Title
Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome

Journal
mBio, 14(2)

ISSN
2161-2129

Authors
Wolff, Richard
Shoemaker, William
Garud, Nandita

Publication Date
2023-04-25

DOI
10.1128/mbio.02502-22

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome

Richard Wolff,[a] William Shoemaker,[a] Nandita Garud[a,b]

[a]Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, California, USA
[b]Department of Human Genetics, UCLA, Los Angeles, California, USA

ABSTRACT The human gut microbiome harbors substantial ecological diversity at the species level as well as at the strain level within species. In healthy hosts, species abundance fluctuations in the microbiome are thought to be stable, and these fluctuations can be described by macroecological laws. However, it is less clear how strain abundances change over time. An open question is whether individual strains behave like species themselves, exhibiting stability and following the macroecological relationships known to hold at the species level, or whether strains have different dynamics, perhaps due to the relatively close phylogenetic relatedness of cocolonizing lineages. Here, we analyze the daily dynamics of intraspecific genetic variation in the gut microbiomes of four healthy, densely longitudinally sampled hosts. First, we find that the overall genetic diversity of a large majority of species is stationary over time despite short-term fluctuations. Next, we show that fluctuations in abundances in approximately 80% of strains analyzed can be predicted with a stochastic logistic model (SLM), an ecological model of a population experiencing environmental fluctuations around a fixed carrying capacity, which has previously been shown to capture statistical properties of species abundance fluctuations. The success of this model indicates that strain abundances typically fluctuate around a fixed carrying capacity, suggesting that most strains are dynamically stable. Finally, we find that the strain abundances follow several empirical macroecological laws known to hold at the species level. Together, our results suggest that macroecological properties of the human gut microbiome, including its stability, emerge at the level of strains.

IMPORTANCE To date, there has been an intense focus on the ecological dynamics of the human gut microbiome at the species level. However, there is considerable genetic diversity within species at the strain level, and these intraspecific differences can have important phenotypic effects on the host, impacting the ability to digest certain foods and metabolize drugs. Thus, to fully understand how the gut microbiome operates in times of health and sickness, its ecological dynamics may need to be quantified at the level of strains. Here, we show that a large majority of strains maintain stable abundances for periods of months to years, exhibiting fluctuations in abundance that can be well described by macroecological laws known to hold at the species level, while a smaller percentage of strains undergo rapid, directional changes in abundance. Overall, our work indicates that strains are an important unit of ecological organization in the human gut microbiome.

KEYWORDS human gut microbiome, macroecology, ecology, metagenomics, strains

The human gut microbiome is a complex ecological community composed of tens of trillions of cells that interact directly and indirectly with one another and the host (1–3). Although the precise species compositions of the gut microbiome differ among hosts, healthy adult guts tend to be both ecologically diverse and temporally stable at the species level under normal circumstances (4–7). The ecological stability of the gut community is critical for the preservation of its functional capacity over time,

Invited Editor Rachel Whitaker, University of Illinois at Urbana-Champaign

Editor John W. Taylor, University of California, Berkeley

Address correspondence to Nandita Garud, ngarud@g.ucla.edu.

The authors declare no conflict of interest.

Received 10 November 2022
Accepted 13 January 2023
Published 21 February 2023

and periods of instability and heightened variability are often associated with environmental perturbations or disease states (8–10).

Much as the gut community as a whole is made up of a diverse array of species, within species, populations of gut microbes harbor many genetic variants (11). A growing body of literature highlights the importance of interhost differences in microbiome genetic composition for various aspects of human health, with specific microbial genotypes and strains being associated with the digestion of certain foods (12), a range of host disease risk factors (13), bile and lipid composition (14), and antibiotic resistance (15). Recent studies have begun to characterize how this genotypic diversity changes over time within hosts (7, 11, 16, 17) and have linked longitudinal changes in genetic composition to specific host phenotypes and metabolite levels (18).

Broadly, dynamic changes in genetic variation in the gut microbiome occur at two distinct levels. First, there are changes in the frequencies of lineages that have clonally diverged since their common ancestor colonized the host due to the evolutionary forces of mutation, drift, selection, and recombination. Typically, such lineages differ from one another at a small number [$\mathcal{O}(1)$ to $\mathcal{O}(10^2)$] of single nucleotide variants (SNVs) (16, 17). Second, there are fluctuations in the relative abundances of conspecific strains that do not share an ancestor within the host. When levels of recombination between such strains are sufficiently low, clonal descendants of the initial colonizers may persist within the host as genetically distinguishable populations differing from one another at a number of sites in their shared, core genome [$\mathcal{O}(10^3)$ to $\mathcal{O}(10^4)$] similar to those of strains drawn from unrelated hosts (11, 16, 19–21). Multiple colonization by conspecific strains is evidently under some degree of ecological constraint as only a few strains (typically between one and four) are ever observed within a host at any one time, a phenomenon dubbed "oligocolonization" (16, 22–24). The mechanisms enabling a small number of strains to colonize a host and increase to a high frequency, but preventing a large number of exogenous strains from doing the same, are not yet known. Interestingly, similar colonization patterns have been observed for a number of other host-associated members of the microbiota, both at different human body sites (25) and in other organisms (26, 27).

In healthy, adult hosts, a large majority of strains persist over periods of months to years (7, 11, 16, 18, 28, 29). Moreover, strains within the gut can remain resilient in the face of large perturbations such as antibiotics (23) and fecal microbiome transplants (30). However, little is known about the magnitude of daily fluctuations in genetic composition at either the strain or the lineage level under ordinary conditions, or how such fluctuations ultimately affect the stability properties of the gut community, as longitudinal studies of genetic diversity in the gut have tended to focus on samples collected at multimonth intervals.

In this work, we seek to understand how the genetic composition of the gut changes over time, from daily to multiyear timescales, in four healthy, adult hosts sampled over the course of 6 to 18 months (31). To do so, we leverage concepts from macroecology to examine the dynamics of strains in these four hosts. Macroecology focuses on characterizing statistical regularities in patterns of abundance and diversity within and between ecological communities. A growing body of work has demonstrated that species-level patterns of diversity in a variety of natural microbial communities are well described by macroecological laws (32–35). Many of these macroecological laws can be recapitulated through intuitive ecological models containing few, if any, free parameters (32, 33, 36). Among these successful models is the stochastic logistic model (SLM), which describes the dynamics of a population experiencing rapid stochastic fluctuations induced by environmental noise around a fixed carrying capacity (37). Whether the populations making up a community exhibit regular, statistically quantifiable dynamics and, if so, whether these dynamics can be explained using simple models are fundamentally macroecological questions. In this work, we find not only that the large majority of strains in these healthy hosts exhibit abundance dynamics consistent with an SLM but also that strain abundance fluctuations follow several macroecological laws known to hold among species (32, 33, 35, 36).

Together, our results indicate that daily fluctuations in overall genetic composition within the gut microbiome are largely stationary and that these fluctuations follow broad

macroecological patterns. Thus, several macroecological properties of the human gut known to hold at higher levels of taxonomic organization, including its stability, appear to emerge at the level of strains.

