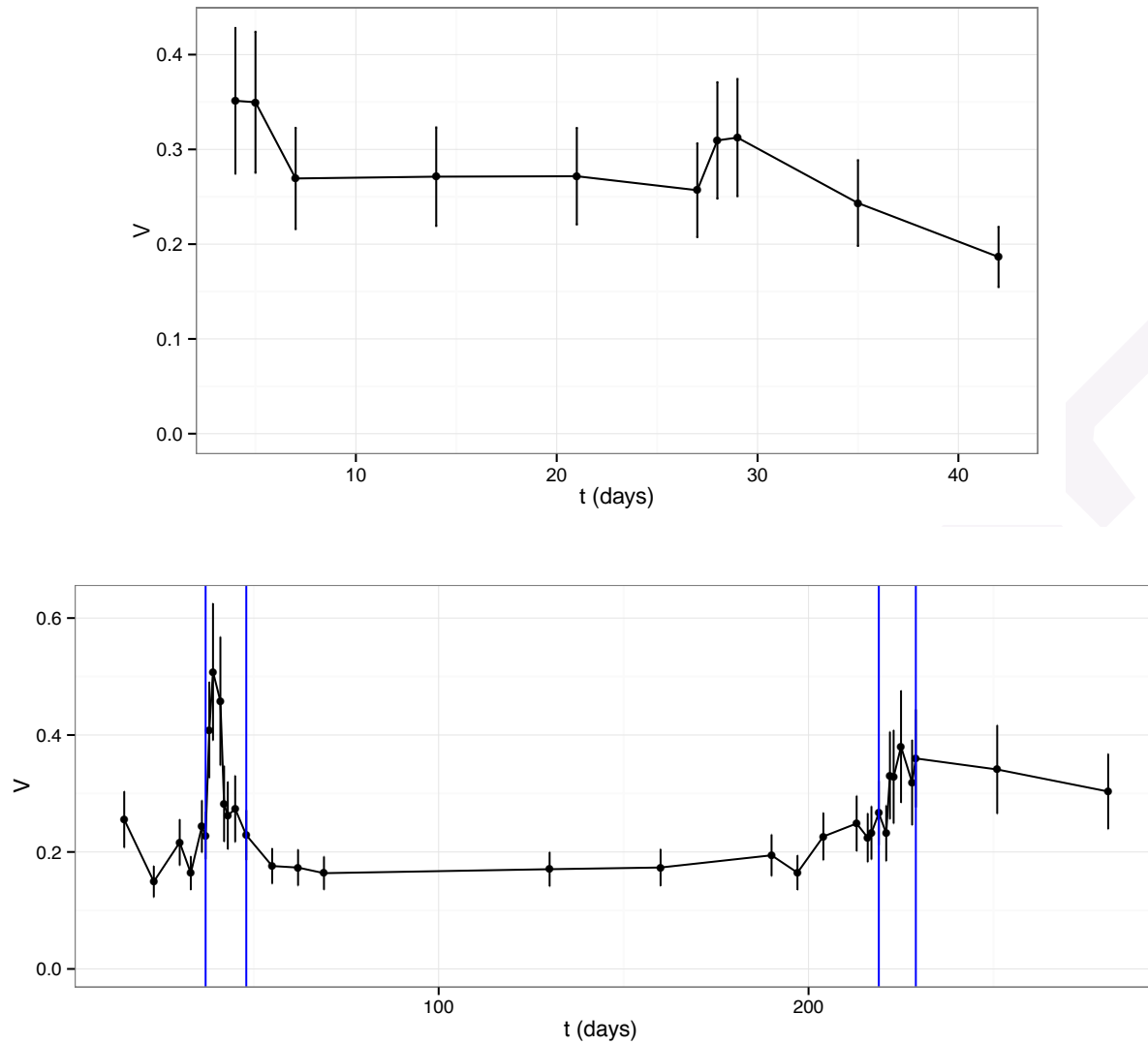
Figure 7. *V* as a function of time for patient P2 of the IBS study (*12*) (upper plot) and patient D in the antibiotics study (*45*) (lower plot). The blue vertical lines in the lower plot are showing the periods of antibiotic treatment.

