ARTICLES
https://doi.org/10.1038/s41564-020-0685-1
nature microbiology

Macroecological dynamics of gut microbiota

Brian W. Ji[1,4], Ravi U. Sheth[1,4], Purushottam D. Dixit[1,2], Konstantine Tchourine[1] and Dennis Vitkup[1,3]✉

The gut microbiota is now widely recognized as a dynamic ecosystem that plays an important role in health and disease. Although current sequencing technologies make it possible to explore how relative abundances of host-associated bacteria change over time, the biological processes governing microbial dynamics remain poorly understood. Therefore, as in other ecological systems, it is important to identify quantitative relationships describing various aspects of gut microbiota dynamics. In the present study, we use multiple high-resolution time series data obtained from humans and mice to demonstrate that, despite their inherent complexity, gut microbiota dynamics can be characterized by several robust scaling relationships. Interestingly, the observed patterns are highly similar to those previously identified across diverse ecological communities and economic systems, including the temporal fluctuations of animal and plant populations and the performance of publicly traded companies. Specifically, we find power-law relationships describing short- and long-term changes in gut microbiota abundances, species residence and return times, and the correlation between the mean and the temporal variance of species abundances. The observed scaling laws are altered in mice receiving different diets and are affected by context-specific perturbations in humans. We use the macroecological relationships to reveal specific bacterial taxa, the dynamics of which are substantially perturbed by dietary and environmental changes. Overall, our results suggest that a quantitative macroecological framework will be important for characterizing and understanding the complex dynamics of diverse microbial communities.

The dynamics of gut bacteria can now be monitored with high temporal resolution using 16S ribosomal RNA amplicon sequencing[1,2]. Recent longitudinal studies have revealed substantial day-to-day variability and marked long-term stability of gut microbiota[3–6]. Several studies have also identified important factors, such as host diet and lifestyle, that contribute to temporal changes in bacterial species abundances[4,7–9]. However, in contrast to macroscopic ecological communities, statistical relationships describing gut microbiota dynamics are not well understood. Macroecological approaches in ecology seek to understand global statistical relationships between species' abundances, their spatial and temporal variability, and taxonomic diversity[10–12]. Although ideas from ecology have been applied to understand static patterns of gut microbial diversity and species abundance distributions[13,14], a comprehensive and quantitative analysis of bacterial macroecological dynamics is currently missing. This makes it interesting and timely to apply macroecological approaches to describe and understand complex dynamics of microbial communities.

In the present study, we sought to investigate dynamic relationships in the gut microbiota using several densely sampled longitudinal studies in humans and mice[3,4,8]. The considered data spanned three independent investigations, and were obtained using different sample collection procedures and sequencing protocols. Bacterial abundances in these studies were tracked daily for several weeks in mice and up to a year in humans; our analysis included four healthy human individuals (A, B, M3 and F4) and six individually housed mice fed either a low-fat, plant polysaccharide (LFPP) diet or a high-fat, high-sugar (HFHS) diet. We used these data to explore short-term abundance changes and long-term drift of gut microbiota, species residence and return times, and the temporal variability of individual bacterial taxa across humans and different diet groups in mice. Collectively, the present study provides a comprehensive description of macroecological dynamics of gut microbiota.

Results

Short-term dynamics of gut microbiota. Following a quantitative framework used previously to examine the ecological dynamics of animal populations[15,16], we first investigated short-term temporal fluctuations of gut microbiota abundances. One of the most basic descriptors of bacterial population dynamics is the short-term abundance change $\mu_k(t)$, defined as the logarithm of the ratio of consecutive bacterial abundances, $\mu_k(t) = \log(X_k(t+1)/X_k(t))$, where $X_k(t)$ is the relative abundance of a bacterial operational taxonomic unit (OTU) k at time t. We calculated $\mu_k(t)$ separately for each OTU k and for each day of the considered temporal microbiota datasets; $\mu_k(t)$ values calculated in this way represent the rate of abundance changes of each OTU averaged over the course of each day. Interestingly, we found that the probability of μ averaged across all OTUs and time points closely followed a Laplace distribution (equation (1)), with a characteristic tent shape in log-transformed probabilities (Fig. 1a–c):

$$p(\mu) = \frac{1}{2b}\exp\left(-\frac{|\mu|}{b}\right) \quad (1)$$

Laplace distributions were highly similar within and between individual humans, and between humans and mice (parameter $b = 0.73 \pm 0.07$, $b = 0.82 \pm 0.1$, mean ± s.d. across all humans and LFPP mice, respectively), indicating the universality of these relationships. Moreover, the Laplace distribution closely approximated the distribution of daily abundance changes for every gut microbiota time series we analysed (see Extended Data Fig. 1), including those defined at various taxonomic resolutions (see Extended Data Fig. 1c). We note that the observed distributions are unlikely to arise due to combining time series data from multiple OTUs with vastly different abundance fluctuations (see Extended Data Fig. 1d)[17,18], or as a result of using relative bacterial abundances, that is, the compositional nature of the temporal datasets (see Extended Data Fig. 3).

[1]Department of Systems Biology, Columbia University, New York, NY, USA. [2]Department of Physics, University of Florida, Gainesville, FL, USA. [3]Department of Biomedical Informatics, Columbia University, New York, NY, USA. [4]These authors contributed equally: Brian W. Ji and Ravi U. Sheth. ✉e-mail: dv2121@cumc.columbia.edu

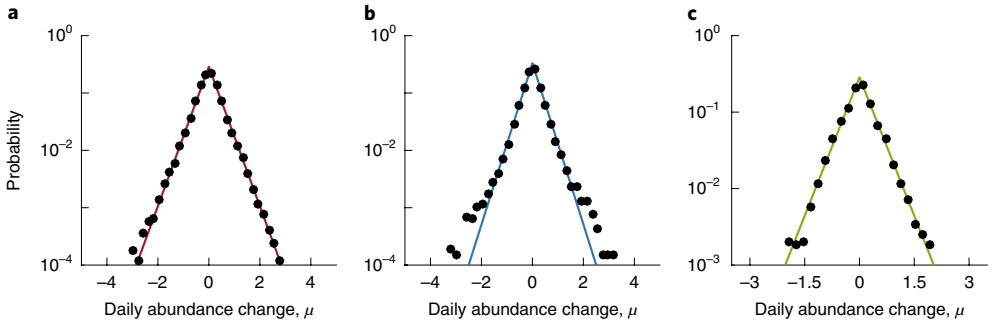


Fig. 1 | Distribution of daily abundance changes of gut bacteria. a–c, Daily abundance changes were defined as $\mu_k(t) = \log(X_k(t+1)/X_k(t))$. The probability of μ calculated over all OTUs and time points closely follows a Laplace distribution (equation (1)), with a characteristic tent shape in log-transformed probabilities. Results are shown for two individuals from different human studies (A (**a**) and M3 (**b**)) and mice fed an LFPP diet (**c**). Laplace exponents are $b = 0.83 \pm 0.1$ for human A (**a**), $b = 0.71 \pm 0.07$ for human M3 (**b**) and $b = 0.82 \pm 0.10$ for LFPP mice (**c**) (mean ± s.d.; $n = 6$ equal subsamples of the data for humans and $n = 3$ animals for mice; see Methods). Solid lines represent MLE fits to the data. Abundance changes in **c** were combined across the three mice on the LFPP diet.

In contrast to the Gaussian distribution (see Extended Data Fig. 1 and Supplementary Fig. 2 for model fits), which is expected when bacterial growth is affected by random multiplicative processes[14,19], the Laplace distribution indicates substantially higher probabilities for large short-term bacterial abundance fluctuations. Laplace distributions of abundance variability may arise due to density-dependent birth and death rates in a migrating population[20] or through emergence of subspecialized environmental niches[21]. Nevertheless, the exact mechanisms and dynamic processes generating these distributions are currently not well understood and need to be investigated further. The symmetry of the Laplace distribution suggests an equal probability for an increase or a decrease in species' abundances, which reflects a zero-sum process due to finite resources in the gut. Interestingly, Laplace distributions, describing short-term changes in species abundances or changes in other characteristic variables, have been observed across many diverse ecological and economic systems. These include species abundance changes in bird communities[15,16], fish populations[22] and tropical rain forest ecosytems[20], as well as fluctuations in sales for publicly traded companies[23] and changes in country-level gross domestic products[24] (see Extended Data Fig. 3a). Similar to these complex ecological and interacting systems, the gut microbiota may often exhibit sudden large-scale abundance fluctuations.

In many complex ecosystems, the short-term abundance fluctuations often depend on species' current abundances[15,16,23,24]. We therefore investigated the relationship between species' abundances and the corresponding standard deviation (s.d.) of daily abundance changes. To this end, for each OTU in the considered human and mouse datasets, we calculated its average abundance on each day. We then sorted all OTUs into bins based on their average abundances, and, for each abundance bin, calculated the average s.d. of daily abundance changes $\mu_k(t)$ (Fig. 2). The analysis revealed that the variability of OTU daily abundance changes decreased approximately linearly with increasing mean daily abundance (Fig. 2). This result was not due to sampling errors associated with finite sequencing depth (see Extended Data Fig. 4), and the decrease in the variability of daily abundance changes was also observed at the single OTU level (see Extended Data Fig. 5). Moreover, the observed behaviour was similar between human and mouse gut microbiota (regression slopes $r = -0.15 \pm 0.01$, -0.17 ± 0.03, mean ± s.d. across humans and mice, respectively). Thus, highly abundant bacteria exhibited substantially smaller relative daily fluctuations compared with bacteria with lower abundances, which may be due to the presence of more stable nutrient niches.

Long-term dynamics and stability of gut microbiota. In addition to short-term dynamics, interesting long-term dynamic trends have also been observed across different macroscopic ecosystems[16,17,25]. To explore the long-term behaviour of gut microbiota, we investigated how the mean-squared displacement of log-relative OTU abundance $<\delta^2(\Delta t)>$ changed with time. Specifically, for each time interval, Δt, and each OTU, k, we calculated the displacement of its log-relative abundance, $\delta(\Delta t)_k = \log(X_k(t+\Delta t)/X_k(t))$. Squared displacements were then averaged, for a fixed interval Δt, across all time points and all OTUs, yielding $<\delta^2(\Delta t)>$, the mean-squared displacement of log-relative abundance. Similar to the behaviour of other diverse ecological communities (see Extended Data Fig. 3c), we found that the long-term drift of gut microbiota abundances was well approximated by the equation of anomalous diffusion (Fig. 3 and see Extended Data Fig. 6):

$$\langle \delta^2(\Delta t) \rangle \propto \Delta t^{2H} \quad (2)$$

where H is the Hurst exponent quantifying the collective rate of microbiota abundance drift over time[26]. In comparison with normal diffusion ($H = 0.5$), a Hurst exponent of $H > 0.5$ indicates a tendency for an increase (or decrease) in abundance to be followed by a further increase (or decrease), whereas a value of $H < 0.5$ indicates a higher degree of stability and a tendency for abundance to revert to its mean. In contrast to short-term fluctuations of bacterial abundances, described by the Laplace distribution (equation (1), Fig. 1), the Hurst exponent in equation (2) quantifies the rate at which the average root mean-squared displacement of abundance increases as a function of time. Both in human and mouse gut microbiota, our analysis revealed small Hurst exponents ($H = 0.09 \pm 0.03$, $H = 0.08 \pm 0.02$, mean ± s.d. across humans and mice, respectively). This suggests that, despite overall stability[5,27,28], gut microbiota exhibit a slow but continuous and predictable long-term abundance drift. Furthermore, the temporal dynamics of individual OTU abundances was also well approximated by the equation of anomalous diffusion (see Extended Data Fig. 7a), with the distribution of Hurst exponents across individual OTUs exhibiting substantial variability (see Extended Data Fig. 7b). This demonstrates the heterogeneity of the long-term stability of different gut bacterial taxa within and across hosts.