RESULTS

To explore how daily fluctuations in nucleotide diversity and strain abundances translate into stability over periods of months to years, we used high-resolution temporal data from four hosts sampled in the BIO-ML project (31) (see Materials and Methods for further details on sampling).

Temporal stability of intraspecific genetic variation. Within hosts, allele frequencies change over time in gut microbial populations due to mutation, drift, selection, and fluctuations in the relative abundances of strains. While studies examining broad cohorts of sparsely longitudinally sampled individuals indicate that the magnitude of intrahost fluctuations only infrequently approaches that of interhost differences over timescales of months to years (7, 11, 16), more finely resolved temporal trends are less well characterized. To evaluate the stability of the gut community, it is crucial to determine whether temporal fluctuations are stationary or directional.

To assess the stability of intraspecific genetic variation over time, we examined temporal trends in the patterns of nucleotide diversity within our four hosts using F_{ST}, a standard measure of genetic differentiation between populations (see Materials and Methods for further details on F_{ST} calculations). If the genetic composition of a species changes directionally, we expect that samples drawn later in the time course will have a higher F_{ST} value relative to the initial time point than earlier samples. If, in contrast, fluctuations are stationary, then later time points should on average be no more diverged from the initial sample than earlier time points.

To contextualize the magnitude of the variation in the genetic composition over time, we normalized our longitudinal measurements of F_{ST} within hosts by the mean F_{ST} value of the species across hosts. We calculated this species-wide mean F_{ST} using shotgun metagenomic data from 250 North American hosts sampled in the Human Microbiome Project (6, 38), allowing us to better capture the extent of interhost diversity. We refer to the resulting normalized F_{ST} statistic, obtained by dividing each intrahost F_{ST} measurement by the mean F_{ST} across hosts, as F'_{ST}. F'_{ST} will approach or exceed a value of 1 when intrahost fluctuations are of the same magnitude as those of interhost differences and will remain close to a value of 0 when the genetic composition of the population is constant over time.

In line with previous work (7, 11), we observe that changes in the genetic composition within the hosts that we examine (*am*, *ao*, *an*, and *ae*) rarely approach the magnitude of interhost differences, as F'_{ST} remained well below 1 at all time points for all but one species examined (see Fig. S5 in Text S1 in the supplemental material). For the one aberrant species, *Faecalibacterium prausnitzii* in host *ao*, F'_{ST} increases steadily before a rapid increase to above 1 at around the time point of 60 days (Fig. 1B). More typical, however, is the example of *Phocaeicola vulgatus* in host *am*, for which F'_{ST} fluctuates but appears to remain near a long-term steady state (Fig. 1A).

To determine quantitatively whether the fluctuations in genetic composition were stationary or directional, we implemented an augmented Dickey-Fuller (ADF) test for each F'_{ST} time series (39). The ADF test tests the null hypothesis that a time series is nonstationary against the alternate hypothesis that the time series is stationary. Rejecting the null hypothesis with the ADF test is thus evidence that the mean and variance in a temporally varying quantity are time invariant. For 34 of the 45 species (76%) considered, we rejected the null hypothesis at a significance level of a P value of 0.05, indicating that the majority of species exhibit stationary F'_{ST} trends.

Our results suggest that fluctuations in allele frequencies are typically stable, at least when coarse-grained across the whole genome, in these hosts. However, large-scale, rapid changes in allele frequency, as observed for *F. prausnitzii*, also occur.

Strain frequencies. Recent empirical studies using isolate, single-cell, and shotgun sequencing data have demonstrated that at any one time, the human gut microbiome

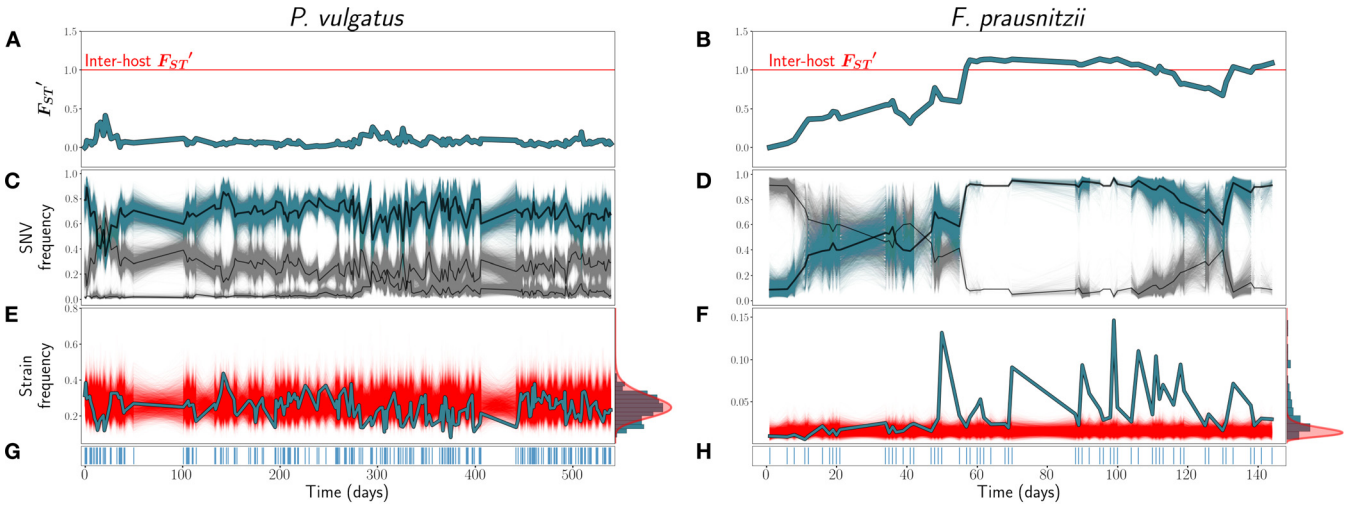

FIG 1 (A and B) F_{ST}' trajectories for *P. vulgatus* (host *am*) (A) and *F. prausnitzii* (host *ao*) (B). (C and D) SNV frequencies of three inferred strains for *P. vulgatus* (C) and two inferred strains for *F. prausnitzii* (D). In black are the inferred strain trajectories. Highlighted in blue are example strains featured further in panels E and F. (E and F) Frequencies of the example strains in blue, with simulations of the corresponding SLM overlaid in red. At the right, the empirical distribution of strain abundances is plotted in blue, and the stationary gamma distribution of abundances (see equation 2) predicted by the SLM is in red. (G and H) Sampling time points. Blue lines indicate that a sample was taken on that day.

is colonized by at most a few distinct conspecific strains (between 1 and 4), which do not share a clonal ancestor within the host (16, 20, 24, 29, 31). Much less is known, however, about the dynamics of strains once they have colonized the host.

To investigate these dynamics, we infer strain genotypes and frequencies using an algorithm adapted from the one described previously by Roodgar et al. (23). This algorithm identifies large clusters of SNVs with tightly correlated frequency trajectories (23), indicative of linkage on a common genomic background. By identifying large clusters, we expect to distinguish the trajectories of deeply diverged strains (for further information on our strain phasing, see Text S1, section 3.2). When no large cluster of tightly linked SNVs was detected within a species, we inferred that only a single strain was present.