Discussion

One of the mains features of this work is to have shown that, independently of its condition, the microbiota follows the Taylor's law. We have seen that the value of the scaling index in each case is always less than the unity (using standard deviation as the measurement for dispersion), which is informing us about the community structure. This means that, in relative terms, the most abundant elements in the population are less volatile to perturbations than the less abundant ones. The explanation for this universal pattern is not clear although some hypothesis have been tested in other studies, as the presence of negative interactions in the population (*56*), or the demonstration that it may depend on reproductive correlation (*57*). Nevertheless, none of these explanations are enough when we are talking about microbiota as the reproduction term is diffuse, the interactions between its components are not only based on competition (*58–60*) and that even that kind of negative interaction may not effectively yield in values less than the unity when referring to a bacterial species (*36*). Anyhow, the values obtained in all cases are very similar among them, which could be suggesting that the community structure is preserved throughout the different scenarios that we have studied.

The second parameter is informing about the noise and can be directly related with the variability or the fluctuation amplitude of the population over time. It is a direct estimator of the stability of the system under study. As we showed above, the healthy subset of each study have lower variability than the non-healthy subset when dealing with adult individuals. Interestingly, the variability parameter was higher in the healthy subset for the study of the discordant twins suffering from kwashiorkor disease (*43*). Taking into account that the infant microbiota is evolving toward a definite, adult state (*62*), it means that the temporal variability would be greater than in a adult who has reached a stability in his gut microbiota, while our results could be directing in the possibility that this variability is necessary in order to reach that adult state `<-review the wording of the whole sentence`. Furthermore, as we wanted to see how this variability behaved over time, we calculated the evolution of

16

this parameter for the samples which had enough time sampling. As can be seen in Figure 6, the variability of the microbiota has some fluctuations over time. It is interesting to note in Figure 7 how this parameter can capture the two antibiotic intakes in one of the patients from the study of Dethlefsen and Relman (*45*), especially that it seems to be some resilience process in the microbiota due to the lower variability increase in the second antibiotic intake.

The primary hypothesis of this work is that, in adults, having a healthy microbiota means that population is stable in time and does not have huge flips or jumps into another states. In order to use the valuable information which gives us the empirical law of Taylor's work, we propose the use of Langevin equation to model how the ranking stability evolves in time. While we can measure directly the component of the noise of the system as their variability, the other main term needs to be inferred from the model. This term, which we have named as 'fitness', is the one that gives the ability to the system to be stable to potential perturbations. In ecological terms, this could mean the nature of interactions that are present among the bacteria, between bacteria and other minority populations as fungi or archaea, between bacteria and the viral component in the microbiota, and the interactions between host and the whole microbiota. Being this a first step to model the temporal stability of the microbiota and due to its complicated nature, we have calculated the fitness term using the Fluctuation Dissipation Theorem as a first approximation (*63*). Thus, the fitness of the microbiota still remains to be modelled in future works in order to make the model more accurate and with a higher predictive power.

By solving the Langevin differential equation, we can obtain a phase diagram where each microbiota sample can be placed according to its fitness and variability in one of two phases according to the ranking stability of the system. As we can see in the phase-space in Figure 3, we are showing three different conditions that could happen. First, we can have a healthy microbiota which could have some fluctuations as showed by one of the subjects of Caporasso et al study (*41*). Because the fitness of this cases will be high enough, the temporal variability

will not place the microbiota in the unstable phase of the diagram. Second, we have a subject from the study of Dethlefsen and Relman (*45*) which is perturbed twice by an antibiotic intake. His microbiota is altered enough to lose its stability, and hence be placed in the unstable part. So located, it is more sensible to potential perturbations as, for example, opportunist infections. Third and last condition, the subject is already in the unstable phase due to some healthy issue as IBS. This can be observed in one of the patients from Durban et al study (*12*). In addition, it was shown that this subject improved its healthy status in the time when the experiment was done, implying that his microbiota also recovered the lost stability.