Both short- and long-term dynamics of gut microbiota contribute to the overall turnover of gut bacterial species. To directly investigate the dynamics of gut microbiota composition, we next calculated the distribution of residence (t_{res}) and return (t_{ret}) times for individual OTUs. Following previous macroecological

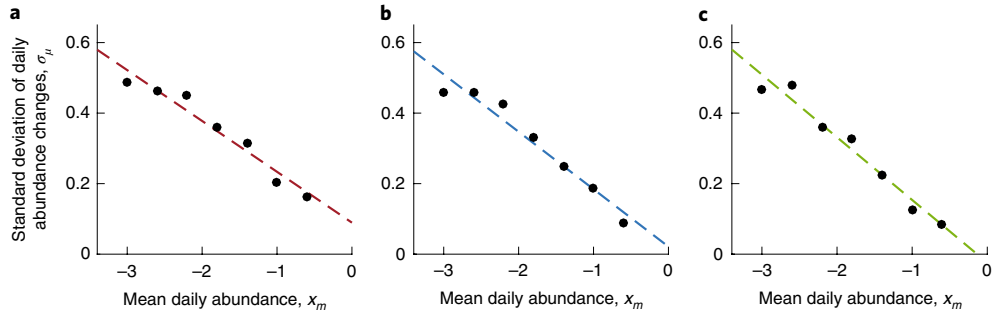


Fig. 2 | Scaling of the variability in daily abundance changes with average bacterial abundance. a–c, Across each OTU k, the s.d. of daily abundance changes (σ_μ) decreases with average daily abundance (x_m), defined as the mean of successive log-abundance values: $x_m = \frac{1}{2}[\log(X_k(t+1)) + \log(X_k(t))]$. The results are shown for two individuals from different human studies (A (**a**) and M3 (**b**)) and mice fed an LFPP diet (**c**). Dashed lines represent least-squares regression fits to the data, with slopes of $r = -0.16 \pm 0.02$, -0.16 ± 0.02 and -0.17 ± 0.03 for A (**a**), M3 (**b**) and LFPP mice (**c**), respectively (mean ± s.d.; $n = 6$ equal subsamples of the data for humans and $n = 3$ animals for mice; see Methods). Abundance changes in **c** were combined across the three mice on the LFPP diet.

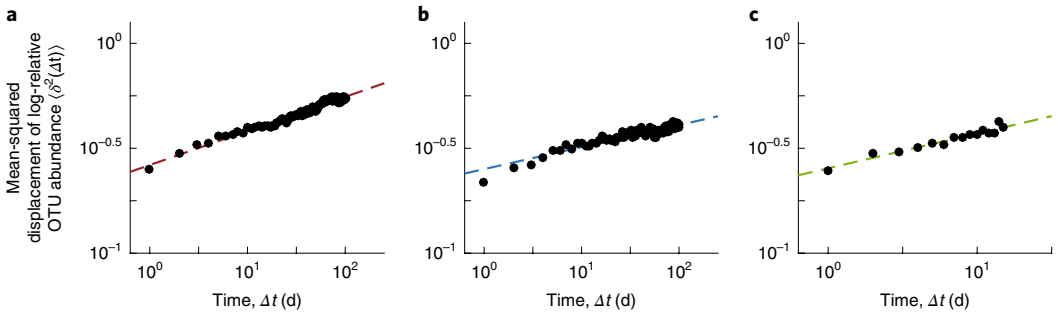


Fig. 3 | Long-term drift of gut microbiota abundances. a–c, In humans and mice, the mean-squared displacement of log-relative OTU abundance ($\langle\delta^2(\Delta t)\rangle$) increases with time as a power law of the form $\langle\delta^2(\Delta t)\rangle \propto \Delta t^{2H}$. Hurst exponents, which quantify the rate of abundance drifts, are $H = 0.07 \pm 0.03$, 0.08 ± 0.02 and 0.08 ± 0.02 for human A (**a**), human M3 (**b**) and LFPP mice (**c**), respectively (mean ± s.d.; $n = 6$ equal subsamples of the data; see Methods). The data in **c** represent an average over the $n = 3$ individual mice on the LFPP diet (see Methods). Dashed lines represent least-squares regression fits to the data.

analyses[16,29,30], we defined residence times as the time intervals between the emergence and subsequent disappearance of corresponding OTUs; analogously, return times were defined as the intervals between disappearance and re-emergence of OTUs. For the considered microbiota datasets, we first calculated the residence and return times for each disappearance/appearance event corresponding to each OTU. We then aggregated the residence and return times across all events and all OTUs to calculate their distributions (Fig. 4). Again, we observed bacteria residence patterns very similar to those previously described in diverse ecological communities[16,29,30] (see Extended Data Fig. 3b). Specifically, the distributions of t_{res} and t_{ret} were described well by power laws (equation (3)), with exponential tails resulting from the finite length of the analysed time series (Fig. 4, and see Extended Data Fig. 8a,b and Supplementary Fig. 2):

$$p(t) \propto t^{-\alpha} e^{-\lambda t} \quad (3)$$

The residence time distributions were similar within and between individual human and mouse gut microbiota ($\alpha_{res} = 2.3 \pm 0.05$, $\alpha_{ret} = 1.2 \pm 0.02$; mean ± s.d. across humans; $\alpha_{res} = 2.2 \pm 0.04$, $\alpha_{ret} = 0.72 \pm 0.03$, across mice on the LFPP diet), suggesting that the processes governing the local emergence and disappearance of gut bacteria are probably independent of the specific host. The power law distribution of residence times may arise, even in an isotropic environment, from the dynamics of births, deaths and species migration patterns defined by the spatial structure of the ecosystem[29]. Notably, the power-law exponents (with values ~2) of the bacteria residence time distributions are similar to those previously observed in macroecological communities[29].

Taylor's law and identification of bacterial OTUs with perturbed dynamics. Having characterized the distributions of daily abundance changes and residence times, we next investigated the temporal variability of individual OTU abundances. One of the most general relationships in ecology, observed across hundreds of different biological communities, is known as Taylor's power law[31–34]. This law connects species' average abundances to their temporal or spatial variances:

$$\sigma_X^2 = CX^\beta \quad (4)$$

where C is a constant, X and σ_X^2 are the mean and variance of species abundance, respectively, and β is a positive scaling exponent. For processes following simple Poisson fluctuations, the parameter $\beta = 1$, whereas for processes with constant per capita growth variability[35], $\beta = 2$. Values of β have been empirically observed to lie between 1 and 2 for the vast majority of investigated plant and animal species[36]. We note that in macroscopic ecological communities, Taylor's law is usually used to describe the behaviour of a single species. In contrast, similar to a previous microbiota analysis[37], we investigated how the temporal abundance variability scales with the average bacterial abundance across all OTUs. Interestingly, our analysis revealed that the temporal variability patterns of gut micro-

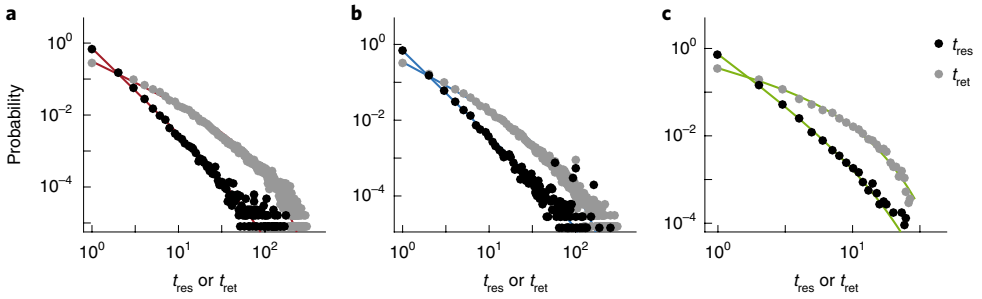


Fig. 4 | Distribution of residence and return times of gut microbiota. a–c, The times t_{res} and t_{ret} were defined as the time intervals during which an OTU was continuously detected at any abundance in the community, or absent from the community, respectively. Probability distributions of t_{res} and t_{ret} follow power laws of the form $p(t) \propto t^{-\alpha} e^{-\lambda t}$, with the exponential tail resulting from the finite length of each time series. Power-law exponents are $\alpha_{res} = 2.3 \pm 0.04$, 2.2 ± 0.07 and 2.2 ± 0.04 for the residence times and $\alpha_{ret} = 1.1 \pm 0.02$, 1.2 ± 0.05 and 0.72 ± 0.03 for the return times (mean ± s.d.; $n = 6$ equal subsamples of the data for humans A (**a**) and M3 (**b**), and $n = 3$ mice on the LFPP diet (**c**); see Methods). Residence and return times were combined across the three individual mice on the LFPP diet. Solid lines represent MLE fits to the data.

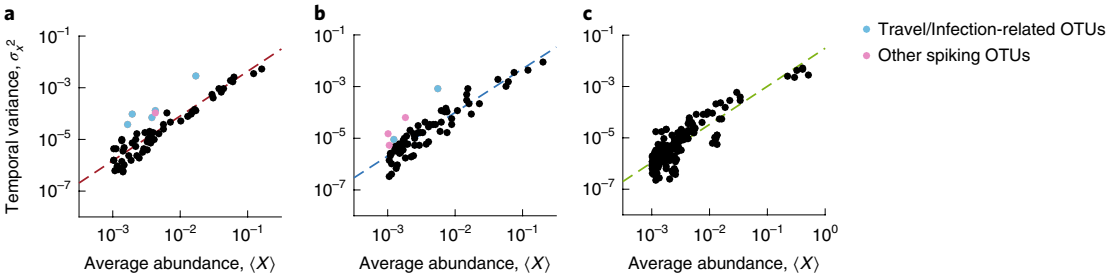


Fig. 5 | Taylor's power law relationships in the gut microbiota. Averages and temporal variances of OTU abundances closely follow Taylor's power law of the form $\sigma_X^2 \propto X^\beta$, with $\beta = 1.66 \pm 0.09$, 1.60 ± 0.08 and 1.49 ± 0.02 for humans A and B, and LFPP mice, respectively (mean ± s.d.; $n = 6$ equal subsamples of the data for humans and $n = 3$ animals for mice; see Methods). Each point corresponds to the average abundance and the corresponding temporal variance of a single bacterial OTU. **a**,**b**, OTUs that exhibited abrupt increases in abundance are indicated as coloured circles (see Methods). Light-blue circles indicate OTUs that exhibited substantial increases in abundance, specifically during periods of travel (human A (**a**)) and enteric infection (human B (**b**)). **c**, Data from each mouse on the LFPP diet are shown together. Dashed lines represent least-squares regression fits to the data.

biota also closely followed Taylor's law (Fig. 5 and see Extended Data Fig. 9a,b), with exponents for human and mouse gut microbiota generally consistent with values observed previously in other ecological communities[36] ($\beta = 1.7 \pm 0.02$ across all four humans, $\beta = 1.49 \pm 0.02$ across LFPP mice). The compositional nature of microbiota datasets did not explain the values of Taylor's law exponents (see Extended Data Fig. 10), and dynamics consistent with Taylor's law have also been observed in a recent short-term analysis of the human vaginal microbiota[37]. It has been previously suggested that competitive interactions between species may result in a Taylor's law exponent in the range between 1 and 2 (ref. [35]). Alternatively, Taylor's law with non-trivial exponents may arise due to stochastic demographics of population growth and decline[36], the presence of species subtypes each with a Gamma-distributed abundance,[38] or a balance between the species' tendency to aggregate and disperse[34,39]. Taylor's law may also arise across a broad range of mathematical models describing population dynamics, without the need for a specific biological mechanism[40,41]. In the future, it will be interesting to investigate and compare the aforementioned theoretical models specifically in the context of microbiota dynamics.

Although Taylor's law described the overall dynamics of gut microbiota well, some specific OTUs clearly deviated from the general trend (Fig. 5). To determine whether their behaviour reflected specific ecological perturbations, we identified all OTUs that exhibited notable and abrupt increases in abundance during previously documented periods of travel in human A and enteric infection in human B[4] (see Methods). Interestingly, these travel- and infection-related OTUs closely matched the outliers from Taylor's law (Fig. 5a,b, blue circles), displaying on average approximately tenfold greater variance than expected based on the overall Taylor's law trend (see Extended Data Fig. 9a,c and Supplementary Table 1). Many of these OTUs were members of the Proteobacteria (in human A: OTU 13, family Enterobacteriaceae; OTU 29, family Pasteurellaceae; OTU 5771, family Enterobacteriaceae; in human B: OTU 13, family Enterobacteriaceae), which were associated with the microbiota perturbations[4] (see Supplementary Table 1). Moreover, other OTUs, primarily belonging to the Firmicutes, that exhibited abrupt changes in abundance (in human A: OTU 25, family Peptostreptococcaceae; in human B: OTU 95, family Ruminococcaceae; OTU 110, family Ruminococcaceae), also displayed higher temporal variability than expected based on Taylor's law (Fig. 5a,b, purple circles; see also Extended Data Fig. 9c and Supplementary Table 1). These results suggest that macroecological relationships can be used to identify and characterize specific microbial taxa that are likely involved in periods of dysbiosis and context-specific environmental perturbations.