Of the 45 species-host pairs examined, 15 (33.3%) harbored multiple strains. Of these 15 pairs, 13 were colonized by two strains, while in 2 separate hosts (*am* and *an*), the species *P. vulgatus* was composed of three strains. The number of fixed differences between strains varied between $\mathcal{O}(10^3)$ and $\mathcal{O}(10^4)$, although due to our conservative filters for both calling SNVs from reads and assigning SNVs to strains, these likely represent underestimates of the true divergence between these strains. For further discussion of interstrain genetic divergence, see Text S1, section 3.1.

As may be expected given our F_{ST}' results, most strains, by visual inspection, exhibit frequency dynamics that are heuristically consistent with stationarity (see Texts S2 to S5 for the strain trajectories of all species). The relative frequencies of most strains appear to fluctuate around a constant value throughout the sampling period. In a typical example of this kind of behavior, the dominant strain of *P. vulgatus* in host *am* fluctuates at around a 60% frequency for more than 500 days (Fig. 1C). However, in a minority of cases, the strain frequencies shifted dramatically throughout the sampling period. The most striking example of this is *F. prausnitzii* in host *ao*, which we have already seen underwent fluctuations in F_{ST}' (Fig. 1B) of the magnitude of interhost differences in genetic composition. In this species, an initially rare strain almost fully supplants the initially dominant strain within the span of 60 days before a partial reversion later in the time course (Fig. 1D). In another case, *Parabacteroides distasonis* in host *ao*, a single strain that initially falls below the detection threshold increases to a detectable abundance midway through the sampling time course (Text S3). Similarly, the minor strain of *P. vulgatus* in host *am* is initially very rare before rapidly increasing in frequency at around day 300 and subsequently reverting to an intermediate steady state below its maximum frequency. This subtle shift in strain frequencies is not, interestingly, detected by our ADF test of F_{ST}' as a departure from stationarity, likely due to

the fact that the absolute magnitude of the shift in allele frequencies between the beginning and the end of the time course is small (less than 5%).

We further note the striking visual correspondence between the F'_{ST} and strain frequency trajectories, as is evident in Fig. 1 for both *P. vulgatus* and *F. prausnitzii*. Because fluctuations in the relative frequencies of cocolonizing strains with respect to one another determine allele frequencies at a very large fraction of all polymorphic sites, genome-wide average diversity statistics like F'_{ST} will reflect strain dynamics. This correspondence is therefore supporting evidence that the overwhelming majority of genetic variation in these species is due to fixed differences between strains rather than among lineages belonging to strains. However, as is evident from the example of the minor strain of *P. vulgatus* in host *am*, subtle but important strain dynamics can also be obscured when considering only genome-wide average patterns of diversity.

Stochastic logistic model. In the sections above, we have seen that the genetic compositions of most species examined exhibit stationary dynamics over the timescale of observation. We hypothesized that this behavior might result from the underlying strains fluctuating around fixed absolute carrying capacities.

To test this hypothesis, we assessed the fit and predictive capacity of the stochastic logistic model (SLM), a model of a population experiencing stochastic fluctuations around a fixed abundance. Recent work in microbial ecology has demonstrated the power of minimal models like the SLM, requiring the fit of no free parameters, to reproduce qualitative and quantitative features of natural microbial community dynamics (32, 33, 36, 37, 40). Here, we tested the capacity of the SLM to forecast future strain behavior when trained on an initial subset of time points. By training on only a subset of initial points, we can assess whether strain dynamics are consistent over time.

Under the assumptions of the SLM, each population, *i*, has a long-term carrying capacity, K_i, and temporal fluctuations in abundance around this value are driven by environmental noise with amplitude σ_i. The dynamics of a population governed by an SLM can be expressed with the following stochastic differential equation:

$$\frac{dx_i}{dt} = \frac{x_i}{\tau_i}\left(1-\frac{x_i}{K_i}\right) + \sqrt{\frac{\sigma_i}{\tau_i}}x_i\eta(t) \tag{1}$$

where τ_i^{-1} is the growth rate and $\eta(t)$ is a Brownian noise term.

Populations following an SLM may experience large fluctuations in abundance over short timescales and may even be temporarily found far from their long-term average value, but these deviations will be transient. Over long timescales, the observed distribution of abundances will converge to a stationary gamma distribution (32):

$$\rho(x_i) = \frac{1}{\Gamma(2\sigma_i^{-1}-1)}\left(\frac{2}{K_i\sigma_i}\right)^{2\sigma_i^{-1}-1} x^{2\sigma_i^{-1}-2}\exp\left(-\frac{2}{K_i\sigma_i}x\right) \tag{2}$$

To determine whether strain trajectories could be described by an SLM, we first obtained time series of strain abundances by multiplying the relative frequencies of the strains inferred in the previous section by the relative abundance of the species to which they belong.

Next, we estimated K_i and σ_i from the first one-third of time points for each strain. K_i and σ_i are not free parameters but rather are functions of the mean and variance of the observed abundances (for details, see Text S1, section 4). To assess quantitatively whether the time series of strains in our cohort could be adequately described by an SLM, we developed and implemented a goodness-of-fit test. This test determines whether the transitions between subsequent time points are consistent with an SLM (for further details, see Text S1, section 5). Qualitatively, if a strain follows an SLM, its average and variance in abundance in the latter two-thirds of the time series should match those of the former one-third, and the strain should have a tendency to revert to its carrying capacity, K_i.

Returning to our case studies, we see that the dominant strain of *P. vulgatus* is well described by the SLM (Fig. 1E). The true abundance trajectory (in blue) explores largely the same space as that of the SLM simulations (in red). Moreover, the empirical distribution of