Specifically, the analysis of the rank stability of the samples of healthy and IBS diagnosed patients studied in our lab (*12*), suggests that the presence of *rank stability islands* among medium-ranked taxa is a interesting feature. The higher stability of these taxa goes against the global meaning of the scaling index. Interestingly, that stability disappears when we look at the IBS patients. One could ask if these taxa are key players in the phase transition of the microbiota, of if they are more susceptible to perturbations than the most abundant. The types of interactions that could be sustaining this particular behavior are not clear, as these non-abundant taxa are not usually included in dynamical studies in order to get the community matrix. Further experiments and data analysis is needed to clarify if this is not a unique event, or it is a widespread feature of stable microbiotas.

However, we have to be aware that the hypothesis above is too simplistic to be directly related with reality. It has been demonstrated that the situation is more complex than the outlook provided separating healthy people from non-healthy people just by compositional terms, as Moya and Ferrer underline in their recent review (*17*). There are several different feasible scenarios in which we can consider the microbiota as stable independently of their compositional evolution over time. For example, depending on their ability to recover the initial composition (resilience), or whether it can recover the original function despite the compo-

274 sition (functional redundancy). What we have shown in this work could be explained as the

275 transitions of a stable microbiota into a dysbiosis state.

276 As a first step toward understanding the microbiota stability, the model presents some lim-

277 itations and there is still work to do. From the biological perspective, many questions arise

278 from this work. We have observed the same pattern in Taylor's parameters in all the different

279 conditions we have studied, but a pertinent question is whether it is really a universal feature

280 in the huge diversity of microbial niches. Furthermore, another relevant question is which

281 mechanisms are involved in maintaining the population structure. The nature of the interac-

282 tions among the elements of the community is surely of great importance in this matter, and

283 it is related to the fitness of the community as has been commented above. How we should

284 address the community fitness is not clear, but works as Tikhonov's (*64*) could help us to aim

285 at the correct direction toward unraveling the complexity of the microbiota.

286 ## Materials and Methods

287 ## Model

288 We model the microbial abundances across time along the lines of Blumm *et al.* (*40*). The

289 dynamics of taxon relative abundances is described by the Langevin equation:

290
$$\dot{x}_i = F_i \cdot x_i^\alpha + V \cdot x_i^\beta \xi_i(t) - \phi(t) \cdot x_i,$$

291 where F_i captures the fitness of the taxon i, V corresponds to the noise amplitude and $\xi_i(t)$

292 is a Gaussian random noise with zero mean $< \xi_i(t) > \, = 0$ and variance uncorrelated in

293 time, $< \xi_i(t)\xi_i(t') > \, = \delta(t'-t)$, . The function $\phi(t)$ ensures the normalization at all times,

294 $\sum x_i(t) = 1$, and corresponds to $\phi(t) = \sum F_i x_i^\alpha + \sum V x_i^\beta \xi_i(t)$. The temporal evolution of

19

the probability that a taxon i has a relative abundance $x_i(t)$, $P(x_i,t)$, is determined by the Fokker-Planck equation:

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

The microbiota evolves towards a steady-state with a time-independent probability depending on the values of α, β, 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:

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

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

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 for the parameters of interest. The ordered phase happens when the solution has a maximum in the physical interval ($0 < x_i < 1$). For larger V, the transition to a disordered phase happens when the maximum shifts to the unphysical region $x_i < 0$, which sets the phase transition region $V(\alpha,\beta,F_i)$. The phase transition region can be calculated analytically in particular cases:

$$F_i^2 = 4\beta\phi_0 V^2 \quad \text{if} \quad \beta = \alpha \neq 1,$$

$$F_i = \beta V^2 \quad \text{if} \quad 2\beta = 1+\alpha,$$

where the first case, simplifies to $F = 3V^2$ if $\beta = 0.75$ and the fitness of this taxon dominates in ϕ_0. In many physical systems (Brownian motion is the classical example), the two terms of the Langevin equation are related. The *Fluctuation Dissipation Theorem* states a general

20

relationship between the response to an external disturbance and the internal fluctuations of the system (*63*). The theorem can be used as the basic formula to derive the fitness from the analysis of fluctuations of the microbiota, assuming that it is in equilibrium (the ordered phase).