Effects of different diets on macroecological dynamics of gut microbiota. It is well established that the dynamics of diverse ecosystems are strongly affected by their environment[42]. Host dietary intake is a major environmental factor influencing gut bacterial abundances[7,8,43] and disease phenotypes[44,45]. Therefore, we next

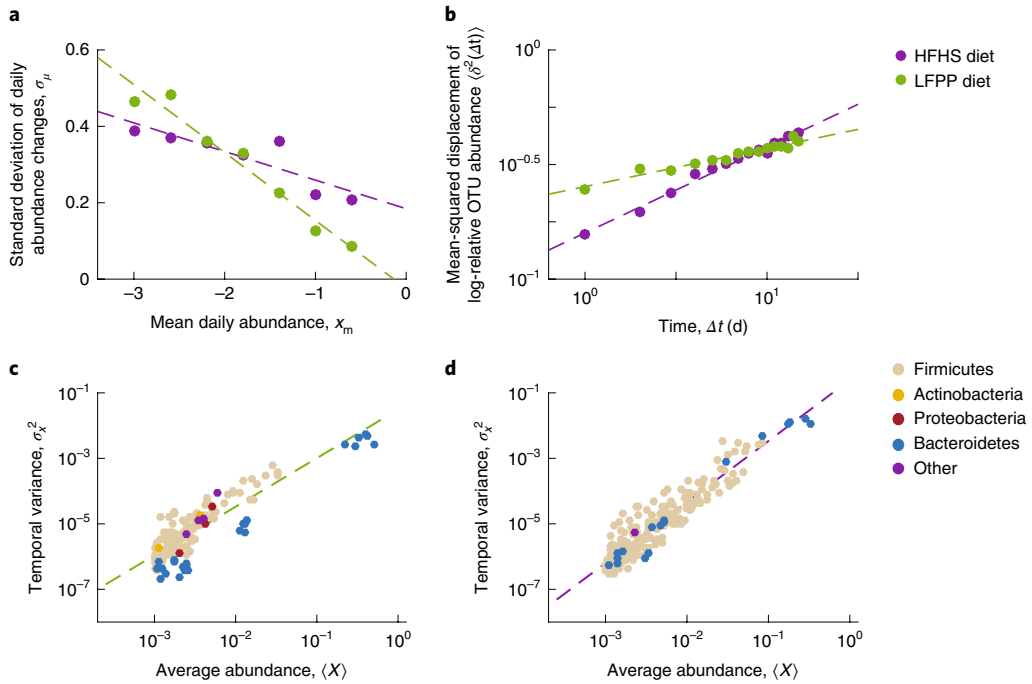


Fig. 6 | Dynamics of gut microbiota in mice fed different diets. a, OTUs in mice fed a low-fat LFPP diet show a stronger dependence of the variability in daily abundance changes (σ_μ) on the average daily abundance (x_m) compared with those fed a high-fat HFHS diet (regression slopes: $r=-0.17\pm0.03$, $r=-0.08\pm0.02$ (mean ± s.d.) for LFPP (green) and HFHS (purple) mice, respectively). Data were combined across the three mice on each diet, and dashed lines represent least-squares regression fits to the data. **b**, OTU abundances in the LFPP mice exhibit a slower long-term abundance drift compared with drift in the HFHS mice ($H=0.08\pm0.02$ and $H=0.19\pm0.02$ for LFPP and HFHS mice, respectively). **c**,**d**, Taylor's law relationships show differences in the overall scaling of the average OTU abundance and the s.d. of the temporal abundance on each diet (LFPP diet (**c**) and HFHS diet (**d**); $\beta=1.49\pm0.02$, $\beta=1.86\pm0.07$, respectively), driven by the temporal behaviour of the Bacteroidetes in the LFPP mice (blue circles). Data were combined from the $n=3$ mice on each diet. Dashed lines represent least-squares regression fits performed using the combined data.

explored how different diets affect the observed macroecological relationships describing gut microbiota dynamics. To this end, we used data from the study of Carmody et al.[8], who investigated faecal bacterial abundances in individually housed mice fed either a low-fat (LFPP) diet or a high-fat (HFHS) diet. Our analysis revealed that the short-term dynamics of gut microbiota were notably affected by the diets. While the s.d. of daily abundance changes declined rapidly with increasing OTU abundance in the LFPP mice (Fig. 6a, green), it remained more homogeneous across OTU abundances in the HFHS mice (Fig. 6a, purple; regression slopes $r=-0.17\pm0.03$ for the LFPP diet, and -0.08 ± 0.02 for the HFHS diet; Z-test of regression coefficients, $P=2.0\times10^{-5}$). The relatively smaller short-term variability of highly abundant species on the LFPP diet probably reflects more stable niches for some bacteria (such as Bacteroidetes) that may primarily catabolize dietary fibres. On the other hand, relatively higher fluctuations of lowly abundant bacteria on this diet may be induced by cross-feeding on catabolic byproducts of highly abundant species. The observed dependence of short-term abundance variability on species abundance is much weaker on the HFHS diet, which may result from a general loss of niche diversity due to the substantially reduced nutrient complexity of that diet. Alternatively, these trends may reflect diet-dependent influence and interactions between the gut microbiota and the host, for example, diet effects on the host immune system[46] or the gastrointestinal tract physiology[47].

In other ecological communities, smaller short-term fluctuations of species abundance do not necessarily lead to increased long-term ecological stability[48,49]. Thus, in addition to short-term fluctuations, we also investigated how different diets affected the long-term abundance drift of gut microbiota. Hurst exponents were notably larger in the HFHS mice, indicating substantially faster long-term drift of bacterial abundance on this diet (Fig. 6b and see Extended Data Fig. 6b; $H=0.19\pm0.02$ for the HFHS diet, and 0.08 ± 0.02 for the LFPP diet; Z-test $P<1\times10^{-10}$). We note that short-term fluctuations in abundance were somewhat higher on the LFPP diet compared with the HFHS diet, which is reflected in a higher y-axis intercept for the diffusion on the LFPP diet (Fig. 6b, green); the higher intercept is due to larger short-term fluctuations of numerous lowly abundant bacteria on the LFPP diet compared with the HFHS diet (Fig. 6a). Despite the intercept differences, the observed diffusion trend continued over long time scales (>100 days, based on the data in Fig. 3). Therefore, the long-term increase of the mean-squared displacement of microbiota abundances is likely to depend primarily on the difference in the respective Hurst exponents.

Previous studies have demonstrated diet-induced compositional shifts of gut microbiota[7,8,43] and a reduced gut bacterial diversity in western populations, attributed in part to altered dietary habits[50–52]. Our analysis shows that different diets not only affect the composition of gut microbiota, but also substantially change their long-term dynamics. The reduced long-term stability on the HFHS diet may result from a higher degree of neutral drift and increased interspecies competition associated with a more homogeneous nutrient environment[53]. Our analysis demonstrated that while the abundance drifts of Bacteroidetes and Firmicutes, two major phyla in the mouse gut, were relatively similar on the HFHS diet (Hurst exponent, $H=0.18\pm0.1$ for Bacteroidetes, and $H=0.18\pm0.03$ for Firmicutes), the Bacteroidetes exhibited a significantly reduced drift on the LFPP diet compared with the Firmicutes ($H=0.03\pm0.06$ and $H=0.09\pm0.02$; Z-test $P=3\times10^{-8}$). This suggests that, although the LFPP diet decreased the long-term abundance drift of all taxa, the stability of the Bacteroidetes was particularly affected by this diet.

Different diets may not only change overall gut microbiota dynamics, but also alter the temporal variability of individual taxa relative to the rest of the community. To understand taxa-specific changes, we next examined Taylor's law in mice on the LFPP diet (Fig. 6c) and the HFHS diets (Fig. 6d). We calculated, for each diet, the temporal mean and variance of each OTU, and then aggregated the data points across multiple mice on the same diet. Interestingly, it was previously demonstrated that the Taylor's law exponent may depend on the environment, at least for some species[54–56]. Our analysis of the gut microbiota dynamics on the different diets is consistent with these observations. Specifically, we found that the power-law exponents were significantly different between the two diets ($\beta = 1.49 \pm 0.02$ for the LFPP diet, $\beta = 1.86 \pm 0.07$ for the HFHS diet; Z-test $P = 1.5 \times 10^{-6}$). The temporal fluctuations of the Bacteroidetes (Fig. 6c,d, blue circles) exhibited significantly lower variability given their abundance on the LFPP diet, but not on the HFHS diet (hypergeometric test, $P = 2.4 \times 10^{-4}$; see Supplementary Table 2 and Methods). Moreover, we observed significantly lower Taylor's law exponents on the LFPP diet, specifically for the Bacteroidetes ($\beta = 1.66 \pm 0.06$ on the LFPP diet, $\beta = 1.95 \pm 0.03$ on the HFHS diet; Z-test $P = 0.0023$), but not for all other bacterial taxa ($\beta = 1.84 \pm 0.2$ for the LFPP diet, $\beta = 1.86 \pm 0.07$ for the HFHS diet; Z-test $P = 0.39$). Notably, the highly abundant Bacteroidetes and their lower temporal variability on the LFPP diet were primarily responsible for the relatively smaller short-term fluctuations of highly abundant bacteria on this diet (Fig. 6a, green). Bacteroidetes are known to metabolize a wide range of dietary fibres present in the LFPP diet[57–59], and are predominantly lost during multigenerational propagation of mice on a low-fibre diet[51]. This suggests that specific members of the Bacteroidetes (in our case, OTU 118, OTU 237, OTU 364, family Porphyromonadaceae; see Supplementary Table 2) may exhibit both smaller temporal variability and slower abundance drift by directly exploiting stable niches that are present on the LFPP diet and probably lost on the HFHS diet. These results demonstrate that macroecological analyses can be used to identify specific taxa with temporal dynamics affected by different diets.

Discussion

The present study demonstrates that, despite an amazing diversity of interactions and organizational complexity, the dynamics of gut microbiota can be described by multiple robust quantitative relationships. The scaling laws revealed by our analysis characterize both short- and long-term microbiota dynamics, and are usually observed across many orders of magnitude in time and bacterial abundances. Furthermore, we show that these relationships are unlikely to arise due to technical noise and the compositional nature of microbiota datasets. Despite the difference of more than six orders of magnitude in the relevant spatial and interaction scales, the statistical relationships described in the present study are strikingly similar to those observed previously in many diverse macroecological systems. This similarity suggests that the temporal processes in both macroscopic and microbial communities may be governed by a universal set of underlying mechanisms and principles.

We anticipate that the quantitative statistical framework developed in macroecology[60–63] will be important for analysing microbiota dynamics. Because the observed statistical relationships describe different aspects of community dynamics, an important goal for future studies will be to unify these observations into an integrated view of microbial ecology, which also takes into account the spatial and environmental dimensions. Moreover, the ability to easily perturb microbiota composition and environment, add and remove particular species, as well as monitor species abundance at high temporal and spatial resolution, suggests an exciting opportunity to use microbiota as a convenient model system to explore general laws of ecology.

We also envision a quantitative ecological framework being important for understanding how host-specific and environmental factors influence the dynamics of human gut and other health-related microbiota. The presented results suggest that the observed macroecological relationships can be used to identify both global changes of microbiota dynamics and specific taxa with abnormal temporal behaviour. Such taxa may serve as biomarkers of disease and clinically relevant perturbations[64,65]. Therefore, it will be important to investigate how the quantitative macroecological relationships revealed in our study vary across large and densely sampled human cohorts[66–68].

Methods

Analysis of 16S rRNA sequences. Raw 16S rRNA sequencing data for humans A and B were obtained from the European Nucleotide Archive (accession no.: PRJEB6518 (ref. [4])). Raw sequencing data from humans M3 and F4 and mice were obtained from the MG-RAST database[69] (4457768.3-4459735.3 for humans; 4597621.3-4599933.3 for mice). Sequences were analysed using USEARCH 8.1 (ref. [70]) with an open clustering approach. For studies including unfiltered sequencing reads, filtering was performed using the –fastq_filter command with the expected maximum number of 2 wrong bases per read. All reads were then truncated to 100 base pairs, with shorter reads discarded. Following a conventional approach, reads were de-replicated and clustered at 97% sequence similarity, using the –cluster_otus command to generate OTUs with a minimum of two sequences. Sequences were then assigned to OTUs using the –usearch_global command, resulting in an OTU table for each dataset. OTU taxonomic assignments were made using the RDP classifier[71]. Sequencing reads from each sample were then rarefied to a depth of 25,000 and 17,000 for the two human studies (A/B, M3/F4), respectively, and 25,000 for the mouse study, using Qiime 1.8 (ref. [72]).

OTU inclusion criteria. To control for known technical factors such as sample preparation and sequencing noise, analysis was restricted to OTUs passing two sets of criteria. First, OTUs were required to be present in over half of the samples within respective subjects. Second, they were required to have a mean relative abundance $>1 \times 10^{-3}$ over the time series. This abundance cutoff corresponded to a mean of 25 reads (humans A and B, and LFPP/HFHS mice) and 17 reads (humans M3 and F4) over respective sampling periods. The final analysis of humans included ~75 OTUs comprising ~90% of the reads with OTU assignments in any given sample. For mice, these criteria resulted in the inclusion of ~70 OTUs in the HFHS diet and ~55 OTUs in the LFPP diet, comprising ~90% of reads with OTU assignments in a given sample. As the HFHS mice initially received an LFPP diet, the analysis of these mice began 5 days after the diet shift. For the calculation of residence and return times, different inclusion criteria were imposed, because these analyses would be biased by a specific prevalence cutoff and were more robust to noise in OTU abundance levels.