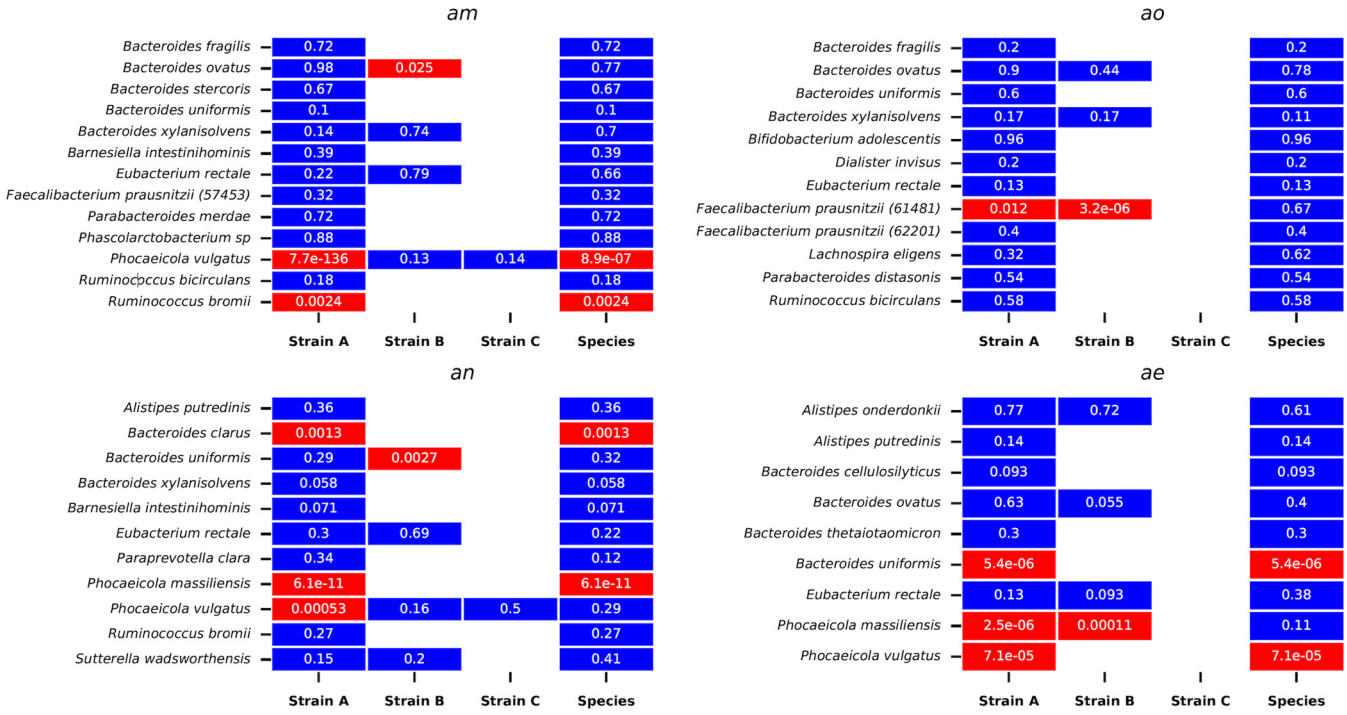

FIG 2 Results of the SLM goodness-of-fit test, by host. Totals of 79% (49/62) of strains and 86.6% (39/45) of species exhibit stochastic logistic dynamics across the sampling interval. The percentage of strains passing the test varied among hosts, with 83% passing in host *am*, 87% passing in *ao*, 75% passing in *an*, and 69% passing in *ae*. The *P* value associated with each strain or species is shown in white within each cell.

abundances across the entire time course appears to approach the stationary gamma distribution (equation 2) predicted by the SLM (Fig. 1E, right). In contrast, the dynamics of the invasive strain of *F. prausnitzii* deviate strongly from the SLM (Fig. 1F).

Overall, 79% (49/62) of strains passed our goodness-of-fit test, with 83% passing in host *am*, 86% passing in *ao*, 73% passing in *an*, and 69% passing in *ae* (Fig. 2). While the dominant strain of *P. vulgatus* passes, both strains of *F. prausnitzii* in host *ao*, as well as the minor strain in host *am*, fail the SLM. Thus, the SLM recapitulates the qualitative behavior of a large majority of strains even when trained on only a subset of initial points while also having the power to discriminate instances in which the dynamics of strains are evidently quite nonstationary.

Moreover, the likelihood of exhibiting SLM dynamics is independent of the presence of other strains. Of the 32 strains for which another conspecific strain was present, 24 (75%) passed the SLM test, while among the 30 singly colonizing strains, 25 (83%) passed the test. Thus, while singly colonizing strains tend to be moderately more likely to exhibit stable dynamics, the difference in pass rates is not statistically significant ($\chi^2 = 0.24$; *P* value = 0.62), indicating that the presence of conspecific strains is not *prima facie* destabilizing for a focal strain.

Next, we conducted an identical test of the SLM at the species level. Overall, 86% (39/45) of species exhibited dynamics consistent with an SLM. In the case of *F. prausnitzii*, for instance, the abundance of the species overall fluctuated stably, obeying the SLM. Thus, despite a partial replacement event, the total abundance of the species remained roughly constant. One interpretation of this observation is that these strains strongly compete with one another for the same species' niche. Interestingly, *F. prausnitzii* is known to experience higher rates of replacement over multiyear timescales than other gut commensals, and these replacements are associated with alterations in the levels of plasma metabolites that affect host immunity (18). However, we emphasize that the replacement was only partial, and the "displaced" strain recovered temporarily to an intermediate abundance. This example highlights the complexity of strain dynamics as well as their potential relevance for host phenotypes.

Interestingly, none of the six species that exhibited non-SLM dynamics also failed our ADF test of intraspecific genetic stationarity. Of these six species, five harbored only a single strain, and in such cases, changes in the abundance of the strain do not imply changes in the levels of intraspecific genetic diversity, as they do in the multistrain case. The only species that exhibited non-SLM dynamics and harbored multiple strains was *P. vulgatus* in host *am*. As noted above, although the minor strain of this species exhibits highly non-SLM dynamics, contributing to the species as a whole failing the SLM test as well, the genome-wide average levels of genetic diversity as measured by F'_{ST} change relatively little. Overall, examining changes in intraspecific genetic diversity and species abundance yields orthogonal information about population dynamics. For instance, the genetic composition of a species may change dramatically while the total abundance of the species fluctuates stably, as with *F. prausnitzii*. Conversely, a species' abundance may change directionally while the genetic composition of the population remains roughly constant, as might occur during a rapid demographic expansion of a population with low initial diversity.

Macroecology of strains. Much as the dynamics of individual strains can largely be recapitulated with a single relatively simple model, we can also attempt to parsimoniously characterize patterns of variability across strains collectively. Such low-dimensional representations of complex community dynamics are the natural purview of macroecology, which attempts to characterize variation within and among communities by observing the statistical patterns of abundance, distribution, and diversity across their constituent members. In many kinds of microbial ecosystems, including the human gut, patterns of species abundance and distribution have been shown to broadly follow a number of macroecological laws, including Taylor's law and a gamma distribution of abundance fluctuations (32, 33, 35). We show here that these macroecological relationships also characterize patterns of variation in the abundance of strains across our cohort.

The first pattern examined is power-law scaling between the mean and variance in abundance, known in ecology as Taylor's law, which can be stated as:

$$\sigma^2_{x_i} \propto \langle x_i \rangle^{\alpha} \tag{3}$$

where $\langle x_i \rangle$ and $\sigma^2_{x_i}$ are the mean and variance of x_i, respectively, and α is the scaling exponent of the power law.

Many mechanisms can give rise to Taylor's law. For instance, when the only source of variability between communities (or, in our case, longitudinal samples) is due to sampling noise, Taylor's law exponent α will equal 1. In contrast, in communities where the scale of fluctuations is independent of abundance, that is, where all populations have identical per-capita fluctuations, α will equal 2 (32, 35) (for further details, see Text S1, section 6). We observed Taylor's law scaling with an exponent of an α of 1.8 among all strains (Fig. 3A), mirroring previous findings at the species level (35). This Taylor's law exponent indicates that higher-abundance strains are proportionally less variable than lower-abundance strains ($\alpha < 2$) and that variation in strain abundance is not driven solely by sampling noise ($\alpha > 1$) but rather reflects true, underlying biological variability. However, the existence of power-law scaling between mean and variance cannot, by itself, conclusively prove that any specific ecological model governs community dynamics. Indeed, the fit of the SLM does not depend on the existence of Taylor's law scaling, or vice versa, as the SLM can hold with arbitrary mean and variance values.