Explain better the Fluctuation Dissipation Theorem

Selection and Methods

The bacteria and archaea taxonomic assignations were obtained by analysing 16S rRNA sequences, which were clustered into operational taxonomic units (OTUs) sharing 97 % sequence identity using QIIME (*47*). WGS data (*43*) were analysed and assigned at strain level by the Livermore Metagenomic Analysis Toolkit (LMAT) (*49*), according to their default quality threshold. Genus, with best balance between error assignment and number of taxa, was chosen as our reference taxonomic level. We have verified that our conclusions are not significantly affected by selecting family or species as the reference taxonomic level (see Figure 8).

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. Caporaso et al. study (*41*) was selected as it has two healthy individuals measured over a very large timespan, with almost daily sampling. Faith et al. study (*42*) was selected due to the BMI differences between subjects, moreover some of them had diets which can be treated as system perturbations. We considered healthy only those individuals who had normal or overweight BMI. Smith et al., study (*43*) was selected for both the age

of the patients and the rare disease. We only worked with the discordant twins, and considered healthy those who were not affected by kwashiorkor in each pair of patients. David et al., study (*44*) was selected for its differential diets. We considered as the healthy part the first time samples of each individual before the diet, and the rest of time points as perturbations. Dethlefsen and Relman work (*45*) was selected due to the interesting treatment of two antibiotic intakes of the same antibiotic by three different subjects. We considered as the healthy part only those times before any antibiotic treatment, and as perturbations the time of antibiotic intakes and the period after that. And finally, we also considered a study made in our group carried by Durban et al., (*12*) in which we considered as healthy subjects those who didn't suffer from irritable bowel disease, and as perturbation the patients who had this disease.

Metadata of each study is provided in Supplementary Tables S1 to S4. All used 16S rRNA gene sequencing except for the study of the discordant kwashiorkor twins (*43*), where both 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 (*46*), 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 (*47*) (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 (*48*) (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,

364 max rejects 500.

Metagenomic sequences processing

366 Shotgun metagenomic (and 16S too) sequences were analyzed with LMAT (Livermore Metage-
367 nomics Analysis Toolkit) software package (*49*) (version 1.2.4, with Feb'15 release of data
368 base *LMAT-Grand*). LMAT was run using a Bull shared-memory node belonging to the team's
369 HPC (high performance computing) cluster. It is equipped with 32 cores (64 threads available
370 using Intel Hyper-threading technology) as it has two Haswell-based Xeons (22 nm technol-
371 ogy), the E5-2698v3@2.3 GHz, sharing half a tebibyte (0.5 TiB, that is, 512 gibibytes) of
372 DRAM memory. This node is also provided with a PCIe SSD card as NVRAM, the Micron
373 P420m HHHL, with 1.4 TB, and 750000 reading IOPS, 4 KB, achieving 3.3 GB/s. The com-
374 puting node was supplied with a RAID-0 (striping) scratch disk area. We used the "Grand"
375 database (*50*), release Feb'15, provided by the LMAT team, where "Grand" refers to a huge
376 database that contains k-mers from all viral, prokaryote, fungal and protist genomes present
377 in the NCBI database, plus Human reference genome (hg19), plus GenBank Human, plus
378 the 1000 Human Genomes Project (HGP) (this represent about 31.75 billion k-mers occu-
379 pying 457.62 GB) (*50*). Previously to any calculation, the full database was loaded in the
380 NVRAM. With this configuration the observed LMAT sustained sequence classification rate
381 was 20 kpb/s/core. Finally, it is worth mentioning that a complete set of Python scripts have
382 been developed as back-end and front-end of the LMAT pipeline in order to manage the added
383 complexity of time series analysis.