Daily abundance changes. Daily abundance changes were defined as $\mu_k(t) = \log(X_k(t+1)/X_k(t))$, where $X_k(t)$ is the relative abundance of a given OTU k on day t. The presented distributions represented the combined data across all OTUs. To obtain the average distributions, we first calculated distributions of daily abundance changes for each OTU across all time points. We then combined the distributions from all considered OTUs. In addition, for human datasets, we investigated whether the resulting Laplace distributions of daily abundance changes depend on aggregating OTUs with different abundance fluctuations. To this end, we first normalized the daily abundance changes for each individual OTU by its s.d. calculated across each temporal dataset. The resulting normalized distributions were then combined across all OTUs and across all human datasets (see Extended Data Fig. 1d). To estimate the variability of the distribution of daily abundance changes within human subjects, each time series was divided into six consecutive time periods of equal length (estimates were generally insensitive to the number of time periods). Within each time frame, daily abundance changes were calculated and maximum likelihood estimation (MLE) was used to fit the Laplace distribution exponent, with the mean and s.d. of these values reported in the main text. For the mouse study, s.d.s reflected variability across the three individual mice on each diet. Mean daily abundance, x_m, was defined as the mean of consecutive log-OTU abundances: $x_m = \frac{1}{2}[\log(X_k(t+1)) + \log(X_k(t))]$. To estimate the variability in daily abundance changes as a function of abundance, abundance changes were binned by values of x_m using a bin size of 0.4, and s.d. (σ_μ) values were then calculated on the binned abundance changes. For comparisons of diets, daily abundance changes were combined across the three mice on each diet. Abundance changes and mean daily abundances were calculated using the base ten logarithm in all figures, while the natural log was used for parameter estimation.

Hurst exponents. The mean-squared displacement of log-OTU abundances was estimated as:

$$\langle \delta^2(\Delta t) \rangle = \frac{1}{N(T-\Delta t)} \sum_k \sum_i [x_k(t_i + \Delta t) - x_k(t_i)]^2$$

where the angle brackets denote a community average (over time points and OTUs). Here, $x_k(t_i)$ is the log-relative abundance of OTU k at time t_i, N is the total number of OTUs, T is the number of days in a time series, and $(T-\Delta t)$ is the number of time windows over which the averaging is performed. A maximum time lag of 100 days and 15 days was chosen for humans and mice, respectively, due to the finite length of each time series. Hurst exponents were then calculated by regressing $<\delta^2(\Delta t)>$ against Δt in log-transformed axes. To estimate the variability of Hurst exponents within human subjects, time series were divided into six equal-length time frames, as was done for daily abundance change calculations. Hurst exponents for individual OTUs were estimated in a similar fashion, but with displacements restricted to time averages. For comparisons of diets, Hurst exponents were additionally averaged over mice within each diet:

$$\langle \delta^2(\Delta t)\rangle_{\text{diet}} = \frac{1}{L}\sum_l \frac{1}{N_l(T_l-\Delta t)}\sum_k\sum_i [x_{l,k}(t_i+\Delta t)-x_{l,k}(t_i)]^2$$

where the outermost summation is over individual mice l ($L=3$) on each diet.

Residence and return times. Residence times ($t_{\text{res},k}$) of an OTU k corresponded to the number of consecutive time points between its appearance ($T_{\text{a},k}$) and disappearance ($T_{\text{d},k}$) in the community: $t_{\text{res},k}=T_{\text{d},k}-T_{\text{a},k}$. Here, $T_{\text{a},k}$ is a time point at which the OTU appeared in the community, with no reads detected on the previous collection date, and $T_{\text{d},k}$ is the closest time point at which the OTU reads were no longer detected. Return times (t_{ret}) were similarly defined as the number of consecutive time points between an OTU disappearance ($T_{\text{d},k}$) and reappearance ($T_{\text{a},k}$) in the community: $t_{\text{ret},k}=T_{\text{a},k}-T_{\text{d},k}$. Only intervals that fell entirely within the time period of the study were included. A series of alternative criteria was also considered to ensure robustness of the residence/return time distributions: (1) to ensure that the results were not biased by sequencing detection sensitivity, the distributions were recalculated using data subsampled to various sequencing depths (down to 1,000 reads per sample); (2) to account for false negatives in read detection, single read counts of zero, interrupting a run of consecutive non-zero abundances, were neglected, that is, an OTU with zero reads at time t was considered to be present in the community if that OTU was also present at times $t-1$ and $t+1$; (3) to control for false positives in read detection, OTU abundances with a single read were neglected and treated as zero counts. Results were qualitatively insensitive to both the sampling depth and the aforementioned alternative read detection criteria. To estimate the variability of distribution parameters within humans, OTUs were randomly assigned into six equal-sized groups. Residence and return times were then calculated within each assigned group, and exponents fitted using MLE, with means and s.d.s calculated across the groups. Within diets, means and s.d.s were calculated across individual mice.

Taylor's power law. The mean abundance X_k and its variance $\sigma^2_{X_k}$ for each OTU k were calculated across the corresponding time series. Taylor's exponents were obtained by performing linear regression of the log-transformed means and variances across OTUs in each subject. To estimate the variability of exponents within subjects, time series were divided into six consecutive time periods of equal length as described previously. Spiking OTUs were defined as those with a single-day abundance greater by more than 25-fold than the average abundance calculated across all other days. Travel-related and infection-related OTUs in humans A and B were identified as those with abundances that spiked by more than 25-fold during the documented periods of travel and infection[7]. For mice, Taylor's law outliers were identified using a likelihood-based approach. Briefly, linear regression on the log-transformed means and variances was performed excluding a single OTU k. The probability of observing the left-out OTU k was then calculated using a Gaussian likelihood function based on estimated residuals. All OTUs with probability less than $\alpha=0.025$ were considered to be outliers. For diet comparisons, means and variances were combined across individual mice within diets groups.

Simulations of sampling errors associated with finite sequencing depth. Read counts were simulated using a multinomial distribution with parameters $\mathbf{p}$ and N. The average abundance vector $\mathbf{p}$ across all OTUs was estimated separately for each human A, B, M3 and F4. The sampling depth N was equal to the total sequencing depth used in each study (25,000 reads per sample for humans A and B; 17,000 reads per sample for humans M3 and F4). Simulations were performed multiple times to generate sample OTU trajectories for each human that solely reflected sampling errors. To account for sporadic sequencing read dropouts, zero counts were introduced into simulations of each OTU to match the empirical frequency of zero counts observed in the real data.

Simulations of macroecological dynamics across a range of community diversities. Initial ($t=t_0$) abundances of $N=65$ species were generated using power-law distributions. The power-law exponents were selected to generate a range of community diversities, quantified by the effective number of species (ENS $=e^H$, where H is the Shannon diversity). To simulate Gaussian daily abundance changes (see Extended Data Fig. 2), the Ornstein–Uhlenbeck process was used to generate 1,000 consecutive time points. For the Taylor's law analyses (see Extended Data Fig. 10), we performed simulations analogous to those of Kilpatrick and Ives[35]. Specifically, abundances of the 65 non-interacting species were simulated for 300 time points, matching the number of OTUs and samples in human A. The simulations were performed using absolute abundances, and the scaling relationships were then calculated using either absolute or relative abundances.

Statistics. All statistical analysis was performed using custom scripts written in MATLAB (https://www.mathworks.com). Comparisons of estimated exponents between mouse diet groups were performed by first calculating the relevant coefficient and associated s.e.m. of the data combined across the three mice in each diet group. Z-tests were then performed comparing the two coefficients associated with each diet group, assuming normality of the s.e.m. Reported P values correspond to one-sided tests.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequencing data used in the present study can be downloaded from the European Nucleotide Archive (accession no. PRJEB6518 for humans A and B; see ref. [4] for metadata) and MG-RAST databases (4457768.3–4459735.3 (https://www.mg-rast.org/linkin.cgi?project=mgp93) for humans M3 and F4 (ref. [3]); 4597621.3–4599933.3 (https://www.mg-rast.org/linkin.cgi?project=mgp11172) for mice[8]). These data were used to generate all figures in the main text, Supplementary information and Extended Data, with the exception of Extended Data Fig. 3, for which the figures were adapted from the original references[16,17,22–24,63]; figures from the original references were digitized and the resulting data points re-plotted.

Code availability

All MATLAB scripts used to perform data analysis and generate figures are available on GitHub (https://github.com/brianwji/Macroecological-Relationships).

Received: 21 June 2018; Accepted: 7 February 2020;
Published online: 13 April 2020

References

1. Faust, K., Lahti, L., Gonze, D., de Vos, W. M. & Raes, J. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **25**, 56–66 (2015).
2. Gohl, D. M. et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* **34**, 942–949 (2016).
3. Caporaso, J. G. et al. Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
4. David, L. A. et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15**, R89 (2014).
5. Faith, J. J. et al. The long-term stability of the human gut microbiota. *Science* **341**, 1237439 (2013).
6. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108**(Suppl 1), 4554–4561 (2011).
7. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
8. Carmody, R. N. et al. Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17**, 72–84 (2015).
9. Smits, S. A. et al. Seasonal cycling in the gut microbiome of the Hadza hunter–gatherers of Tanzania. *Science* **357**, 802–806 (2017).
10. Brown, J. H. & Maurer, B. A. Macroecology: the division of food and space among species on continents. *Science* **243**, 1145–1150 (1989).
11. von Humboldt, A., Bopland, A. & Ross, T. *Personal Narrative of Travels to the Equinoctial Regions of America: During the Years 1799–1804* (Benediction Classics, 2012).
12. Mcgill, B. The what, how and why of doing macroecology. *Global Ecol. Biogeogr.* **28**, 6–17 (2019).
13. Li, L. & Ma, Z. S. Testing the neutral theory of biodiversity with human microbiome datasets. *Sci. Rep.* **6**, 31448 (2016).
14. Shoemaker, W. R., Locey, K. J. & Lennon, J. T. A macroecological theory of microbial biodiversity. *Nat. Ecol. Evol.* **1**, 107 (2017).
15. Keitt, T. H., Amaral, L. A. N., Buldyrev, S. V. & Stanley, H. E. Scaling in the growth of geographically subdivided populations: invariant patterns from a continent-wide biological survey. *Phil. Trans. R. Soc. B* **357**, 627–633 (2002).
16. Keitt, T. H. & Stanley, H. E. Dynamics of North American breeding bird populations. *Nature* **393**, 257–260 (1998).
17. Niwa, H. S. Random-walk dynamics of exploited fish populations. *ICES J. Mar. Sci.* **64**, 496–502 (2007).
18. Allen, A. P., Li, B. L. & Charnov, E. L. Population fluctuations, power laws and mixtures of lognormal distributions. *Ecol. Lett.* **4**, 1–3 (2001).