The next pattern considered is the abundance fluctuation distribution (AFD), the overall distribution of abundances of a population over time. It is known that in a variety of microbial ecosystems, the AFD of many species tends to approach a gamma distribution (32). As discussed above, a population governed by stochastic logistic dynamics will tend toward a gamma distribution of abundances over long timescales (see, for instance, the histogram of abundances for the dominant strain of *P. vulgatus* in Fig. 1E, right). Given the generally excellent fit of the SLM to the population time series, the abundances of strains might generically be expected to each individually follow a gamma distribution. In Fig. 3B, we see that the distributions of strain abundances are indeed, on average, well described by a gamma distribution (black dots and blue line), although some individual strains (gray lines)

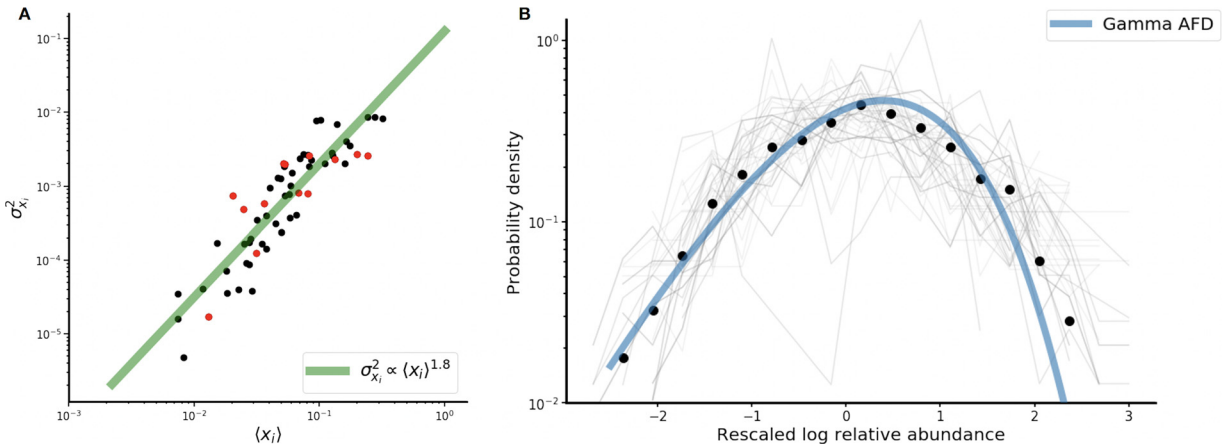

FIG 3 (A) Scaling of the variance in strain abundance with the mean abundance obeys Taylor's law with an exponent of 1.8. Black dots are strains passing the SLM, while red dots are strains failing the SLM. (B) Strain abundances approximately follow a gamma distribution, which is the stationary distribution of the SLM. Black circles are the average probability densities of the rescaled abundances across all strains, and the blue line is the gamma fit of the bin means of the rescaled distributions. Light-gray lines are the individual rescaled abundance distributions for each strain individually (62 in total).

deviate somewhat from the gamma stationary distribution. Recalling that the SLM of a given strain is uniquely determined by its mean and variance, it is apparent that the collapse of the AFDs to a single gamma distribution is in fact a consequence of the strong constraint that Taylor's law places on these quantities across strains.

DISCUSSION

In this study, we sought to characterize the within-species population dynamics of the human gut microbiomes of four healthy hosts. Previous efforts have shown that within-host changes in the genetic composition of gut microbial populations over time tend to be small compared to interhost differences (7, 11). We build on this result by demonstrating that at a daily temporal resolution, intraspecific diversity tends to fluctuate around a long-term average value within the hosts examined over periods of years. We show, crucially, that the abundance fluctuations of a large majority of strains that we detect can be predicted by the stochastic logistic model (SLM) of growth, a model that also recapitulates fluctuations at the species level (32, 33, 37). Finally, we find that empirical patterns of strain abundance variation in these hosts follow macroecological laws, which have also previously been demonstrated to hold at the species level, including Taylor's law and a gamma abundance fluctuation distribution (32, 33, 35). Together, our results indicate that many of the broad properties of the gut observed at higher taxonomic levels of organization, such as its ecological and functional stability, may in fact emerge at the level of strains.

While the SLM was able to sufficiently describe strain dynamics for the majority of strains across species, its success was not universal, and deviations from this typical pattern were also informative. In one host, for instance, two strains of *F. prausnitzii* appear to undergo rapid strain replacement and fail the test. Whether this replacement was due to shifting environmental conditions or direct interstrain competition is unclear. Regardless, our work indicates that any successful description of gut microbial dynamics must incorporate the possibilities of both coexistence and rapid replacement. Over very long timescales, in fact, strain replacement may dominate the stable dynamics that we observe here. Previous work (16, 28) suggests that over the course of decades, a large fraction of strains are ultimately replaced. One hypothesis is that this timescale reflects a waiting time for large environmental perturbations such as antibiotics (23, 41) or bowel cleanses (42), but this is just one of many hypotheses. Indeed, this hypothesis is partially challenged by Roodgar et al. (23), where the strain content of an adult gut was perturbed during a course of antibiotics but ultimately largely recovered to its pretreatment state. This antibiotic study

is a powerful demonstration of the stability of strains even in the face of large perturbations. Investigating the possible explanations for the discrepancy between years-long and decades-long population dynamics at the strain level is an important problem that can be addressed with more extended timescales of observation.

While, in this work, we characterize the population dynamics of strains as ecological units, strains are by no means internally genetically homogeneous. The deep divergences that we detect between conspecific strains [$\mathcal{O}(10^3)$ to $\mathcal{O}(10^4)$ SNVs] are in fact genetic backgrounds, representing timescales of divergence likely far preceding the colonization of the host. However, individual lineages bearing these backgrounds can differ from one another both at sites in the core genome and in gene content, and the relative frequencies of these different lineages with respect to one another can change due to evolution. Previous studies have shown that at the level of lineages belonging to a strain, populations of gut microbes can experience both rapid selective sweeps (16, 31) and diversification into stably coexisting lineages (17).

How evolution impacts the ecological dynamics of strains and how, in turn, these ecological dynamics constrain and channel evolution are active areas of research (43). In the context of the SLM, these ecoevolutionary feedbacks can be viewed as tuning a strain's carrying capacity, K_i; growth rate, τ_i^{-1}; and sensitivity of the growth rate to environmental perturbation, σ_i. Naively, it is expected that evolution would tend to increase the carrying capacity while minimizing the sensitivity of growth to abiotic fluctuations, but evolutionary modifications driving changes in one quantity may affect the other. The observed power-law scaling between the mean and variance in abundance (Taylor's law) is, in essence, a constraint on K_i given σ_i, and vice versa. The SLM thus not only describes ecological dynamics but also, in conjunction with the empirical observation of macroecological laws, provides a useful framework for investigating the ecological effects of adaptation.