Taxa level selection

385 We selected genus as taxonomic level for the subsequent steps of our work. In order to ensure
386 that, between adjacent taxonomic levels, there were not crucial differences which could still
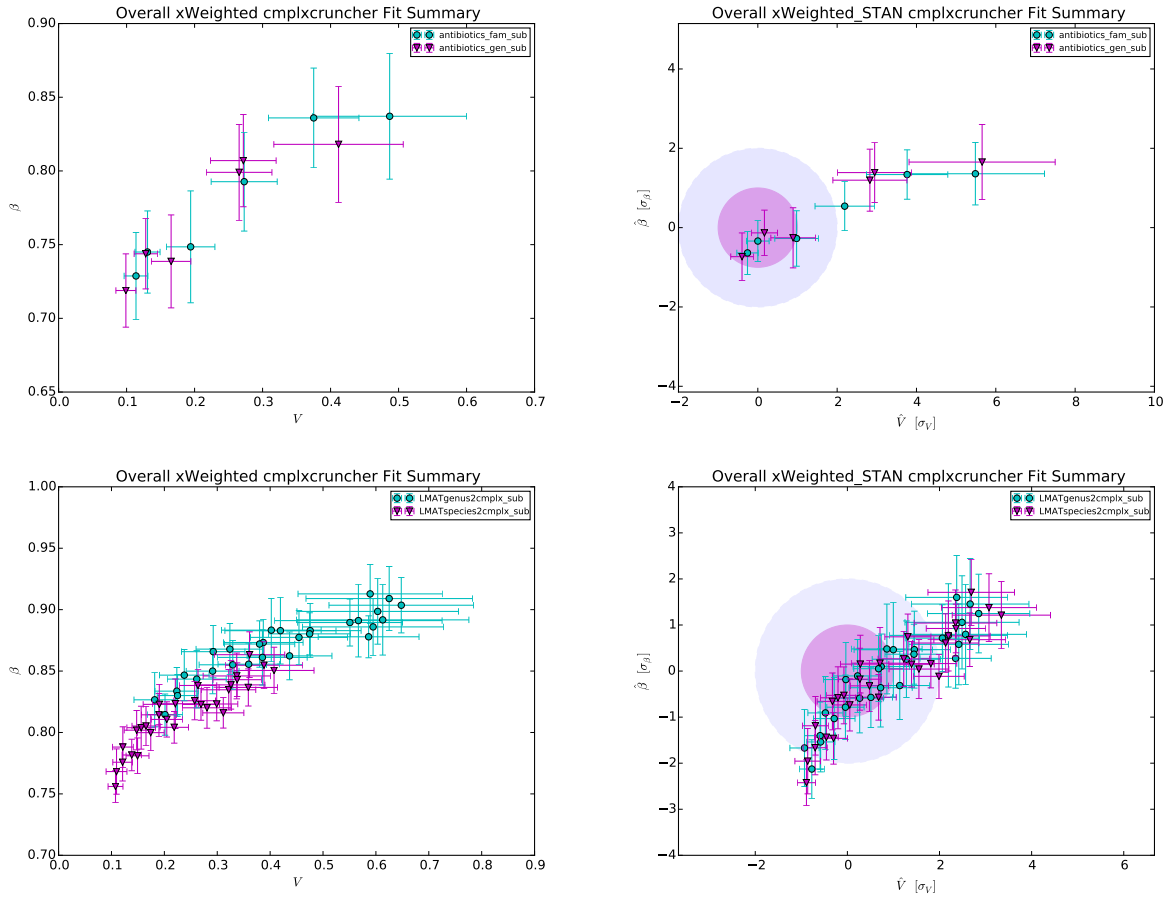
Figure 8. Overview of comparison of different approaches based on adjacent taxonomic levels using plots in the Taylor-parameters space. For 16S (former row of subfigures), the levels are family vs. genus, whereas for SMS (latter row of subfigures) levels are genus vs. species. The left column shows the 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 (*45*) with 16S data, we tested the differences between genus