19. Mitzenmacher, M. A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1**, 226–251 (2003).
20. Azaele, S., Pigolotti, S., Banavar, J. R. & Maritan, A. Dynamical evolution of ecosystems. *Nature* **444**, 926–928 (2006).
21. Amaral, L. A. N., Sergey, V. B., Havlin, S., Salinger, M. A. & Stanley, H. E. Power law scaling for a system of interacting units with complex internal structure. *Phys. Rev. Lett.* **80**, 1385–1388 (1998).
22. Sun, J., Cornelius, S. P., Janssen, J., Gray, K. A. & Motter, A. E. Regularity underlies erratic population abundances in marine ecosystems. *J. R. Soc. Interface* https://doi.org/10.1098/rsif.2015.0235 (2015).
23. Stanley, M. H. R. et al. Scaling behaviour in the growth of companies. *Nature* **379**, 804–806 (1996).
24. Plerou, V., Amaral, L. A. N., Gopikrishnan, P., Meyer, M. & Stanley, H. E. Similarities between the growth dynamics of university research and of competitive economic activities. *Nature* **400**, 433–437 (1999).
25. Hekstra, D. R. & Leibler, S. Contingency and statistical laws in replicate microbial closed ecosystems. *Cell* **149**, 1164–1173 (2012).
26. Metzler, R., Jeon, J. H., Cherstvy, A. G. & Barkai, E. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* **16**, 24128–24164 (2014).
27. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350**, 663–666 (2015).
28. Gibbons, S. M., Kearney, S. M., Smillie, C. S. & Alm, E. J. Two dynamic regimes in the human gut microbiome. *PLoS Comput. Biol.* **13**, e1005364 (2017).
29. Bertuzzo, E. et al. Spatial effects on species persistence and implications for biodiversity. *Proc. Natl Acad. Sci. USA* **108**, 4346–4351 (2011).
30. Suweis, S. et al. On species persistence-time distributions. *J. Theor. Biol.* **303**, 15–24 (2012).
31. Taylor, L. R. Aggregation, variance and mean. *Nature* **189**, 732–735 (1961).
32. Taylor, L. R. & Woiwod, I. P. Temporal stability as a density-dependent species characteristic. *J. Anim. Ecol.* **49**, 209–224 (1980).
33. Taylor, L. R., Woiwod, I. P. & Perry, J. N. Density-dependence of spatial behavior and rarity of randomness. *J. Anim. Ecol.* **47**, 383–406 (1978).
34. Eisler, Z., Bartos, I. & Kertesz, J. Fluctuation scaling in complex systems: Taylor's law and beyond. *Adv. Phys.* **57**, 89–142 (2008).
35. Kilpatrick, A. M. & Ives, A. R. Species interactions can explain Taylor's power law for ecological time series. *Nature* **422**, 65–68 (2003).
36. Anderson, R. M., Gordon, D. M., Crawley, M. J. & Hassell, M. P. Variability in the abundance of animal and plant-species. *Nature* **296**, 245–248 (1982).
37. Ma, Z. S. Power law analysis of the human microbiome. *Mol. Ecol.* **24**, 5428–5445 (2015).
38. Kendal, W. Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. *Ecol. Complex.* **1**, 193–209 (2004).
39. Taylor, L. R. & Taylor, R. A. Aggregation, migration and population mechanics. *Nature* **265**, 415–421 (1977).
40. Giometto, A., Formentin, M., Rinaldo, A., Cohen, J. E. & Maritan, A. Sample and population exponents of generalized Taylor's law. *Proc. Natl Acad. Sci. USA* **112**, 7755–7760 (2015).
41. Zhao, L., Sheppard, L. W., Reid, P. C., Walter, J. A. & Reuman, D. C. Proximate determinants of Taylor's law slopes. *J. Anim. Ecol.* **88**, 484–494 (2019).
42. Brose, U. & Hillebrand, H. Biodiversity and ecosystem functioning in dynamic landscapes. *Phil. Trans. R. Soc. Lond. B* **277**, 2339–2345 (2016).
43. Wu, G. D. et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
44. Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
45. Devkota, S. et al. Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10$^{-/-}$ mice. *Nature* **487**, 104–108 (2012).
46. Kim, K. S. et al. Dietary antigens limit mucosal immunity by inducing regulatory T cells in the small intestine. *Science* **351**, 858–863 (2016).
47. Dey, N. et al. Regulators of gut motility revealed by a gnotobiotic model of diet–microbiome interactions related to travel. *Cell* **163**, 95–107 (2015).
48. Huisman, J. & Weissing, F. J. Biodiversity of plankton by species oscillations and chaos. *Nature* **402**, 407–410 (1999).
49. McCann, K. S. The diversity–stability debate. *Nature* **405**, 228–233 (2000).
50. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
51. Sonnenburg, E. D. et al. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016).
52. Sonnenburg, E. D. & Sonnenburg, J. L. Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* **20**, 779–786 (2014).
53. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8**, 15–25 (2010).
54. Banerjee, B. Variance to mean ratio and the spatial distribution of animals. *Experientia* **32**, 993–994 (1976).
55. Davis, P. M. & Pedigo, L. P. Analysis of spatial patterns and sequential count plans for stalk borer (Lepidoptera: Noctuidae). *Environ. Entomol.* **18**, 504–509 (1989).
56. Downing, J. A. Spatial heterogeneity: evolved behaviour or mathematical artefact? *Nature* **323**, 255–257 (1986).
57. Sonnenburg, E. D. et al. Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. *Cell* **141**, 1241–1252 (2010).
58. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
59. Gurry, T. et al. Predictability and persistence of prebiotic dietary supplementation in a healthy human cohort. *Sci. Rep.* **8**, 12699 (2018).
60. Marquet, P. A. et al. Scaling and power-laws in ecological systems. *J. Exp. Biol.* **208**, 1749–1769 (2005).
61. Real, L.A. & Brown, J. H. (eds.) *Foundations of Ecology* (Univ. Chicago Press, 1991).
62. Hubbell, S. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
63. Azaele, S. et al. Statistical mechanics of ecological systems: neutral theory and beyond. *Rev. Mod. Phys.* **88**, 035003 (2016).
64. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784 (2017).
65. Yassour, M. et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci. Transl. Med.* **8**, 343ra381 (2016).
66. Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017).
67. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).
68. Zeevi, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell* **163**, 1079–1094 (2015).
69. Meyer, F. et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.* **9**, 386 (2008).
70. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
71. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
72. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

Acknowledgements

We thank G. Plata, H. Wang, E. Alm, J. Sonnenburg and E. Sonnenburg for very helpful scientific discussions. D.V. acknowledges funding from the National Institutes of Health (NIH; grant nos. R01GM079759, R35GM131884 and R01DK118044). B.W.J. was supported in part under a Ruth L. Kirschstein National Research Service Award Institutional Research Training grant (no. T32GM007367), and by the MD–PhD programme at Columbia University. R.U.S. was supported by a Fannie and John Hertz Foundation Fellowship and a National Science Foundation Graduate Research Fellowship (no. DGE-1644869).

Author contributions

D.V. conceived the study, and oversaw the project, data analysis and interpretation. B.W.J., R.U.S. and K.T. performed the data analysis, interpretation and modelling. P.D.D. contributed to the data interpretation and analysis. All the authors wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

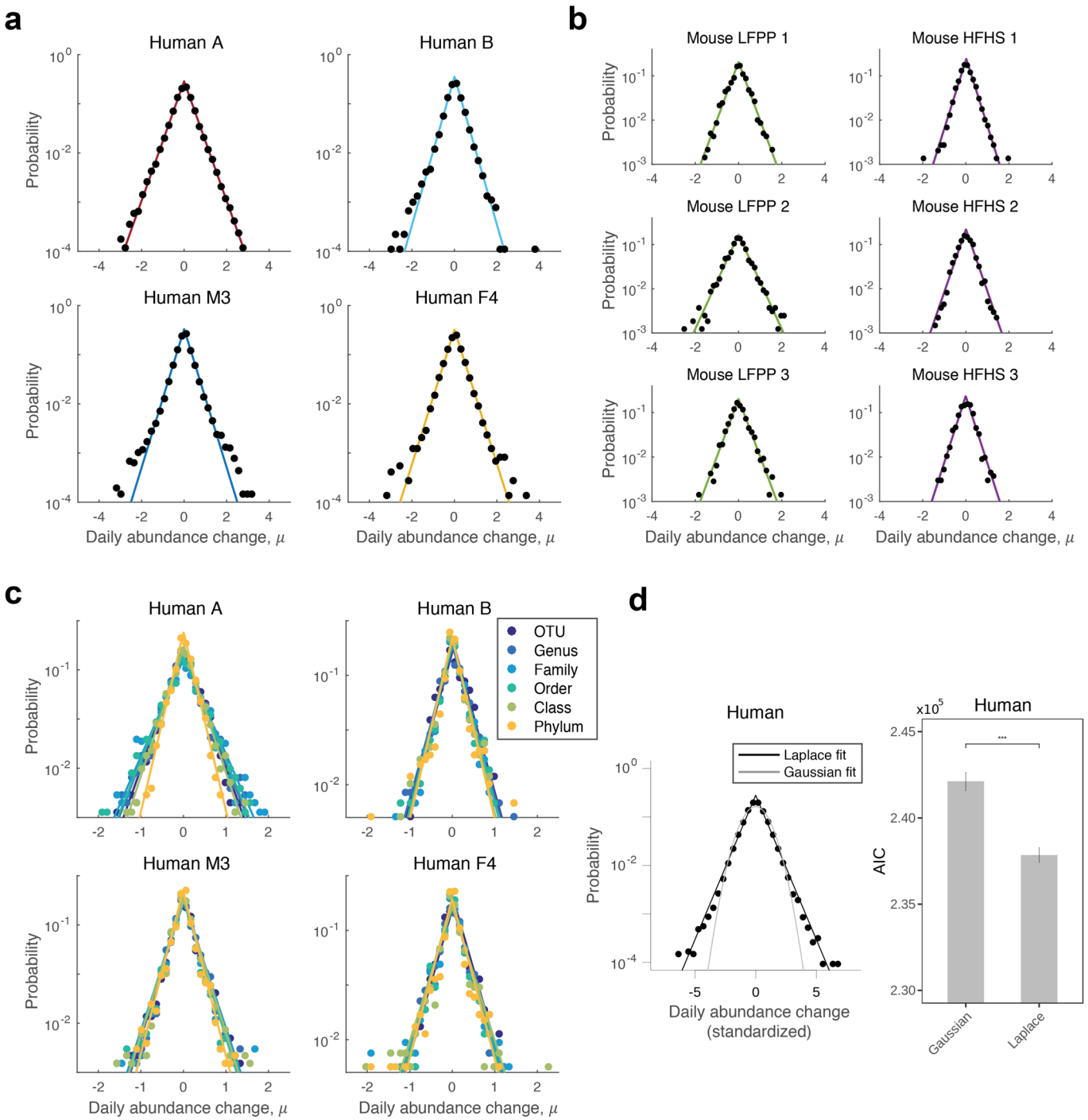
Extended data is available for this paper at https://doi.org/10.1038/s41564-020-0685-1.

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-0685-1.

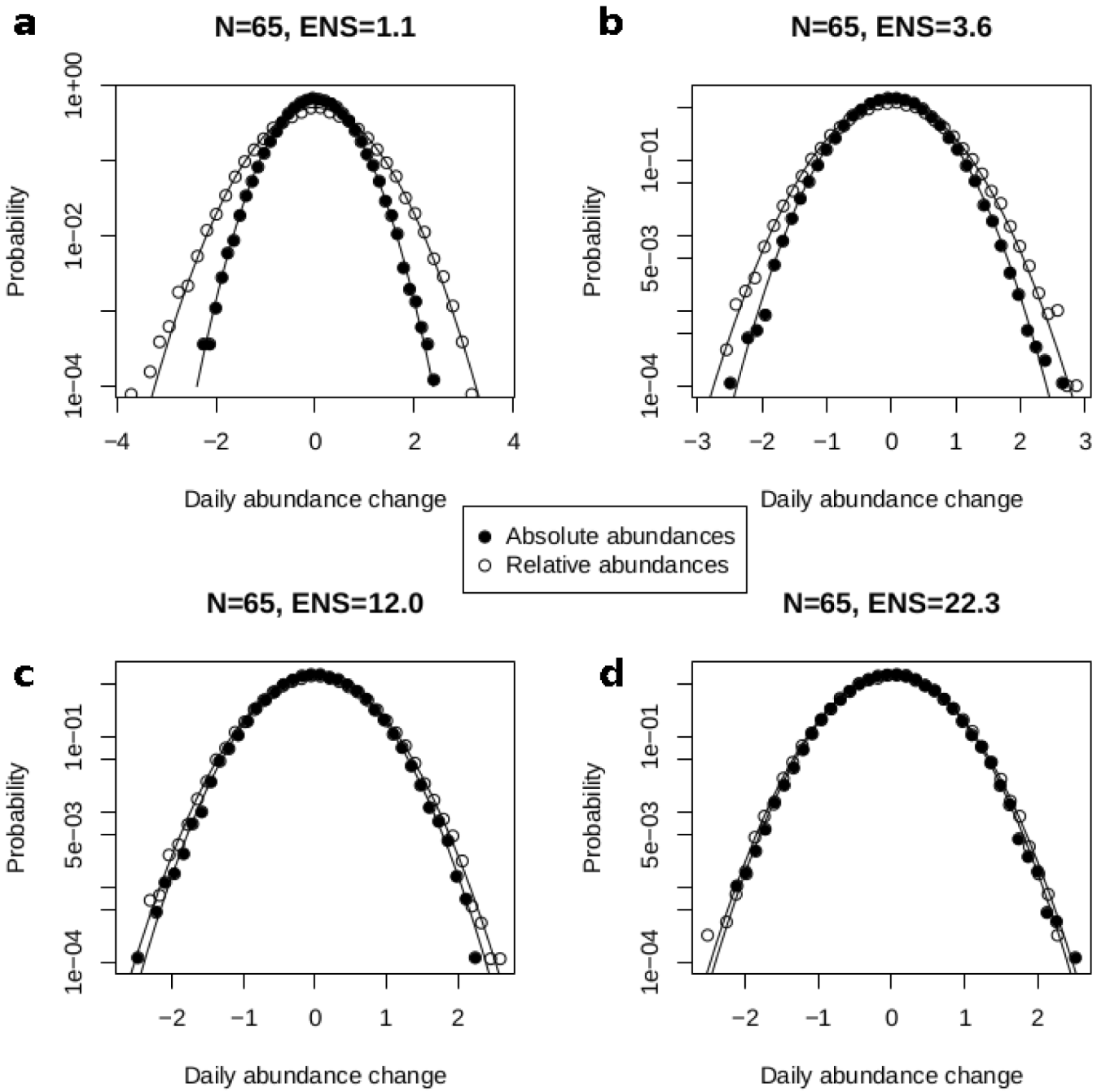
Correspondence and requests for materials should be addressed to D.V.