The SLM is ultimately a phenomenological model, not a mechanistic one, and its success at the strain level does not explain why strains coexist. How and why closely related strains coexist in the human gut are two of the central biological questions raised by our results. Spatial segregation between strains, perhaps occupying different colonic crypts, or partitioning luminal and mucosal niches, could contribute to the observed pattern of strain coexistence (44–46), much as it does among *Cutibacterium acnes* strains inhabiting different pores on the facial microbiome (25). However, the spatial structure is far from the only mechanism that can foster coexistence between strains. For instance, differences in genetic content at polysaccharide utilization loci may contribute to intrahost metabolic niche differentiation, potentially favoring the coexistence of closely related strains (47). Moreover, stably coexisting strains have been reported under laboratory conditions as well (26, 48). In these experiments, strains may coexist by finely partitioning some aspect of the abiotic environment, by engaging in ecological interactions (e.g., cross-feeding), or by some combination of both. Indeed, recent theoretical work suggests that even subtle differences in resource uptake rates under high- and low-nutrient conditions may, in the presence of a temporally variable environment, lead to the coexistence of small numbers of closely related strains (49). Investigating which of these mechanisms promotes strain coexistence in the human gut microbiome and identifying the relevant genomic architectures are important avenues for future research.

We note that the four hosts examined here are not a representative sample of the full diversity of human lifestyles. For instance, all hosts were between the ages of 21 and 37 years and resided in the United States at the time of sampling. The proportion of strains exhibiting stable or unstable dynamics may vary in different cohorts, but the tests that we developed here will nonetheless be useful in identifying such differences. A future avenue of work will be to assess the generality of these findings in different cohorts, using new time series, which are at least as long and densely temporally sampled as the BIO-ML data analyzed here and are of a similar quality, including diseased or perturbed cohorts, which may exhibit quite different dynamics.

Finally, our work highlights the importance of strains in understanding community structure and dynamics in the human gut microbiome (26, 49). The ambiguity

surrounding the bacterial species concept is well known (50), and reasonable alternatives have been proposed (51), but operationally, species are nonetheless the predominant focus of attention in gut microbiome ecology. This focus is reasonable, as within-host strain structure is a comparatively recent discovery (16, 24, 29), and 16S rRNA gene sequencing provides an inexpensive, high-throughput means to examine community dynamics. However, it is reasonable to propose that for the human gut, and perhaps other microbial ecosystems, many higher-level macroecological patterns of abundance and diversity may originate at the level of strains.

MATERIALS AND METHODS

Longitudinal data. To investigate the temporal dynamics of the human gut microbiome, we analyzed densely sampled shotgun metagenomic time series data from four hosts from the BIO-ML project (accession number: PRJNA544527) (31). By analyzing shotgun metagenomic sequences, we capture longitudinal patterns of intraspecific genetic variation for many bacterial species in these communities simultaneously. A total of 402 samples were drawn from these four individuals (hosts *am*, *ao*, *an*, and *ae*), with 206 samples coming from host *am*, 74 from *ao*, 63 from *an*, and 59 from *ae* (see Table S1 and Fig. S2 in Text S1 in the supplemental material for further details on sampling). All four hosts were healthy adults between the ages of 21 and 37 years, three of whom were male and one of whom was female, and all were residing in the United States at the time of sampling (see Data Set S1 for further details). Crucially, for our purposes, these hosts were sampled at a very fine temporal resolution, with a median interval between successive samples of either 1 or 2 days in each host, over a period of 5 months (host *ao*) to 18 months (host *am*).

Aligning reads. To call single nucleotide variants (SNVs) and gene content, we aligned shotgun metagenomic reads to a panel of species that were prevalent and abundant within each host using MIDAS (52) (see Text S1, section 2, for further details on the bioinformatic pipeline employed). In total, we detected 45 species across the four hosts that met our coverage and prevalence criteria. To reflect a more recently published taxonomy (53), the names of three species (*Phocaeicola vulgatus*, *Phocaeicola massiliensis*, and *Lachnospira eligens*) were amended from the names of these species native to MIDAS (*Bacteroides vulgatus*, *Bacteroides massiliensis*, and *Eubacterium eligens*).

Stationarity of intraspecific genetic variation. To calculate F_{ST}, a measure of subpopulation differentiation, we used the estimator:

$$F_{ST} = \frac{\pi_{BT} - \pi}{\pi_{BT}} \tag{4}$$

where π is the nucleotide diversity within a population and π_{BT} is the level of diversity between populations.

The nucleotide diversity, π, is a classical population genetic measure of polymorphism, representing the average number of pairwise SNV differences between randomly chosen members of a population. To determine π for a given species within a sample, we used the estimator:

$$\pi = \frac{1}{|G|}\sum_{i=1}^{|G|} \frac{r_i}{d_i}\frac{a_i}{d_i - 1} + \frac{a_i}{d_i}\frac{r_i}{d_i - 1} \tag{5}$$

where r_i is the count of the reference allele at site i, a_i is the count of the alternate allele, d_i is the depth of coverage, and $|G|$ is the total number of sites in the genome. This quantity was calculated after first excluding sites with low read depths ($<5\times$), as reliable estimates of true allele frequency cannot be made for such sites. Our π calculations follow the same methodology as the one described previously by Schloissnig et al. (11) in an early, foundational work characterizing patterns of genetic diversity in gut microbial populations within and across hosts and are meant to be directly comparable.

Similarly, π_{BT}, the diversity between time points, was calculated as:

$$\pi_{BT} = \frac{1}{|G|}\sum_{i=1}^{|G|} \left(\frac{r_{t_1 i}}{d_{t_1 i}}\frac{a_{t_2 i}}{d_{t_2 i}} + \frac{a_{t_1 i}}{d_{t_1 i}}\frac{r_{t_2 i}}{d_{t_2 i}} \right) \tag{6}$$

where r_{ji}, a_{ji}, and d_{ji} are the reference allele count, alternate allele count, and depth of coverage of site i in sample j, respectively, and $|G|$ is the total number of sites in the genome.

We calculated π for each species in each sample in each host and π_{BT} for each sample relative to the initial time point.

To calculate the F_{ST} between the initial time point (sample i) and sample j, we used the formula:

$$F_{ST}^{(ij)} = \frac{\pi_{BT} - \frac{(\pi_i + \pi_j)}{2}}{\pi_{BT}} \tag{7}$$

Finally, we obtained our normalized statistic $F_{ST}^{'}$ by dividing $F_{ST}^{(ij)}$ by the species mean interhost F_{ST}, estimated from a panel of 250 North American subjects sequenced in the Human Microbiome Project (6, 38). To determine this species mean F_{ST}, we first calculated the pairwise F_{ST} for each pair of samples in which a species appeared and then took the mean of these values.

To implement the augmented Dickey-Fuller test on each time series of $F_{ST}^{'}$ values, we used the adfuller function from the python statsmodels library (54).

Strain inference. To phase strains, we use a modified version of the allele frequency trajectory clustering algorithm developed by Roodgar et al. (23). While the approach of Roodgar et al. was appropriate for their purposes, namely, detecting selective sweeps of linked variants that deviated substantially from the overall background, our clustering scheme is designed to detect only large clusters of SNVs (minimally >1,000 SNVs within a cluster) whose linkage patterns are consistent with perfect linkage on a single haplotype background. Our choice of 1,000 SNVs as a cutoff was informed by previous work estimating the typical scale of genetic divergence between strains found in different hosts (see Text S1, section 3.1, for further information). While lineages can and do diverge as a result of diversifying evolution within hosts (17, 31), by imposing a minimum cluster size of 1,000 core-genome SNVs, we expect largely to exclude cases of within-host diversification. When no cluster of >1,000 SNVs was detected, only a single strain was inferred to be present.