and family levels. The latter dataset tested was the kwashiorkor discordant twins study (*43*) for both genus and species taxonomic levels. The Figures 8 (overview) and 9 (detail) plot the comparison between studies (and so, 16S and SMS) and between adjacent taxonomic levels.
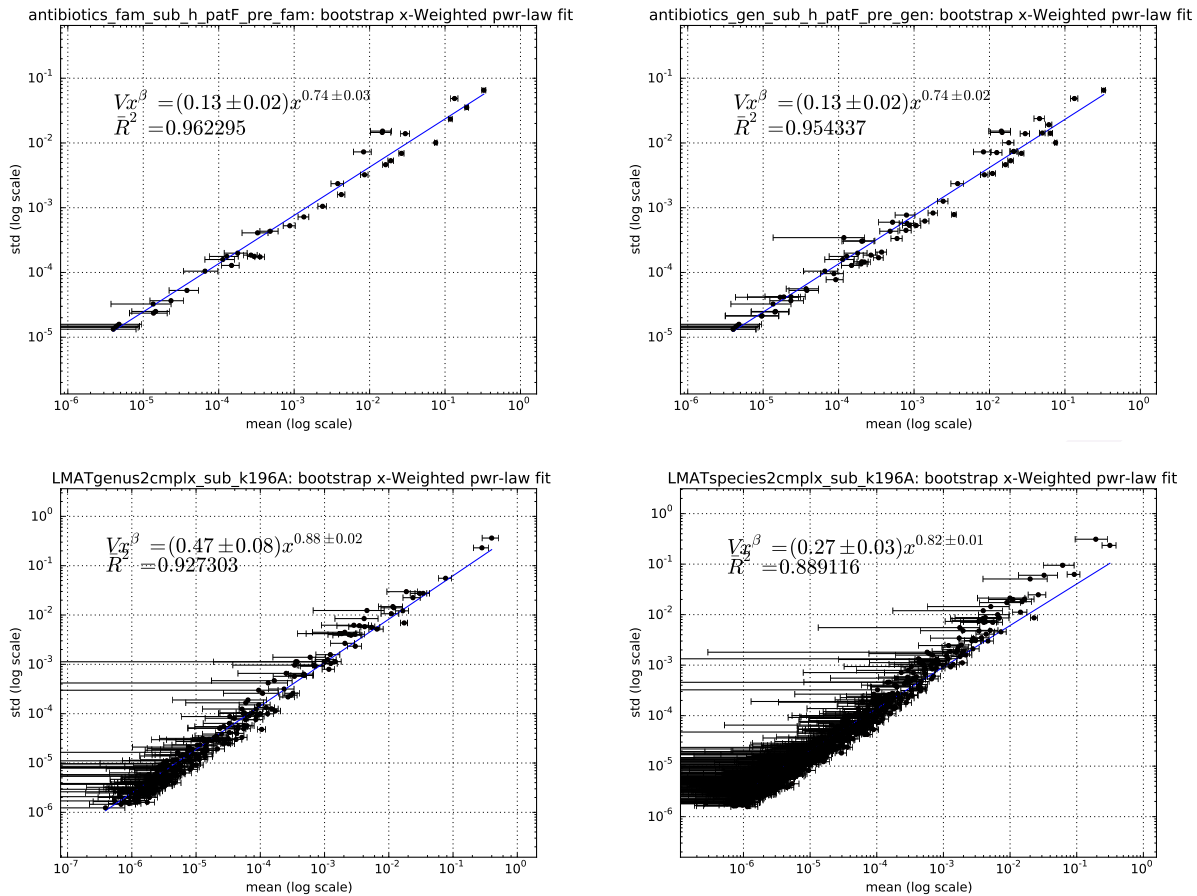
Figure 9. Detail of comparison of different approaches based on adjacent taxonomic levels using plots of X-weighted power-law fits (see Material and Methods). The former row of subfigures shows examples for 16S, whereas the latter row of subfigures plots examples for SMS. The left column shows results 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).