Reprints and permissions information is available at www.nature.com/reprints.

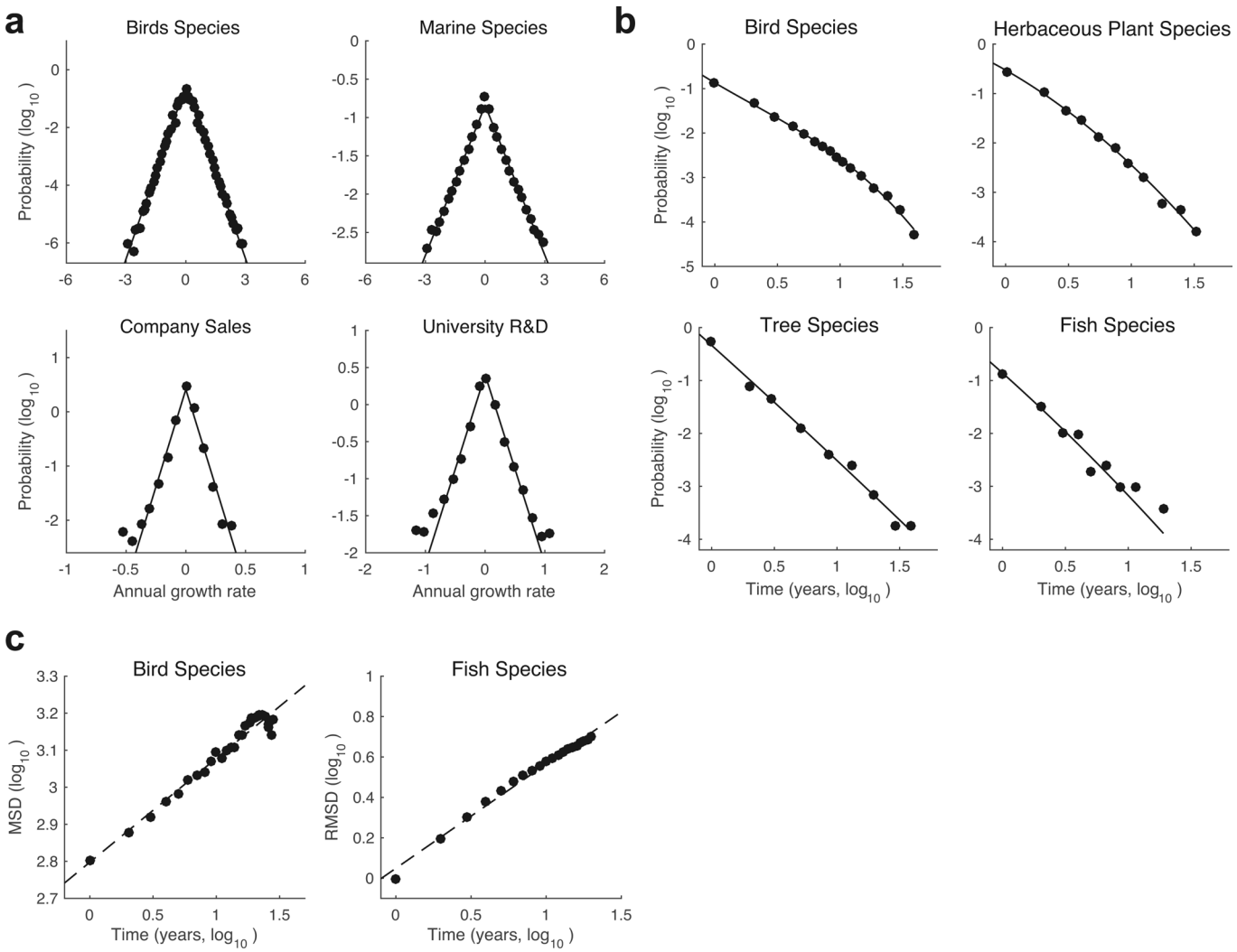
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



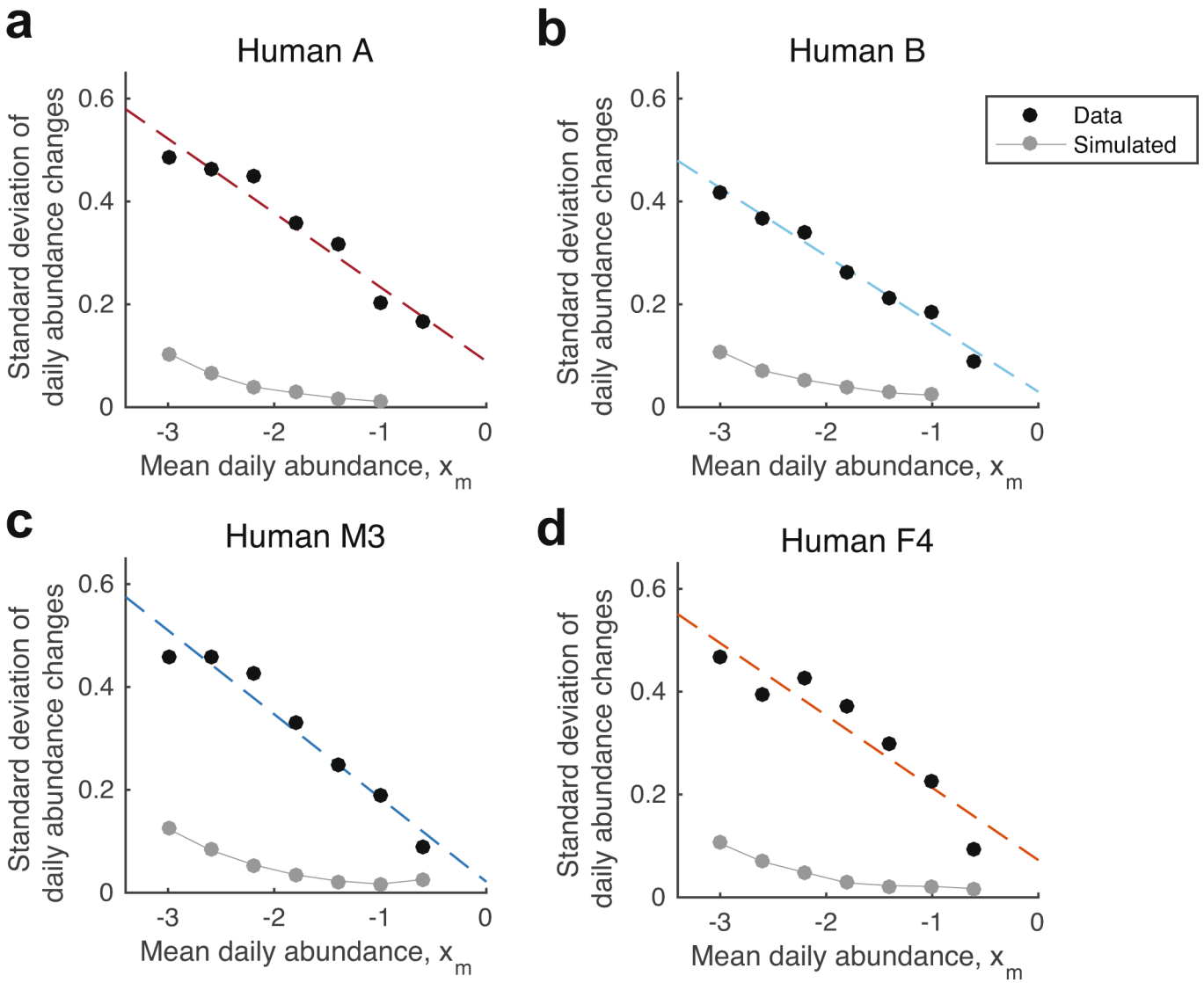
Extended Data Fig. 1 | Distribution of daily abundance changes for human and mouse gut microbiota. Daily abundance changes for an OTU were calculated as the logarithm of the ratio of successive abundances, $\mu = \log(X(t+1)/X(t))$, where $X(t)$ is the relative abundance of the OTU on day t. The distributions of abundance changes for the analyzed bacterial communities closely follow the Laplace distribution: $p(\mu) = (1/2b)\exp(-|\mu|/b)$ in **a**, humans ($b = 0.83 \pm 0.1$, 0.67 ± 0.1, 0.71 ± 0.07, 0.73 ± 0.05, human A, B, M3, F4, respectively; mean ± s.d., across n = 6 equal subsamples of the data, see Methods) and **b**, mice ($b = 0.82 \pm 0.1$, 0.67 ± 0.03, mice on the LFPP and HFHS diets, respectively; mean ± s.d., across n = 3 individual mice). **c**, Daily abundance changes of gut microbiota calculated at different taxonomic resolutions. Microbiota abundance changes were calculated as the logarithm of the ratio of successive abundances, $\mu = \log(X(t+1)/X(t))$, where $X(t)$ corresponds to the sum of abundances on day t for all OTUs within the same taxonomic group. OTUs were defined at the level of 97% sequence similarity of 16S rRNA. **d**, Distribution of normalized daily abundance changes in the human gut microbiota. To obtain the distribution, daily abundance changes for individual OTUs were first normalized by their corresponding standard deviations. The distributions of resulting normalized abundance changes were then combined across all OTUs and across humans A, B, M3, and F4. Barplot shows the Akaike Information Criterion (AIC) calculated based on the maximum likelihood estimate (MLE) fits to the data using the Gaussian and Laplace distributions. The bars show the mean AIC calculated across 10,000 bootstrap samples of the abundance changes data, errors represent the standard deviation, and the triple asterisk *** represents $p<10^{-4}$; across the 10^4 bootstrap samples, the AIC were always smaller for the Laplace distributions, indicating better fits to the data. In all panels, solid lines show MLE fits to the data.



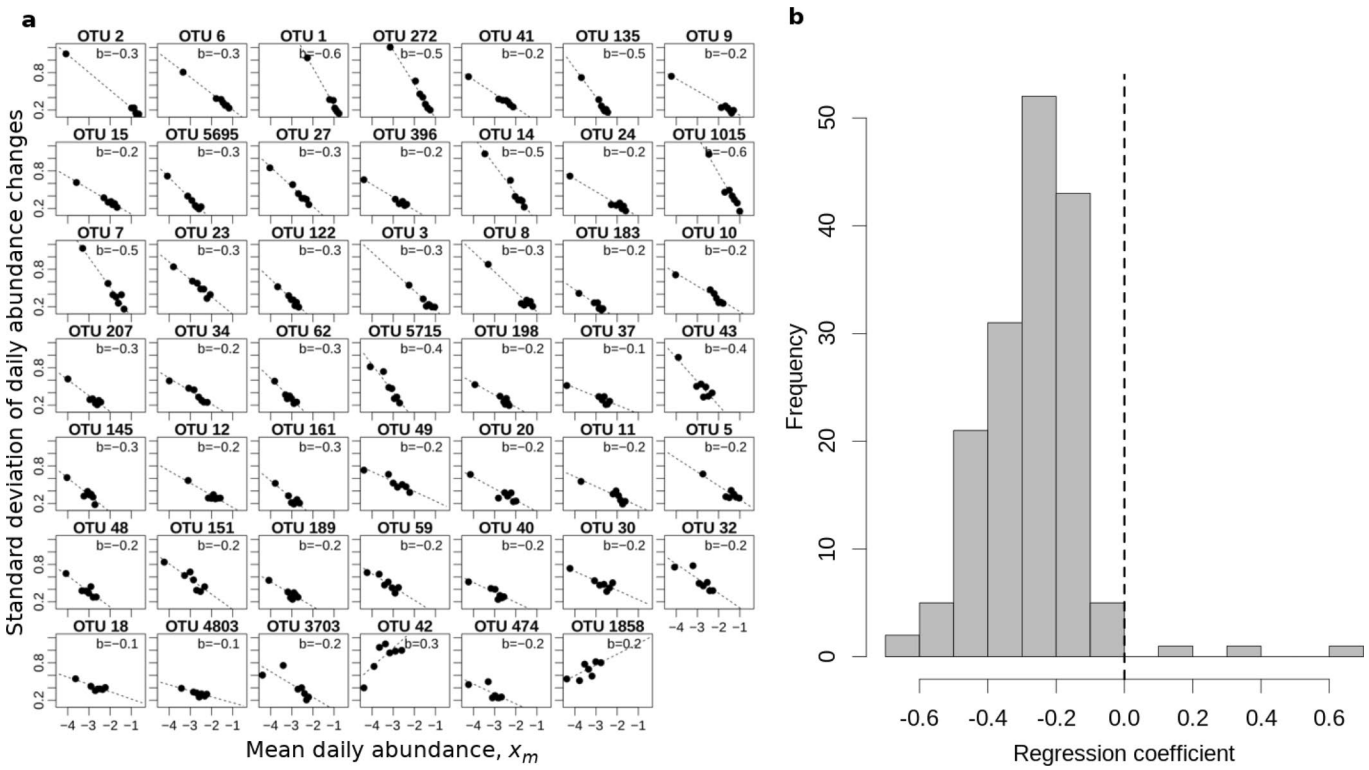
Extended Data Fig. 2 | Compositional effects on the distribution of daily abundance changes for bacterial communities with various diversities. Simulations of bacterial community dynamics (with N=65 OTUs) were performed using the Ornstein-Uhlenbeck process. Black dots represent the resulting distributions of daily absolute abundance changes, while hollow dots represent the distributions of daily relative abundance changes. In panels (**a**-**c**), steady-state OTU abundances in each community were sampled from a power law, where the power law exponent was selected to generate bacterial communities with different diversities; community diversities were quantified by the effective number of species (ENS = e^H, where H is the Shannon diversity). In (**d**), the simulated steady-state OTU abundances were equal to those in the real data. In all panels, solid lines represent Gaussian distribution fits to the simulated data.