Macroecology. To fit Taylor's law (Fig. 3A), we used the polyfit function from the python numpy library to fit a power-law regression between the mean and the variance of each strain's abundance distribution. To obtain the log-rescaled gamma AFD (Fig. 3B), each strain's abundance distribution was first log rescaled and then normalized to have zero mean and unit variance. We then binned each strain's rescaled abundance distribution into 20 evenly spaced bins and fit the gamma AFD to the bin-wise mean across strains. To perform the gamma AFD fit, we adapted code from J. Grilli (32).

Data availability. All necessary metadata, as well as the source code for the MIDAS metagenomic pipeline, downstream analyses, and figures, are available on GitHub at https://github.com/garudlab/StrainStability.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, PDF file, 2.7 MB.

TEXT S2, PDF file, 2.6 MB.

TEXT S3, PDF file, 1.5 MB.

TEXT S4, PDF file, 1.7 MB.

TEXT S5, PDF file, 2.5 MB.

DATA SET S1, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank Colin Kremer for his crucial comments on the manuscript, Van Savage for early discussions on this work, Jacopo Grilli for code and important discussions, as well as members of the Garud lab for continued feedback. Finally, we thank two anonymous reviewers for their constructive feedback.

This work was supported by the NSF Postdoctoral Research Fellowships in Biology Program under grant no. 2010885 (W.S.) as well as the Paul Allen Foundation, the UCLA Hellman Fellowship, and the Research Corporation for Science Advancement (N.G.).

We declare no competing interests.

REFERENCES

1. Sender R, Fuchs S, Milo R. 2016. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol 14:e1002533. https://doi.org/10.1371/journal.pbio.1002533.
2. Venturelli OS, Carr AC, Fisher G, Hsu RH, Lau R, Bowen BP, Hromada S, Northen T, Arkin AP. 2018. Deciphering microbial interactions in synthetic human gut microbiome communities. Mol Syst Biol 14:e8157. https://doi.org/10.15252/msb.20178157.
3. Daniel N, Lécuyer E, Chassaing B. 2021. Host/microbiota interactions in health and diseases—time for mucosal microbiology! Mucosal Immunol 14:1006–1016. https://doi.org/10.1038/s41385-021-00383-w.
4. Coyte KZ, Schluter J, Foster KR. 2015. The ecology of the microbiome: networks, competition, and stability. Science 350:663–666. https://doi.org/10.1126/science.aad2602.
5. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. 2012. Diversity, stability and resilience of the human gut microbiota. Nature 489:220–230. https://doi.org/10.1038/nature11550.
6. Integrative HMP (iHMP) Research Network Consortium. 2019. The Integrative Human Microbiome Project. Nature 569:641–648. https://doi.org/10.1038/s41586-019-1238-8.
7. 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 341:1237439. https://doi.org/10.1126/science.1237439.
8. Gonze D, Lahti L, Raes J, Faust K. 2017. Multi-stability and the origin of microbial community types. ISME J 11:2159–2166. https://doi.org/10.1038/ismej.2017.60.
9. Relman DA. 2012. The human microbiome: ecosystem resilience and health. Nutr Rev 70:S2–S9. https://doi.org/10.1111/j.1753-4887.2012.00489.x.
10. Levy R, Magis AT, Earls JC, Manor O, Wilmanski T, Lovejoy J, Gibbons SM, Omenn GS, Hood L, Price ND. 2020. Longitudinal analysis reveals transition barriers between dominant ecological states in the gut microbiome. Proc Natl Acad Sci U S A 117:13839–13845. https://doi.org/10.1073/pnas.1922498117.
11. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, Waller A, Mende DR, Kultima JR, Martin J, Kota K, Sunyaev SR, Weinstock GM, Bork P. 2013. Genomic variation landscape of the human gut microbiome. Nature 493:45–50. https://doi.org/10.1038/nature11711.
12. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature 464:908–912. https://doi.org/10.1038/nature08937.
13. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, Weinberger A, Fu J, Wijmenga C, Zhernakova A, Segal E. 2019. Structural variation in the gut microbiome associates with host health. Nature 568:43–48. https://doi.org/10.1038/s41586-019-1065-y.
14. Lynch JB, Gonzalez EL, Choy K, Faull KF, Jewell T, Arellano A, Liang J, Yu KB, Paramo J, Hsiao EY. 2022. Turicibacter modifies host bile acids and