X-weighted power-law fit

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

The basic idea of bootstrapping is that inference about a population from sample data (sample → population) can be modeled by resampling the sample data and performing inference on (resample → sample) (*66*). To adapt this general idea to our problem, we resample the x-data array using its errors array. That is, for each replicate, a new x-data array is computed based on:

$$x_i^* = x_i + v_i$$

where v_i is a Gaussian random variable with mean $\mu_i = 0$ and standard deviation $\sigma_i = \text{SEM}_i$, as defined previously. For each replicate a complete un-weighted power-law fit is performed, where to choose between fitting power laws ($y = Vx^\beta$) using linear regression on log-transformed (LLR) data versus non-linear regression (NLR) we mainly follow *General Guidelines for the Analysis of Biological Power Laws* (*65*). The parameters of the X-weighted fit are then estimated by averaging through all the replicate fits performed, and their errors are estimated by computing the standard deviation also for all the fits. At the end of each step, the relative error is calculated by comparing the fit parameters estimation in the last step with the previous one. Finally, both the coefficient of determination of the fit and the coefficient

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 Figures 4 and 5. 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.

Standardization

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

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

$$\widehat{V} = \frac{1}{W_1} \sum_{i=1}^{h} V_i \omega_i = \sum_{i=1}^{h} V_i \omega_i$$

438   as $W_1 = \sum_i^h \omega_i = 1$, since ω_i are normalized weights calculated as:

439
$$\omega_i = \frac{\frac{1}{\sigma_{V_i}^2}}{\sum_i^h \frac{1}{\sigma_{V_i}^2}}$$

440   being σ_{V_i} the estimation of the uncertainty in V_i obtained together with V_i from the X-weighted

441   power-law fit described in Section , for healthy individuals.

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

443
$$\widehat{\sigma}_V = \sqrt{\frac{1}{W_1 - \frac{W_2}{W_1}} \sum_{i=1}^h \left[\omega_i \left(V_i - \hat{V} \right)^2 \right]}$$

444   being $W_2 = \sum_i^h \omega_i^2$, which finally yields to:

445
$$\widehat{\sigma}_V = \sqrt{\frac{1}{1 - \sum_i^h \omega_i^2} \sum_{i=1}^h \left[\omega_i \left(V_i - \hat{V} \right)^2 \right]}$$

Acknowledgments

Funding Information

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. **Moran NA, Sloan DB.** 2015. The Hologenome Concept: Helpful or Hollow? PLoS Biol **13**:1–10.

4. **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.

5. **Spencer 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.

6. **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.

7. **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.

8. **Kimura I, Ozawa K, Inoue D, Imamura T, Kimura K, Maeda T, Terasawa K, Kashihara 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.

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

10. **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.

11. **Brown JM, Hazen SL.** 2015. The Gut Microbial Endocrine Organ: Bacterially Derived Signals Driving Cardiometabolic Diseases. Annu Rev Med **66**:343–359.

12. **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.

13. **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-naive microbiome in new-onset Crohn's disease. Cell Host Microbe **15**:382–392.

14. **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.

15. **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.

16. **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.

17. **Moya A, Ferrer M.** 2016. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends Microbiol **24**:402–413.

18. **Giloteaux L, Goodrich JK, Walters WA, Levine SM, Ley RE, Hanson MR**. 2016. Reduced diversity and altered composition of the gut microbiome in individuals with myalgic encephalomyelitis/chronic fatigue syndrome. Microbiome. **4**:30

19. **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.

20. **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, Eeckhaudt L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J.** 2016. Population-level analysis of gut microbiome variation. Science (80-) **352**:560–564.

21. **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.

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

23. **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.

24. **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.

25. **Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, Liu Y-Y.** 2016. Universality of human microbial dynamics. Nature **534**:259–262.

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

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

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

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

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

561 31. **Anderson RM, Gordon DM, Crawley MJ, Hassell MP.** 1982. Variability in the abun-
562 dance of animal and plant species. Nature **18**: 245–248

563 32. **Živković J, Tadić B, Wick N, Thurner S.** 2006. Statistical indicators of collective be-
564 havior and functional clusters in gene networks of yeast. Eur Phys J B **50**:255–258.

565 33. **Kendal WS.** 2003. An Exponential Dispersion Model for the Distribution of Human
566 Single Nucleotide Polymorphisms. Mol Biol Evol **20**:579–590.

567 34. **Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, Ma ZS, Shi P.** 2014. Spatial het-
568 erogeneity and co-occurrence patterns of human mucosal-associated intestinal micro-
569 biota. ISME J **8**:881–93.

570 35. **Kaltz O, Escobar-Paramo P, Hochberg M, Cohen** JE. 2012. Bacterial microcosmos
571 obey Taylor's law: Effects of abiotic and biotic stress and genetics on mean and variance
572 of population density. Ecol Process **1**:5.

573 36. **Ramsayer J, Fellous S, Cohen JE, Hochberg ME.** 2012. Taylor's Law holds in exper-
574 imental bacterial populations but competition does not influence the slope. Biol Lett
575 **8**:316–319.

576 37. **Pérez-Cobas AE, Artacho A, Ott SJ, Moya A, Gosalbes MJ, Latorre A.** 2014. Struc-
577 tural and functional changes in the gut microbiota associated to Clostridium difficile
578 infection. Front Microbiol **5**:1–15.