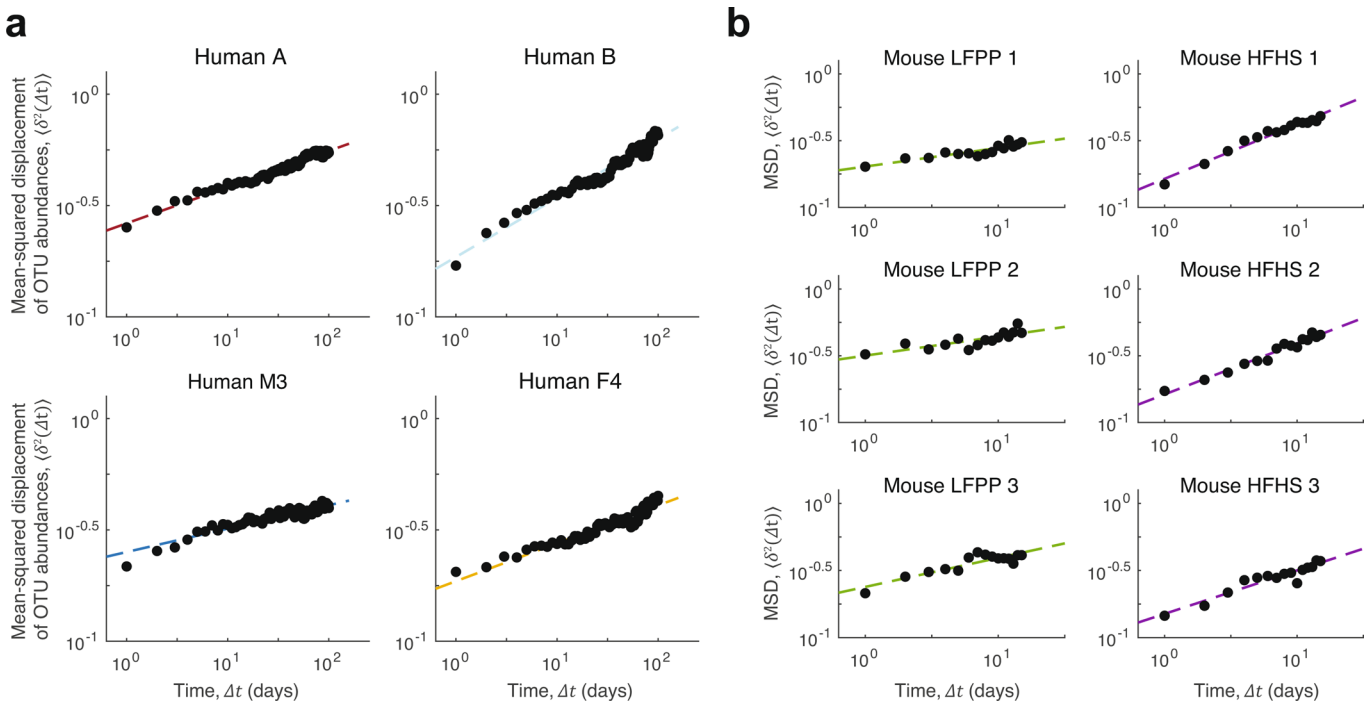
Extended Data Fig. 3 | Growth rate distributions in diverse ecological communities and economic systems. a, Annual growth rate distributions of North American bird populations[16], marine species abundances[22], publicly-traded company sales[23] and university R&D expenditures[24]. Figures were adapted from their original text. Distributions of company sales and R&D expenditures were re-plotted for companies with initial dollar sales of more than one billion dollars and universities with large R&D expenditures (see refs. [23,24] for details). Lines are provided for visual purposes only. **b**, Residence time distributions of species from diverse ecosystems. Figures were adapted from ref. [63], with original data describing North American breeding bird species, estuarine fish species, and plant species collected from both prairie and forest ecosystems. Lines are provided for visual purposes only. **c**, Long-term drift of bird species abundances[16] and North Atlantic fish stock abundances[17]. Figures were adapted from their original text.



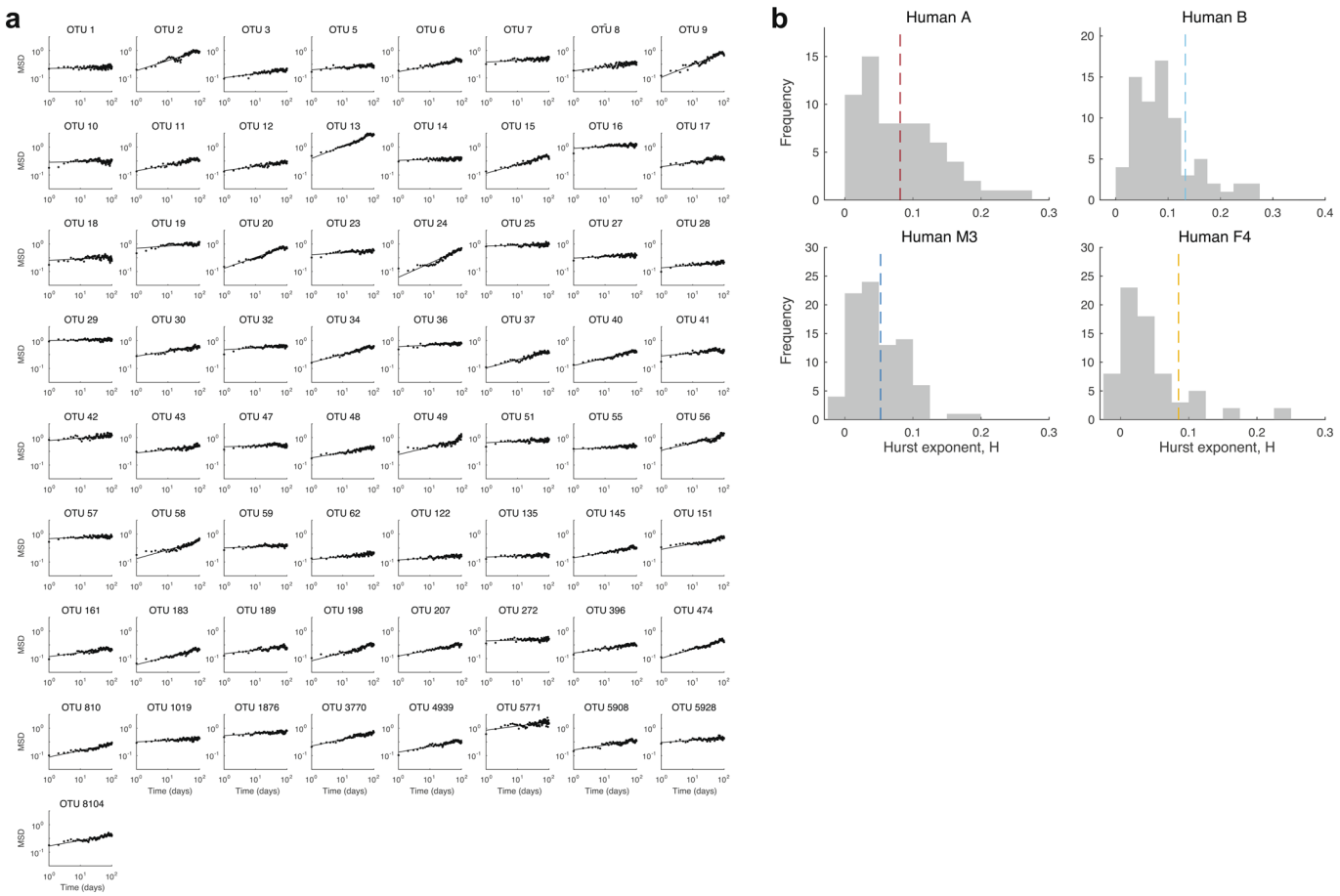
Extended Data Fig. 4 | Variability in daily abundance changes for human gut microbiota. a-d, Black data points show the standard deviation of daily OTU abundance changes as a function of average daily OTU abundances $x_m = \frac{1}{2}[\log(X_k(t+1)) + \log(X_k(t))]$ in humans. Gray data points were generated by simulating time series data in which OTU abundance changes originated exclusively from Poissonian sampling noise. The simulations were performed by sampling sequencing read counts from the underlying OTU abundance distributions empirically observed in humans A, B, M3, and F4 (Methods). Random zero counts were added for each OTU to match the frequency of its zero counts in the real data. Sequencing reads were sampled to the same depths as in the real data. The simulations demonstrate that the observed decrease in the variability of OTU abundance changes as a function of the average daily OTU abundances is not a result of simple Poissonian sampling effects.



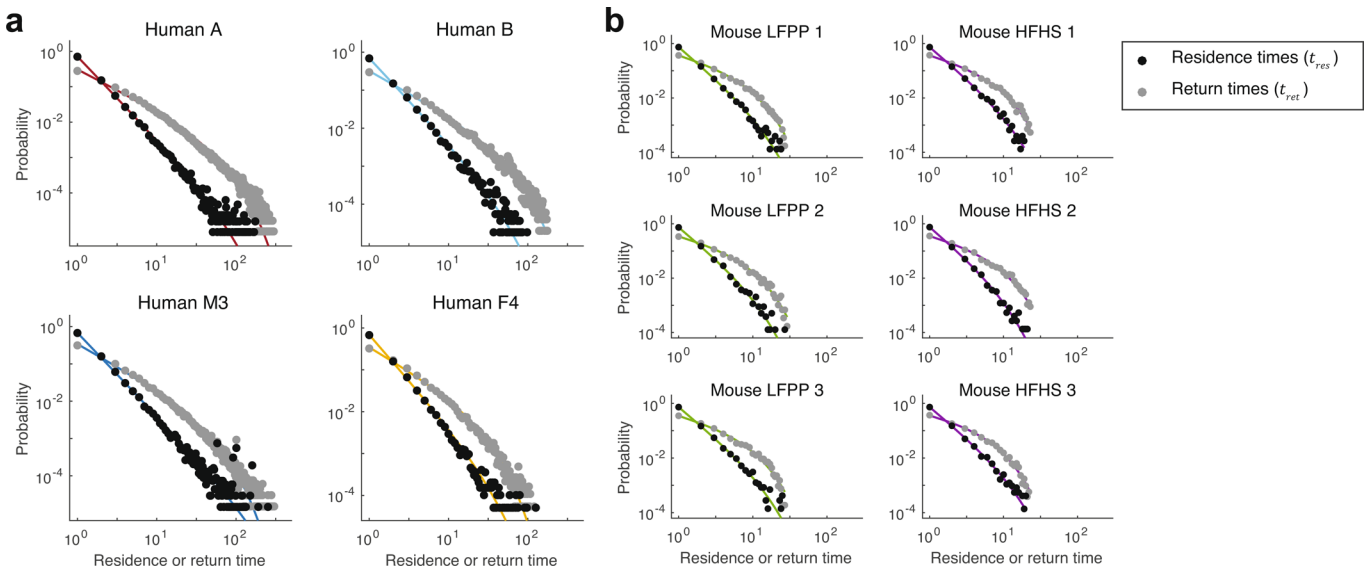
Extended Data Fig. 5 | Standard deviations of daily abundance changes decrease with increasing average daily abundances for individual OTUs. Standard deviations of changes in daily OTU abundance as a function of the average OTU abundance; the standard deviations were calculated, for each OTU separately, across bins of various daily OTU abundances. The abundance bins were selected to have an equal number of time points in each bin. Panel **a** shows the relationships for each OTU in Human A. Dotted lines represent regression fits to the data for each OTU. Panels are sorted according to the p-value of the regression fits, and the slopes of the fits are shown in the top right corner of each panel (n=7). The 48 out of the total of 65 OTUs shown here are significant, based on the FDR cutoff of 10% (the Benjamini-Hochberg method). **b**, Distribution of slopes of the linear regression fits across 154 OTUs and the four human datasets. Only the OTUs with regression p-values significant based on the FDR cutoff of 10% are shown.