lipids in a strain-specific manner. bioRxiv. https://doi.org/10.1101/2022.06.27.497673.
15. Benz F, Hall AR. 2022. Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids. bioRxiv. https://doi.org/10.1101/2022.07.06.498992.
16. Garud N, Good B, Hallatschek O, Pollard K. 2019. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol 17: e3000102. https://doi.org/10.1371/journal.pbio.3000102.
17. Zhao S, Lieberman TD, Poyet M, Kauffman KM, Gibbons SM, Groussin M, Xavier RJ, Alm EJ. 2019. Adaptive evolution within gut microbiomes of healthy people. Cell Host Microbe 25:656–667.e8. https://doi.org/10.1016/j.chom.2019.03.007.
18. Chen L, Wang D, Garmaeva S, Kurilshikov A, Vich Vila A, Gacesa R, Sinha T, Lifelines Cohort Study, Segal E, Weersma RK, Wijmenga C, Zhernakova A, Fu J. 2021. The long-term genetic stability and individual specificity of the human gut microbiome. Cell 184:2302–2315.e12. https://doi.org/10.1016/j.cell.2021.03.024.
19. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. 2020. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol 18: 491–506. https://doi.org/10.1038/s41579-020-0368-1.
20. Yassour M, Vatanen T, Siljander H, Hämäläinen A-M, Härkönen T, Ryhänen SJ, Franzosa EA, Vlamakis H, Huttenhower C, Gevers D, Lander ES, Knip M, DIABIMMUNE Study Group, Xavier RJ. 2016. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci Transl Med 8:343ra81. https://doi.org/10.1126/scitranslmed.aad0917.
21. Quince C, Nurk S, Raguideau S, James R, Soyer OS, Summers JK, Limasset A, Eren AM, Chikhi R, Darling AE. 2021. STRONG: metagenomics strain resolution on assembly graphs. Genome Biol 22:124. https://doi.org/10.1186/s13059-021-02419-7.
22. Verster AJ, Ross BD, Radey MC, Bao Y, Goodman AL, Mougous JD, Borenstein E. 2017. The landscape of type VI secretion across human gut microbiomes reveals its role in community composition. Cell Host Microbe 22:411–419.e4. https://doi.org/10.1016/j.chom.2017.08.010.
23. Roodgar M, Good BH, Garud NR, Martis S, Avula M, Zhou W, Lancaster SM, Lee H, Babveyh A, Nesamoney S, Pollard KS, Snyder MP. 2021. Longitudinal linked read sequencing reveals ecological and evolutionary responses of a human gut microbiome during antibiotic treatment. Genome Res 31: 1433–1446. https://doi.org/10.1101/gr.265058.120.
24. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. 2017. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Res 27:626–638. https://doi.org/10.1101/gr.216242.116.
25. Conwill A, Kuan AC, Damerla R, Poret AJ, Baker JS, Tripp AD, Alm EJ, Lieberman TD. 2022. Anatomy promotes neutral coexistence of strains in the human skin microbiome. Cell Host Microbe 30:171–182.e7. https://doi.org/10.1016/j.chom.2021.12.007.
26. Goyal A, Bittleson LS, Leventhal GE, Lu L, Cordero OX. 2022. Interactions between strains govern the eco-evolutionary dynamics of microbial communities. Elife 11:e74987. https://doi.org/10.7554/eLife.74987.
27. Russell SL, Cavanaugh CM. 2017. Intrahost genetic diversity of bacterial symbionts exhibits evidence of mixed infections and recombinant haplotypes. Mol Biol Evol 34:2747–2761. https://doi.org/10.1093/molbev/msx188.
28. Yaffe E, Relman DA. 2020. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. Nat Microbiol 5:343–353. https://doi.org/10.1038/s41564-019-0625-0.
29. Zheng W, Zhao S, Yin Y, Zhang H, Needham DM, Evans ED, Dai CL, Lu PJ, Alm EJ, Weitz DA. 2022. High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. Science 376: eabm1483. https://doi.org/10.1126/science.abm1483.
30. Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, Huerta-Cepas J, Nieuwdorp M, Salojärvi J, Voigt AY, Zeller G, Sunagawa S, de Vos WM, Bork P. 2016. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science 352:586–589. https://doi.org/10.1126/science.aad8852.
31. Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, Perrotta AR, Berdy B, Zhao S, Lieberman TD, Swanson PK, Smith M, Roesemann S, Alexander JE, Rich SA, Livny J, Vlamakis H, Clish C, Bullock K, Deik A, Scott J, Pierce KA, Xavier RJ, Alm EJ. 2019. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. Nat Med 25:1442–1452. https://doi.org/10.1038/s41591-019-0559-3.
32. Grilli J. 2020. Macroecological laws describe variation and diversity in microbial communities. Nat Commun 11:4743. https://doi.org/10.1038/s41467-020-18529-y.
33. Descheemaeker L, de Buyl S. 2020. Stochastic logistic models reproduce experimental time series of microbial communities. Elife 9:e55650. https://doi.org/10.7554/eLife.55650.
34. Shoemaker WR, Locey KJ, Lennon JT. 2017. A macroecological theory of microbial biodiversity. Nat Ecol Evol 1:107. https://doi.org/10.1038/s41559-017-0107.
35. Ji BW, Sheth RU, Dixit PD, Tchourine K, Vitkup D. 2020. Macroecological dynamics of gut microbiota. Nat Microbiol 5:768–775. https://doi.org/10.1038/s41564-020-0685-1.
36. Ho P-Y, Good BH, Huang K. 2022. Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas. Elife 11:e75168. https://doi.org/10.7554/eLife.75168.
37. Lim JJ, Diener C, Gibbons SM. 2022. Growth phase estimation for abundant bacterial populations sampled longitudinally from human stool metagenomes. bioRxiv. https://doi.org/10.1101/2022.04.23.489288.
38. Human Microbiome Project Consortium. 2012. A framework for human microbiome research. Nature 486:215–221. https://doi.org/10.1038/nature11209.
39. Fuller WA. 2009. Introduction to statistical time series. John Wiley & Sons, Chichester, United Kingdom.
40. Zaoli S, Grilli J. 2021. A macroecological description of alternative stable states reproduces intra- and inter-host variability of gut microbiome. Sci Adv 7:eabj2882. https://doi.org/10.1126/sciadv.abj2882.
41. Korpela K, Costea P, Coelho LP, Kandels-Lewis S, Willemsen G, Boomsma DI, Segata N, Bork P. 2018. Selective maternal seeding and environment shape the human gut microbiome. Genome Res 28:561–568. https://doi.org/10.1101/gr.233940.117.
42. Tropini C, Moss EL, Merrill BD, Ng KM, Higginbottom SK, Casavant EP, Gonzalez CG, Fremin B, Bouley DM, Elias JE, Bhatt AS, Huang KC, Sonnenburg JL. 2018. Transient osmotic perturbation causes long-term alteration to the gut microbiota. Cell 173:1742–1754.e17. https://doi.org/10.1016/j.cell.2018.05.008.
43. Good BH, Martis S, Hallatschek O. 2018. Adaptation limits ecological diversification and promotes ecological tinkering during the competition for substitutable resources. Proc Natl Acad Sci U S A 115:E10407–E10416. https://doi.org/10.1073/pnas.1807530115.
44. Tropini C, Earle KA, Huang KC, Sonnenburg JL. 2017. The gut microbiome: connecting spatial organization to function. Cell Host Microbe 21: 433–442. https://doi.org/10.1016/j.chom.2017.03.010.
45. Karita Y, Limmer DT, Hallatschek O. 2022. Scale-dependent tipping points of bacterial colonization resistance. Proc Natl Acad Sci U S A 119:e2115496119. https://doi.org/10.1073/pnas.2115496119.
46. Jin S, Wetzel D, Schirmer M. 2022. Deciphering mechanisms and implications of bacterial translocation in human health and disease. Curr Opin Microbiol 67:102147. https://doi.org/10.1016/j.mib.2022.102147.
47. Feng J, Qian Y, Zhou Z, Ertmer S, Vivas EI, Lan F, Hamilton JJ, Rey FE, Anantharaman K, Venturelli OS. 2022. Polysaccharide utilization loci in Bacteroides determine population fitness and community-level interactions. Cell Host Microbe 30:200–215.e12. https://doi.org/10.1016/j.chom.2021.12.006.
48. Lenski R. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J 11:2181–2194. https://doi.org/10.1038/ismej.2017.69.
49. Fridman Y, Wang Z, Maslov S, Goyal A. 2022. Fine-scale diversity of microbial communities due to satellite niches in boom and bust environments. PLoS Comput Biol 18:e1010244. https://doi.org/10.1371/journal.pcbi.1010244.
50. Tang L, Li Y, Deng X, Johnston RN, Liu G-R, Liu S-L. 2013. Defining natural species of bacteria: clear-cut genomic boundaries revealed by a turning point in nucleotide sequence divergence. BMC Genomics 14:489. https://doi.org/10.1186/1471-2164-14-489.
51. Tikhonov M. 2017. Theoretical microbial ecology without species. Phys Rev E 96:e032410. https://doi.org/10.1103/PhysRevE.96.032410.
52. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. 2016. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res 26:1612–1625. https://doi.org/10.1101/gr.201863.115.
53. García-López M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, Kyrpides NC, Hahnke RL, Göker M. 2019. Analysis of 1,000 type-strain genomes improves taxonomic classification of Bacteroidetes. Front Microbiol 10:2083. https://doi.org/10.3389/fmicb.2019.02083.
54. Seabold S, Perktold J. 2010. statsmodels: econometric and statistical modeling with python, p 92–96. *In* van der Walt S, Millman J (ed), Proceedings of the 9th Python in Science Conference.