579 38. **Ding T, Schloss PD.** 2014. Dynamics and associations of microbial community types
580 across the human body. Nature **509**:357–360.

39. **Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, Koenig SSK, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J.** 2012. Temporal dynamics of the human vaginal microbiota. Sci Transl Med **4**:132ra52.

40. **Blumm N, Ghoshal G, Forró Z, Schich M, Bianconi G, Bouchaud J-P, Barabási A-L.** 2012. Dynamics of Ranking Processes in Complex Systems. Phys Rev Lett **109**:128701.

41. **Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R.** 2011. Moving pictures of the human microbiome. Genome Biol **12**:R50.

42. **Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI.** 2013. The long-term stability of the human gut microbiota. Science (80-) **341**:1237439.

43. **Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS, Concannon P, Mychaleckyj JC, Liu J, Houpt E, Li J V, Holmes E, Nicholson J, Knights D, Ursell LK, Knight R, Gordon JI.** 2013. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science **339**:548–54.

44. **David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling A V, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ.** 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature **505**:559–63.

45. **Dethlefsen L, Relman DA.** 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci **108**:4554–61.

46. **Gordon, A., Hannon, G.J.** 2010. FASTX-Toolkit. FASTQ/A shortreads pre-processing tools v0.0.13. http://hannonlab.cshl.edu/fastx_toolkit/ (last accessed 26 Jul 2016).

47. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK,**

Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley G a, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone C a, Mcdonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters W a, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. correspondence QIIME allows analysis of high- throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. Nat Publ Gr **7**:335–336.

48. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO**. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Acids Res. **41**:D590-D596

49. **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.

50. **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.

51. **Eisler Z, Bartos I, Kertész J.** 2008. Fluctuation scaling in complex systems: Taylor's law and beyond1. Adv Phys **57**:89–142.

52. **Jørgensen B, Martinez JR, Tsao M.** 1994. Asymptotic behaviour of the variance function. Scand J Stat **21**:223–243.

53. **Fronczak A, Fronczak P.** 2010. Origins of Taylor's power law for fluctuation scaling in complex systems. Phys Rev E **81**:066112.

54. **Kendal, W.S., Jorgensen,B.** Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. Phys. Rev. E **83** 066115.

55. **Kendal WS, Jørgensen B.** 2011. Tweedie convergence: A mathematical basis for Taylor's power law, 1/f noise, and multifractality. Phys Rev E **84**:066120.

56. **Kilpatrick a M, Ives a R.** 2003. Species interactions can explain Taylor's power law for ecological time series. Nature **422**:65–68.

57. **Ballantyne IV F, J. Kerkhoff A.** 2007. The observed range for temporal mean-variance scaling exponents can be explained by reproductive correlation. Oikos **116**:174–180.

58. **Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, Sander C, Xavier JB.** 2013. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol **9**:e1003388.

59. **Fisher CK, Mehta P.** 2014. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PLoS One **9**:e102451.

60. **Bucci V, Tzen B, Li N, Simmons M, Tanoue T, Bogart E, Deng L, Yeliseyev V, Delaney ML, Liu Q, Olle B, Stein RR, Honda K, Bry L, Gerber GK.** 2016. MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. Genome Biol **17**:121.

61. **Cohen JE, Xu M, Schuster WSF**. 2013. Stochastic multiplicative population growth predicts and interprets Taylor ' s power law of fluctuation scaling Stochastic multiplicative population growth predicts and interprets Taylor ' s power law of fluctuation scaling. Proc R Soc B Biol Sci **280**:20122955.

62. **Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE.** 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci **108**:4578–4585.

63. **Weber, J.** 1956. Fluctuation Dissipation Theorem. Phys. Rev. **101**:1620-6

64. **Tikhonov M.** 2016. Community-level cohesion without cooperation. Elife **5**.

65. **Xiao X, White EP, Hooten MB, Durham SL**. 2011. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. Ecology **92**:1887-1894

66. **Wu, C.F.J.** 1986. Jackknife, bootstrap and other resampling methods in regression analysis. (with discussions) *The Annals of Statistics* **14**:1261-1350