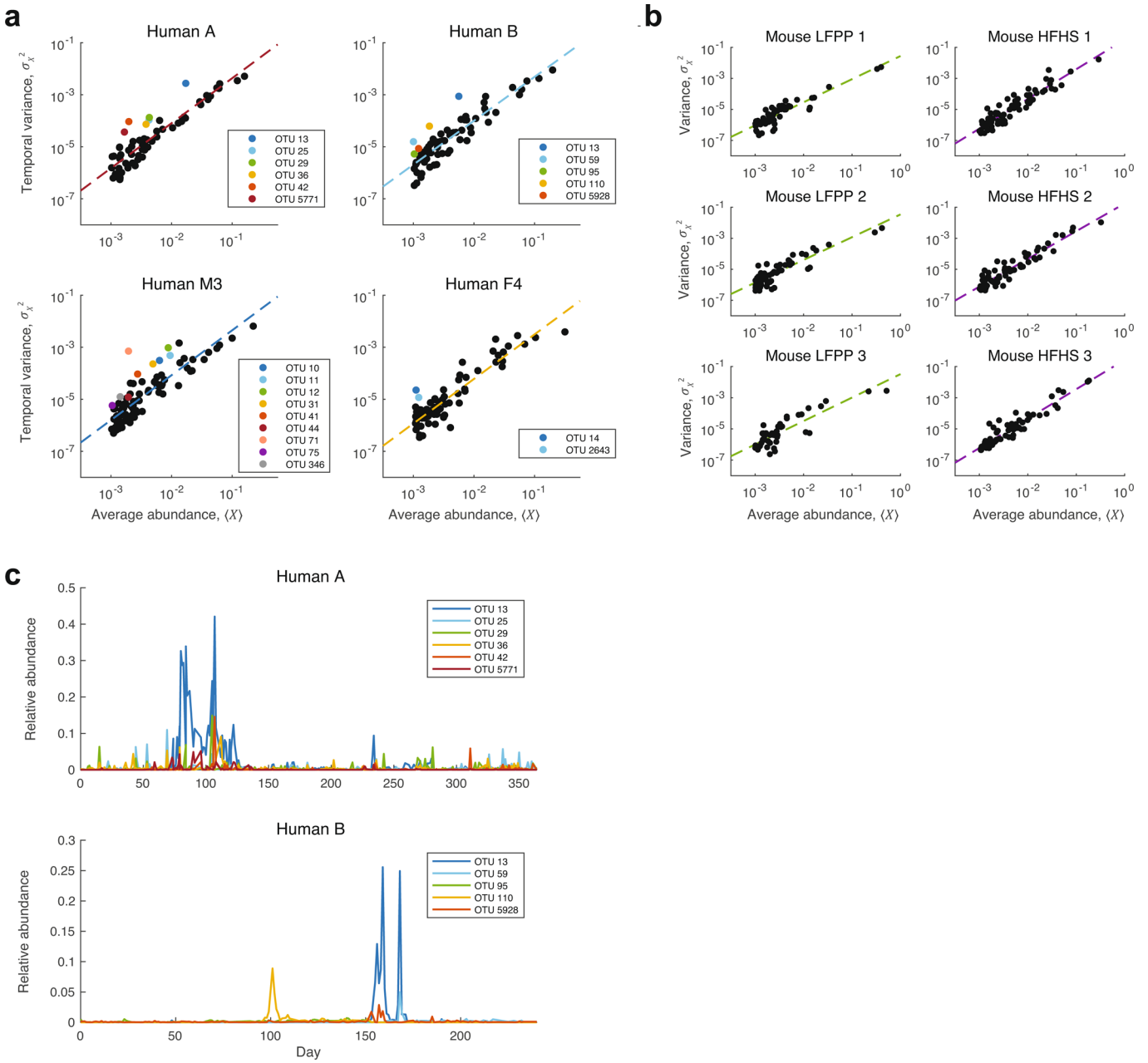
Extended Data Fig. 6 | Long-term drift of gut microbiota abundances in humans and mice. a,**b**, Mean-squared displacements of log-relative OTU abundances $<\delta^2(\Delta t)>$ as a function of time Δt. Dashed lines represent regression fits to the data using the equation of abnormal diffusion $<\delta^2(\Delta t)> \propto \Delta t^{2H}$, where H is the Hurst exponent characterizing the diffusion process. The diffusion Hurst exponents are $H = 0.07 \pm 0.03$, 0.10 ± 0.04, 0.08 ± 0.02, 0.1 ± 0.07 for humans A, B, M3, and F4, respectively, and $H = 0.08 \pm 0.02$, 0.19 ± 0.02 for the LFPP and HFHS mice (mean ± s.d., n = 6 equal subsamples of the data for humans and n = 3 animals for mice, see Methods).



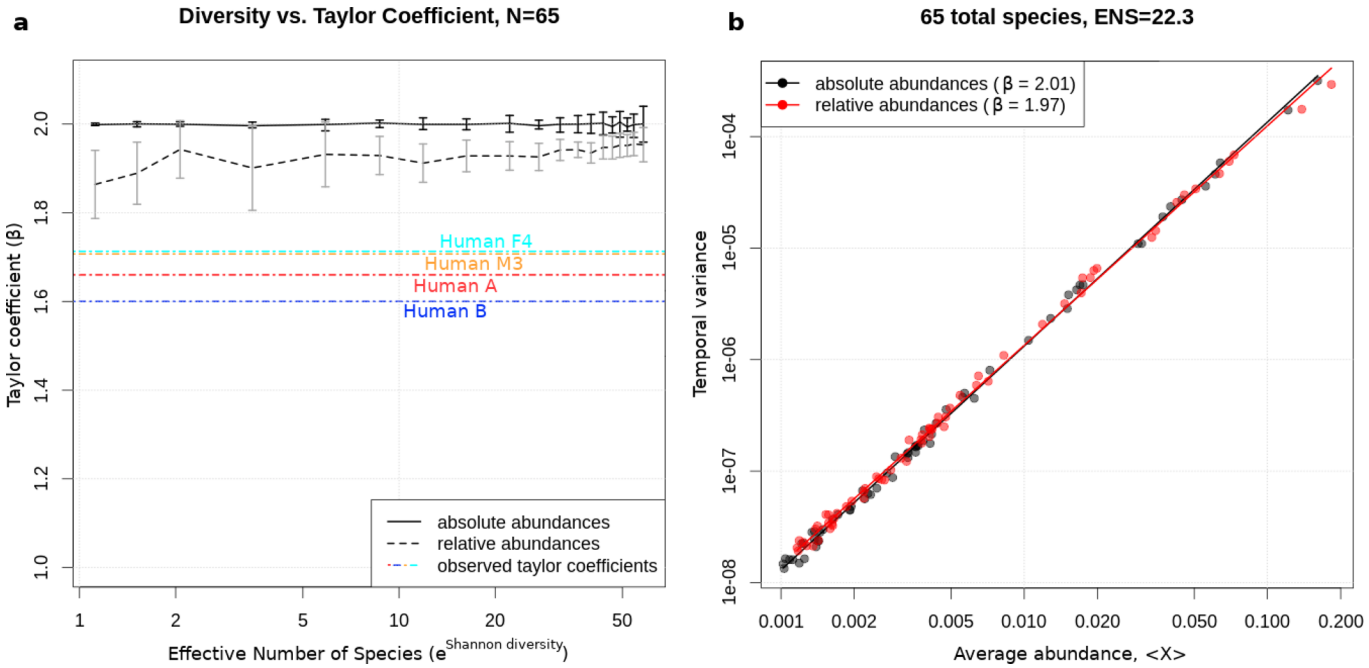
Extended Data Fig. 7 | Long-term drift of individual OTU abundances. a, Mean-squared displacements of log-relative abundance for individual OTUs $<\delta^2(\Delta t)>$ as a function of time Δt. Dashed lines represent regression fits to data, using the abnormal diffusion equation $<\delta^2(\Delta t)>\propto\Delta t^{2H}$, where H is the Hurst exponent characterizing the diffusion process. Panels correspond to individual OTUs from human A. The Hurst exponents were determined using least squared regression fits to n = 100 data points for each OTU (see Methods). **b**, Distributions of Hurst exponents across individual OTUs in humans A, B, M3, F4. The Hurst exponents describing the abundance drift of the entire bacterial communites in each human are indicated by dashed lines.



Extended Data Fig. 8 | Distributions of residence and return times of gut microbiota in humans and mice. Distributions of residence and return times for **a**, human and **b**, mouse gut microbiota. Solid lines represent fits to the data using power laws with exponential tails of the form $p(t) \propto t^{-\alpha} e^{-\lambda t}$. In humans, the power law exponents are $\alpha_{res} = 2.3 \pm 0.04$, 2.2 ± 0.05, 2.2 ± 0.07, and 2.14 ± 0.08 for the residence times and $\alpha_{ret} = 1.1 \pm 0.02$, 0.15 ± 0.03, 1.2 ± 0.05, and 1.09 ± 0.07 for the return times (A, B, M3, F4 respectively; mean ± s.d., n = 6 equal subsamples of the data, see Methods). In mice, $\alpha_{res} = 2.2 \pm 0.04$, 2.2 ± 0.03 and $\alpha_{ret} = 0.72 \pm 0.03$,$0.67 \pm 0.06$ for the LFPP and HFHS diet groups, respectively (mean ± s.d., across n = 3 mice).



Extended Data Fig. 9 | Taylor's power law relationships in human and mouse gut microbiota. Temporal abundance variances σ_X^2 as a function of average species abundances X in (**a**) human and (**b**) mouse gut microbiota. Dashed lines represent least-squares fits to the data using Taylor's power law of the form $\sigma_X^2 \propto X^\beta$. Each dot represents the mean and temporal abundance variance for a single OTU. **a**, The power law exponents in humans are $\beta = 1.66 \pm 0.09, 1.60 \pm 0.08, 1.71 \pm 0.07, 1.71 \pm 0.07$ for humans A, B, M3, and F4, respectively (mean ± s.d., n = 6 equal subsamples of the data, see Methods). Colored dots denote specific OTUs whose abundances on any day exceeded the average abundance over all other days by more than 25-fold (Supplementary Table 1). **b**, The power law exponents in mice are $\beta = 1.49 \pm 0.02$ and 1.86 ± 0.07 for the LFPP and HFHS diets, respectively (mean ± s.d., across n = 3 mice). **c**, The temporal profile of relative abundances of spiking OTUs identified in (**a**) for the two humans (A and B), whose lifestyles were documented over the time series. Major events affecting the gut microbiota of these individuals included travel of individual A to a developing country near day 100, and an enteric infection in individual B near day 150.



Extended Data Fig. 10 | Effects of the compositional nature of microbiota data on Taylor's law exponents. To investigate possible effects of the compositional nature of microbiota data on Taylor's law exponents, simulations were carried out in absolute abundances using a model identical to the one used by Kilpatrick and Ives[35]; in the model, Taylor's law exponents of 2 were expected. After converting absolute bacterial abundances, obtained in the simulations, to relative abundances, Taylor's law exponents were recalculated to assess the effects of data compositionality. **a**, Steady-state OTU abundances for each simulated community were drawn from a power law, with the power law exponents selected to generate a range of community diversities; the community diversities were quantified by the effective number of species (ENS = e^H, where H is the Shannon diversity). Across ENS values (x axis), the resulting Taylor's exponents (y axis), calculated using absolute and relative abundances, are shown by solid and dashed black lines, respectively. Error bars represent the standard deviation across n = 20 independent simulations. The colored semi-dotted lines represent the Taylor's exponents observed in real data in the four human datasets analyzed in our study. **b**, A representative simulation in which steady-state OTU abundances were equal to the mean abundances in Human A.

nature research

Corresponding author(s): Dennis Vitkup

Last updated by author(s): 16/12/2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data and code available on github: https://github.com/brianwji/Macroecological-Relationships/tree/master/matData
Data analysis	Data and code available on github: https://github.com/brianwji/Macroecological-Relationships/tree/master/matData

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data used in this study can be downloaded from the ENA (https://www.ebi.ac.uk/ena/data/view/PRJEB6518 for humans A and B) and MG-RAST databases (https://www.mg-rast.org/linkin.cgi?project=mgp93 for humans M3 and F4; https://www.mg-rast.org/linkin.cgi?project=mgp11172 for mice). Data and code available on github: https://github.com/brianwji/Macroecological-Relationships/tree/master/matData

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n/a
Data exclusions	n/a
Replication	n/a
Randomization	n/a
Blinding	n/a

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging