Towards a potential landscape framework of microbiome dynamics

William K. Chang[1], Dave VanInsberghe[2], and Libusha Kelly[1]

[1]Systems and Computational Biology Department, Albert Einstein College of Medicine
[2]Department of Environmental and Civil Engineering, Massachusetts Institute of Technology

November 14, 2019

Abstract

Microbiome dynamics influence the health and functioning of human physiology and the environment. These dynamics are driven in part by interactions between large numbers of microbial taxa, making large-scale prediction and modeling a challenge. Here, we identify states and dynamical features relevant to macroscopic processes, such as infection in the human body and geochemical cycling in the oceans, by modeling the dynamics as stochastic motion on a potential energy-like landscape. We show that gut disease processes and marine geochemical events are associated with reproducible transitions between community states, defined as topological features of the landscape. We find a reproducible two-state succession during recovery from cholera in the gut microbiomes of multiple patients. Recurrence of the late disease state prolongs disease duration. We find evidence of dynamic stability in the gut microbiome of a human subject after experiencing diarrhea during travel, in contrast to residual instability in a second human subject after clinical recovery from *Salmonella* infection. Finally, we find the structure of marine *Prochlorococcus* communities in the western Atlantic and north Pacific oceans to smoothly vary with temperature and depth. However, annual water column cycling in the Atlantic drives periodic state transitions across depths. Our approach bridges the small-scale fluctuations in microbiome composition and large-scale changes in state and phenotype, improves analyses of how changes in community composition associate with phenotype without requiring experimental characterization of underlying mechanisms, and provides a novel assessment of microbiome stability and its relation to human and environmental health.

Importance

Time series of microbial communities are difficult to analyze due to the large number of interacting taxa. We developed a novel analysis based on topology to detect compositional states and state transitions in microbial time series. Our method generalizes across biological systems and can identify gut microbiome dynamics associated with recovery from disease in multiple patients on the order of weeks, and marine bacterial dynamics driven by geochemical cycling on the order of years. We furthermore propose a novel definition of ecological stability that distinguishes between complete and incomplete recovery from infection in human gut microbiomes. Our method requires minimal assumptions regarding biological mechanisms. Overall, our analysis complements current methods for identifying key ecological processes in microbial communities, and suggests further developments in modeling that may improve prediction of microbial dynamics.

Introduction

Complex microbial ecosystems ('microbiomes') inhabit a diversity of environments in the biosphere, including the global ocean [42], soil [12], and the human gut [43]. Large-scale alterations in the composition of microbiomes is often associated, whether as driver or consequence, with environmental processes such as seasonal geological cycling and nutrient fluctuations [14]; physiological processes

such as menstrual cycles [15]; and clinical phenotypes such as irritable bowel syndrome [3]. Analysis and prediction of the large-scale dynamics of microbiome composition is thus a pressing issue in multiple fields of study.

As with many biological systems, understanding of the dynamics of microbiomes is complicated by their high dimensionality. Numerous variables define the state of a microbiome; these include frequencies of microbial taxa and their genetic alleles, which are decoupled due to genomic plasticity and horizontal gene transfer [31, 32], and environmental conditions such as temperature, pH, and biochemical concentrations. A microbiome thus has a vast number of potential configurations in which it may, in principle, fluctuate on a short time scale. By contrast, systemic phenotypes, such as human gut infections or aquatic algal blooms, persist for much longer than bacterial generation time, and community compositions may be diverse within a phenotype [14]. Furthermore, due to the diverse biology of microbiomes across habitats, it may be desirable to have a quantitative framework that can be generalized across biological systems.

One approach to analyzing microbiome dynamics has been to infer the network of underlying pairwise interactions between taxa by calculating the inverse covariance matrix from time series data, often as a basis for modeling population dynamics using Lotka-Volterra equations [13, 25, 41]. Such approaches are useful for predicting fine-grained taxon-taxon interactions of importance, and are challenged by the compositional nature of microbiome data [39] and possible role of higher-order interactions [4]. A complementary coarse-grained approach is to cluster samples according to compositional similarity, and conceptualize dynamics as stochastic transitions between clusters [2, 9]. Such approaches can be used to identify large-scale shifts in compositional state, with the implicit assumption that each temporal sample can be assigned to one of a finite number of discrete categories.

We propose to supplement these methods with a *potential landscape* approach. Potential landscapes provide a framework for modeling the dynamics of high-dimensional, complex systems such as microbiomes by representing the configurations of a dynamical system—here, the possible compositions of a microbiome—as coordinates in phase space, where similar configurations are located close together. The system dynamics are considered as stochastic motion on the resultant manifold, with topological features corresponding to the probable configurations of the system and trajectories between them. Features of the potential landscape, such as valleys and peaks, represent more and less probable compositional states, respectively, and are related to notions of attractors and basins of attraction in dynamical systems theory. This approach predicts that, over time, the system evolves along the contours of the landscape towards a local minimum of the potential; thus, the shape of the landscape in principle predicts the dynamics. Thus, the landscape encodes the underlying interactions between components, in our case microbial taxa, without explicit assumptions regarding underlying biological mechanisms (Fig. 1A). In biology, attractors and basins of attraction have been found in theoretical and experimental studies to correspond to states of population survival and extinction [6, 7, 35]; cell phenotypes in differentiating stem cells [44, 46] and transformed cancer cells [23, 27]; and probable states of brain activity [19]. These results show that topological features of the potential landscape can be thought of as metastable states associated with phenotypes of biological and clinical relevance as well as the dynamics of phenotypic transitions, and that revealing the potential landscape may have implications for modeling and predicting the dynamics of complex biological systems. Similarly to clustering, the composition can be approximated by the metastable state or states to which it belongs at a given time, and the trajectory of the system in phase space over time can be approximated by a succession of such states. However, it is important to note that this definition of system states derives from an underlying continuous potential landscape, and thus differs from clustering methods.

To characterize features of the microbial potential landscape, we used topological data analysis (TDA), specifically the Mapper algorithm [34, 40] to infer the topological features of the potential landscapes for three published microbial time series data sets, two human gut microbiomes—one of stool samples collected from seven cholera patients from disease through recovery [22], one from two mostly healthy adult males [8]—and one of marine *Prochlorococcus* communities spanning multiple depths collected from one site in the Atlantic Ocean (BATS) and one in the Pacific (HOT) [28]. We selected these data sets in part to test our method by recapitulating biology known from the original studies, and in part to discover novel features not addressed by prior methods. Briefly, Mapper represents the underlying distribution of data in a metric space as an undirected graph, where each vertex comprises a non-exclusive subset of data points spanning a patch of phase space. An edge is drawn between each two vertices that share at least one data point (Fig. 1A), representing

connectivity between patches. We complement Mapper with a novel graph-theoretical analysis to estimate the value of the potential over each patch of phase space represented by a vertex, determine local minima, and define metastable community states (Fig. 1B). In both human gut and marine systems, we find that significant physiological and environmental events, including recovery from infection and geochemical cycling, correspond to recurrent successions of state transitions. We show that these successions are an informative coarse-grained view of microbiome dynamics, with implications for the assessment of ecological resilience.

Results

Dynamics of human gut microbiome recovery from cholera infection

We found the cholera phase space to be partitioned by clinical phenotype, i.e. diarrhea or recovery (Fig. 2A). The original study [22] recognized phases of progression according to equal-time divisions of the diarrhea and recovery periods, respectively, of each patient. Our identification of disease substates,in contrast, is based on community composition and integrated across data from all patients. We found the diarrhea region was further subdivided into two states, 2 and 7 (Fig. 2B). Patients C, E, and G occupied state 7 for prolonged durations immediately before clinical recovery; patients A, B, and F stably occupied state 7 for approximately 20 hours, but switched to other states for the last few time points before clinical recovery (Fig. 2C). In the case of patient A, the final few time points were associated with state 5, which represented an intermediate region of the phase space between the diarrhea- and recovery-associated neighborhoods. These results suggest that state 2 constituted a universal 'early' diarrhea state, and state 7 a universal 'late' diarrhea state, with distinct community compositions. The original study noted taxa which consistently changed in abundance between the start and end of the diarrhea phase, for example *Streptococcus* and *Fusobacterium* [22], here we show that these compositional shifts are observable on the whole-community scale.

Generally, patients occupied state 7 for longer than they did state 2, suggesting that the stability of the late state in a given patient influences disease duration. To quantify stability, we calculated a temporal correlation function for each state-patient pair during the diarrhea phase. Monotonically decreasing correlation functions indicate metastability; slopes become more negative with decreasing stability. While this analysis revealed that all patients transiently occupied state 2, with greatest persistence in patient C, patients A, C, and E had non-monotonic correlation functions for state 7, coinciding with prolonged times to recovery compared to the rest of the cohort, with patients B and F exhibiting the expected monotonic decrease (Fig. 2D). This indicated that patients A, C, and E repeatedly entered and exited state 7, suggesting that prolonged diarrhea in these three patients may have been additionally influenced by the instability or inaccessibility of alternative, healthy states, and that (re-)assembly of the healthy microbial community constitutes a non-trivial step in recovery.

Dynamics of two healthy adult microbiomes with transient diarrhea

In contrast to the cholera data set, the two healthy adult gut microbiome time series from David *et al.* [8] were separated by subject (Fig. 3A). Despite being clinically healthy for most of the observation period, both subjects' microbiomes experienced perturbations: subject A traveled from his residence in the United States to southeast Asia, twice experiencing traveller's diarrhea; and subject B, also based in the US, suffered an acute infection by *Salmonella*. Previous studies [8, 18] noted that, while the microbiome of A returned to its original state after travel, recovery from *Salmonella* left the microbiome of B in an alternative state. Confirming this, we found that subject A occupied the same regions of phase space before and after travel, while subject B occupied disjoint regions before and after infection. We further found that the post-*Salmonella* samples of subject B distributed over several connected components, showing that the gut microbiome of subject B remained in flux across several distinct compositional substates even after being clinically marked as having recovered (Fig 3B).

The large connected components representing the pre- and post-travel healthy samples of subject A and the pre-*Salmonella* healthy samples of subject B were each divided into several states (Supplementary Information Fig. 1), suggesting that the clinical 'healthy' phenotype of an individual is a probability over multiple compositionally distinct states. The existence of states in

microbiome phase space proposes a novel metric for microbiome resilience: comparing the distribution of samples across states between time windows. Subject A occupied states with identical probability before and after travel, exhibiting resilience; in contrast, subject B post-infection did not restore the pre-infection probability across states, despite some samples sharing states with pre-infection healthy samples (Fig. 4A). Thus, the restoration of the microbial community to a 'healthy' state cannot be confirmed with a single time point.

Temporal correlation functions further showed that subject A, as well as subject B before infection, repeatedly visited the same set of states; in contrast, subject B after infection transiently occupied several states without repetition (Fig. 4B). This shows that not only did the microbiome of subject B enter an alternative state, or probability across states, post-infection, but that this alternative state was not fully stabilized. It is possible that the pre-infection probability across states was restored in subject B after the end of the observational period.

Recurrent seasonal dynamics of *Prochlorococcus* communities in the Pacific and Atlantic

Compared to the phase spaces of human gut microbiomes, which may be discretized by individual or phenotype, the *Prochlorococcus* phase space was organized by gradients of depth (Fig. 5A) and temperature (Supporting Fig. 4), indicating that, in these environments, small changes to environmental conditions result in small changes to community structure. The phase space possessed multiple states (Fig 5B), with state 4 largely representing shallow fractions of the water column $\leq$ 100m; states 2, 3, and 6 deeper fractions; and state 1 intermediate depths. State 5 represented an infrequently-occupied region sampled only by the 140m fraction at BATS on January 27, 2004, and by the 125m fraction at HOT on January 31, 2008. As such, state 5 possibly constitutes an alternative state for deep water fractions in mid-winter. Communities differing in depth rarely shared compositions, and transitioned between states, in many cases periodically across calendar years (Fig. 5C), showing that some communities experienced abrupt periodic shifts in environmental conditions due to geochemical events.

Despite the graduated variation of composition with depth and temperature, the range of compositional dissimilarity across the range of environmental conditions is sufficient to constrain given depth fractions to a neighborhood of phase space, such that shallow- and deep-fraction *Prochlorococcus* communities rarely occupy the same compositional states over time (Fig. 5C). However, it is known that the BATS water column undergoes an annual late winter upwelling [28], intermixing communities that otherwise inhabit different depths, and homogenizing environmental conditions across depths. We predicted that mixing would drive communities at all depths at BATS to converge on a common state, while no convergence would be observed at HOT. Accordingly, we observed a transition to state 1 by all depths at BATS in January of each year. After June, depths 1-20m and 120-200m relax toward states characteristic of shallow and deep depth fractions, respectively, while state 1 persists longer in intermediate depths 40-100m. By contrast, no such upwelling occurs at HOT, and the probability of a given depth fraction occupying any state remains uniform over the calendar year; the distribution is especially stationary for shallow depths (Fig. 5C). This periodicity was also evident in periodic correlation functions for BATS, and non-periodic for HOT (Fig. 5D).

Robustness of potential estimation

Given that the data sets analyzed here are among the largest longitudinal microbiome data sets currently available, we asked whether the biological hypotheses could have been obtained from sparser data sets. We focused on our finding that microbiome phase spaces are structured by latent variables representing host phenotypes or environmental conditions, and examined whether this structuring was robust to data rarefaction. We found that the partitioning of the phase space by clinical phenotype in the case of the cholera patients, by subject in the case of the two healthy adult humans, and the gradation by depth in the case of *Prochlorococcus* communities, are robust to all rarefaction tests performed. In the case of cholera patients, nodes remained divided into those representing mostly samples from the diarrhea phase and those representing the recovery phase, with edges being more dense between nodes of the same phenotype than those of different phenotypes (Supporting Information Fig. 3). In the case of the two healthy adult humans, nodes were consistently dominated by samples from one subject, with edges being

more dense between nodes representing the same subject than those representing different subjects (Supporting Information Fig. 4). For the *Prochlorococcus* data set, nodes aggregating samples from similar depth fractions were more densely connected than those representing disparate depths (Supporting Information Fig. 5).

Discussion

We identified unrecognized dynamics governing large-scale phenotypes in microbial time series data by using TDA to infer the shape of a potential landscape from 16S and ITS ribosomal RNA time series data. Our results reveal the role of latent physiological and environmental variables [29], such as disease phenotype and phase of geochemical cycles, in organizing microbiomes over time. We observed common dynamics across instances of ecological processes in the two gut and one environmental timeseries datasets we studied. Using our approach, one can thus begin to infer general mechanisms that determine large scale phenotypes of clinical and environmental importance. The elements of our method—the definition of a metric phase space using the square root of the Jensen-Shannon divergence, the representation of the phase space using TDA, and the characterization of topological features using the adapted kNN density estimator and shortest graph distance searches—are specifically advantageous for analyzing high-dimensional compositional data. Compared to representational methods such as PCA, our method benefits from using all distance information; and compared to clustering techniques, our method does not require specifying the number of states, such as required in k-means.

While subjects in both human gut data sets experienced transient infection by bacterial pathogens, the large-scale dynamics differed between the two groups. We found that multiple cholera patients followed a trajectory of early- to late-stage disease states. In contrast, the two healthy subjects from the year-long data set experienced apparently random jumps between states during *Salmonella* infection and traveler's diarrhea, respectively. This discordance between the two human gut microbiome datasets suggests that microbial infections can potentially be classified into 'ordered' and 'disordered' types. Ordered infections are characterized by a reproducible trajectory through phase space, while disordered infections are characterized by unpredictable progression through phase space. The latter case represents a version of the 'Anna Karenina principle,' meaning individual microbiomes are more dissimilar during a particular perturbation than during health [45], while the former represents an inversion of the principle. Scale is likely important in this distinction: independent of the deterministic or stochastic nature of the perturbation induced by an infection, if its magnitude is smaller than 'baseline' fluctuations of the healthy microbiome, variations between individuals will remain the dominant variable in organizing the phase space. If the magnitude of the perturbation is larger, it may overwhelm individual variability and cause the phase space to instead appear organized by phenotype. Thus, data on the variability of healthy microbiomes over time between and within individuals will be crucial to characterizing the impact of a given disease on the microbiome.

Our analysis of the David *et al.* data set shows that the microbiome of a healthy individual transitions between states over time. While key dominant taxa may persist, no single large-scale compositional state defines healthy physiology. However, an individual microbiome may occupy states with the same probability during two separate 'healthy' time windows. Integrating the information over time for each of the healthy periods, the physiological phenotype can be inferred to be stable despite the system state being dynamic. Put differently, if one interprets states as microstates of the microbiome composition, a systemic clinical or environmental phenotype could then be regarded as a *macrostate*, and a resilient 'healthy' microbiome will remain in a stable macrostate over time.

This notion of resilience as identical probability across states before and after a perturbation can be generalized to a notion of dynamic stability, defined as stationary probability across states over time. Dynamically stable microbiomes do not necessarily stabilize within a single state, but revisit a given set of states with fixed probability. Our temporal correlation analysis shows that dynamically stable microbiomes, such as subject A and subject B pre-infection from the study in [8], are characterized by non-monotonic temporal correlation functions, indicating the microbiome revisits the same states over time. In contrast, unstable microbiomes, such as subject B post-infection, exhibit monotonically decaying correlation functions, indicating the microbiome transiently occupies compositional states without recurrence. Dynamical instability can persist after infection even in the microbiome of an individual clinically marked as having recovered from

infection, as in the case of subject B, revealing additional nuances to the association between stability and health in human microbiomes. The ability to assess resilience from data in the absence of detailed knowledge of the underlying network of microbe-microbe interactions complements model-based methods that analytically solve for fixed points and linear stability [5].

For the two human gut microbiome data sets, we observe some of the same phenomena as the original studies: for the seven cholera patients, certain taxa were differentially abundant throughout the progression of disease [22]; and for subject B of the two healthy males, that the pre-*Salmonella* microbiome composition was not recovered by the end of the experiment [8]. In the first case, we remark that differential abundance of individual taxa does not necessarily imply the existence of large-scale compositional states consistent across patients and disease phases, such as we describe here. In the second case, we additionally found multiple states in the pre- and post-perturbation healthy phases of both subjects, and showed that restoration of a healthy and resilient microbiome is associated with the recovery not of a specific composition but of a distribution across compositional states.

We point out several caveats regarding our method. First, though we defined the phase space using the Jensen-Shannon distance, other metrics may be used, and the results of analysis using different metrics for the same data should be compared in future applications. Second, due to the lack of an established protocol for selecting Mapper hyperparameters, we used a heuristic method to choose their values for our analyses. A more rigorous optimization method is desirable, especially one developed against synthetic data from *de novo* simulations where the 'ground truth' of the parameters, and thus the shape of the potential landscape, are known *a priori*. Third, we use Mapper to create a representation of the potential landscape, but the landscape and question of whether it is effective to model microbiome dynamics in a given case using a potential landscape are independent of Mapper and TDA, and other methods may be used. Fourth, we assume the data accurately represent the compositions of the sampled communities, when in fact challenges exist with translating sequencing data into compositions [16, 17]; addressing these challenges is outside the scope of this manuscript.

In real ecosystems such as those under study, several factors may complicate the basic prediction of the potential landscape that real ecosystems evolve toward configurations of lowest potential, and thus limit predictive power. First, real systems are open to their environment and subject to external perturbation; the dynamics of an ecosystem experiencing strong driving forces may deviate from that predicted by the potential landscape. In addition, strong stochastic fluctuations in microbial populations may weaken the predictive power of the potential landscape; however, in this case, the potential landscape may still form an informative 'deterministic skeleton' of the dynamics [1]. Third, high dimensionality may also increase the number and complexity of paths by which the system evolves toward lower potential. Finally, the time scales of sampling may differ from those that are predictable by the potential landscape; for example, the potential landscape may well predict the dynamics of gut microbiome relaxation after a meal on the time scale of hours, but this may not be captured by daily sampling. Nolting [30] and Abbott [1] discuss some of these factors in detail. As above, analysis of synthetic data generated by theoretical population dynamics models may help elucidate the limits of potential landscape inference and prediction.

In addition to offering a novel quantitative description of microbiome states and dynamics, we hope our analysis will, in time, facilitate predictive modeling of the dynamics and forecasting of major state transitions in the microbiome. As an example, our approach to identifying states from microbial time series can be used to infer state transition probabilities under different conditions, and thus can serve as a basis for fitting the parameters of Markov chain models [9, 11]. Alternatively, the theory of critical transition forecasting [6, 7, 26, 37, 36] is closely linked to the concept of the potential: as perturbations destabilize a system, it ascends the potential gradient and eventually reaches a tipping point from where it can rapidly enter into an alternative stable state. Topological analyses, in turn, may enable characterization of the system state and potential based on past observations, and real-time estimation of its stability and state transition probability. Both of these approaches allow modeling and prediction of major dynamical events without detailed knowledge of underlying mechanisms, and may prove pivotal to understanding complex, data-rich biological systems not limited to microbiomes, but also including, for instance, gene regulatory networks and animal ecosystems.

Methods

Human gut microbiome data and preprocessing

The publicly available data that we re-analyzed here were generated by David *et al* [8] accessible on the European Nucleotide Archive (ENA) under the accession number ERP006059, and by Hsiao *et al* [22] on the NCBI Short Read Archive (SRA) under the accession number PRJEB6358. The downloaded reads were trimmed with V-xtractor version 2.1 [21] to ensure the amplicon sequences could be aligned across consistent fractions of the 16S rRNA variable regions. Trimmed reads were then clustered into OTUs at a Levenshtein distance of two using CrunchClust version 43 [20] and classified up to the family level using MOTHUR version 1.36.1 [38] and Silva release 128 [33] reference sequences.

Prochlorococcus data

Data from Malstrom *et al* [28] was obtained from the Biological and Chemical Oceanography Data Management Office (`https://www.bco-dmo.org`), accession number 3381.

Mapper

Conceptually, the Mapper algorithm accepts as input a matrix of distances or dissimilarities between data, and aims to represent the shape of the distribution of data points in high-dimensional phase space as an undirected graph. In this graph, vertices represent neighborhoods of phase space spanned by subsets of adjacent data points, and edges represent connectivity between neighborhoods. In brief, it does this by dividing the data into overlapping subsets that are similar according to the output of at least one filter function that assigns a scalar value to each data point, performing local clustering on each subset, and representing the result as an undirected graph, where each vertex represents a local cluster of data points, and edges between vertices represent at least one shared data point between clusters.

Distance matrix

We interpreted microbiome relative abundances to be probability distributions, and thus used the square root of the Jensen-Shannon divergence as a metric [24]. However, it is important to note that any other metric can be used in place of the Jensen-Shannon distance, such as an Euclidean calculated from centered [25] or isometric [39] log-transformed relative abundances.

Filter functions and binning

For the filter functions used by Mapper to bin data points, we performed principal coordinate analysis (PCoA, also known as classical multidimensional scaling) in two dimensions on the pairwise distance matrix, and used the ranked values of principal coordinates (PCo) 1 and 2 as the first and second filter values for Mapper, following Rizvi *et al.* [34]. PCo ranks are an appropriate filter for our purposes, as it assigns similar filter values to points that are relatively close together in the original phase space. We wish to note that while PCoA leads to loss of information, the following local clustering step is performed using subsets of distances from the original distance matrix, and is thus not affected. The data points were then binned by overlapping intervals of the two ranked principal coordinates. For hyperparameters specifying these bins and their overlaps, see Table 1.

Local clustering

The algorithm first performs hierarchical clustering from all pairwise distances between data points within a bin of filter values. Then, it creates a histogram of branch lengths using a predefined number of bins, and uses the first empty bin in the histogram as a cutoff value, separating the hierarchical tree into single-linkage clusters. The algorithm thus finds a separation of length scales within each neighborhood of phase space represented by a bin of the filter values. We used the default number of histogram bins, 10, for each data set (Table 1).

Creating the undirected Mapper graph

The final output is produced by representing each local cluster of data points as a vertex, and drawing an edge between each pair of vertices that share at least one data point. When plotting, the size of each vertex represents the number of data points therein.

Selection of hyperparameters

The Mapper algorithm is relatively new, and there are currently no standard protocols to optimize the values of the hyperparameters. For our purposes, it was important that the algorithm achieved a sufficiently high resolution in partitioning data, but also adequately represented connections between regions of phase space. We thus used the following heuristic to set the number of intervals and percent overlap for each data set.

1. The largest vertex in the resultant Mapper graph should represent no more than $\approx 10\%$ of the total number of data points in the set;
2. the number of connected components representing only one data point should be minimized.

We acknowledge that a heuristic determination of appropriate hyperparameter values leaves much to be desired; as such, we recommend future in-depth theoretical explorations of how the Mapper output depends on the choice of hyperparameters.

Potential estimation

We estimated the potential for each vertex by calculating the k-nearest neighbors (kNN) density [10] for each constituent data point i:

$$\mathrm{kNN}(i,k) = \frac{\sum_j^k d_{ij}}{k} \tag{1}$$

where d_{ij} is the distance between points i and j, choosing k equal to 10% of the number of samples in each data set, rounded to the nearest integer. kNN varies inversely with density, making it a proxy for the potential. For a vertex V representing n points, we define its potential as

$$U(V) = \frac{\sum_{i\in V}^n \mathrm{kNN}(i,k)}{n^2} \tag{2}$$

The n^2 term in the denominator compensates for the differing sizes of vertices.

State assignment

We then defined states as topological features of the landscape surrounding local minima of U. We designated each vertex with lower U than its neighbors to be a local minimum of the potential. Connected vertices tied for minimum U were each assigned to be a local minimum. To approximate a gradient, we converted the undirected Mapper graph to a directed graph, with each edge pointing from the the vertex with greater U to the one with lower U. For each non-minimum vertex, we found the graph distance d_g to each local minimum constrained by edge direction. We defined the *state* B_x of a minimum V_x as the set of vertices V with uniquely shortest graph distance to V_x:

$$V \in B_x \text{ if } d_g(V,V_x) < d_g(V,V_y) \tag{3}$$

for all $x \neq y$ and $V_y \in M$, where M is the set of all local minima (Fig 1B). Vertices equidistant to multiple minima were defined to be unstable regions unassigned to any state. Multiple connected minima were defined as belonging to the same state. Notably, one data point may be associated with multiple vertices and states, or an unstable region and at least one state: we interpreted this to mean that the point is near a saddle point separating states, and as the 'true' coordinates of the saddle point are unknown, the data point is assigned to *all* such states and/or an unstable region with uniform weight.

Calculating the temporal correlation function

Given that a system occupied state B_x at time t, we defined the temporal correlation to be the expectation that it will still (or again) occupy state B_x at time $t + \tau$:

$$f_x(t) = \begin{cases} 1 & \text{if system is associated with state } B_x \text{ at time } t \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

$$\text{corr}_x(\tau) = \langle f_x(t + \tau) \rangle \tag{5}$$

We calculated the correlation function for each state x visited by a subject during a characteristic period and for all sampled intervals of length τ, where $f_x(t) = 1$. For the cholera data set, we calculated correlation functions for each state visited by each subject over the disease period. For the data set of two healthy adult males, we calculated correlation functions for each state visited by each subject in each healthy period, either before or after infection. For the *Prochlorococcus* data set, we calculated correlation functions for each state at each depth fraction at either site. Where a data point is associated with multiple states, we weigh the association with each state as $f'_x(t) = \frac{1}{p} f_x(t)$, with p the total number of unique states associated with the system at time t, with the unassigned/unstable state regarded as a single distinct state.

Rarefaction test

We created random subsets of each data set representing 90%, 50%, and 10% of the original data points, repeating 10 times for each data set and downsampling ratio. We then created Mapper graphs representing the rarefied data using the same hyperparameters as for each of the full data sets. We colored the vertices to indicate the same features as for the full data sets: for the cholera data set, by fraction of samples belonging to the diarrhea or recovery phase; for the two healthy adult gut microbiomes data set, by fraction of samples obtained from each subject; and for the *Prochlorococcus* data set, by the mean depth from which samples originated. We ordered the vertices by feature value and used a circularized linear layout algorithm, such that vertices with similar feature values are adjacent. Finally, we used shading to display edge densities.

Software and data

The main repository for the study can be found on GitHub, at `http://github.com/kellylab/microbial-landscapes`.

An open-source implementation of Mapper in R, `TDAmapper`, was used for the main analysis and can be found at `http://github.com/wkc1986/TDAmapper`. This package was forked from the original implemented by Daniel Müllner which is maintained by Paul T. Pearson and can be found at `https://github.com/paultpearson/TDAmapper`.

Funding

L.K. is supported in part by a Peer Reviewed Cancer Research Program Career Development Award from the United States Department of Defense (CA171019).

Author's contributions

W.K.C. designed and performed the analysis. D.V. processed and performed OTU calling on the data from Hsiao *et al.*[22] and David *et al.*[8]. W.K.C., D.V., and L.K. wrote the manuscript.

References

[1] Karen C. Abbott and Ben C. Nolting. "Alternative (un)stable states in a stochastic predator–prey model". In: *Ecological Complexity* 32 (Dec. 2017), pp. 181–195. ISSN: 1476945X. DOI: 10.1016/j.ecocom.2016.11.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S1476945X16301039 (visited on 03/23/2018).

[2] J. Paul Brooks et al. "Changes in vaginal community state types reflect major shifts in the microbiome". In: *Microbial Ecology in Health and Disease* 28.1 (Jan. 1, 2017), p. 1303265. ISSN: null. DOI: 10.1080/16512235.2017.1303265. URL: https://doi.org/10.1080/16512235.2017.1303265 (visited on 08/05/2019).

[3] C. Casén et al. "Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD". In: *Alimentary Pharmacology & Therapeutics* 42.1 (July 2015), pp. 71–83. ISSN: 1365-2036. DOI: 10.1111/apt.13236.

[4] Hasan Celiker and Jeff Gore. "Clustering in community structure across replicate ecosystems following a long-term bacterial evolution experiment". In: *Nature Communications* 5 (Aug. 8, 2014). ISSN: 2041-1723. DOI: 10.1038/ncomms5643. URL: http://www.nature.com/doifinder/10.1038/ncomms5643 (visited on 12/18/2014).

[5] Katharine Z. Coyte, Jonas Schluter, and Kevin R. Foster. "The ecology of the microbiome: Networks, competition, and stability". In: *Science* 350.6261 (Nov. 6, 2015), pp. 663–666. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aad2602. URL: http://www.sciencemag.org/content/350/6261/663 (visited on 11/07/2015).

[6] L. Dai et al. "Generic Indicators for Loss of Resilience Before a Tipping Point Leading to Population Collapse". In: *Science* 336.6085 (June 1, 2012), pp. 1175–1177. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1219805. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1219805 (visited on 09/12/2014).

[7] Vasilis Dakos and Jordi Bascompte. "Critical slowing down as early warning for the onset of collapse in mutualistic communities". In: *Proceedings of the National Academy of Sciences* 111.49 (Dec. 9, 2014), pp. 17546–17551. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1406326111. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1406326111 (visited on 11/08/2016).

[8] Lawrence A. David et al. "Host lifestyle affects human microbiota on daily timescales". In: *Genome Biology* 15 (2014), R89. ISSN: 1474-760X. DOI: 10.1186/gb-2014-15-7-r89. URL: http://dx.doi.org/10.1186/gb-2014-15-7-r89 (visited on 08/12/2016).

[9] Daniel B. DiGiulio et al. "Temporal and spatial variation of the human microbiota during pregnancy". In: *Proceedings of the National Academy of Sciences* 112.35 (Sept. 1, 2015), pp. 11060–11065. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1502875112. URL: https://www.pnas.org/content/112/35/11060 (visited on 01/03/2019).

[10] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern classification*. Google-Books-ID: YoxQAAAAMAAJ. Wiley, 2001. 688 pp. ISBN: 978-0-471-05669-0.

[11] Mathieu Faure and Sebastian J. Schreiber. "Quasi-stationary distributions for randomly perturbed dynamical systems". In: *The Annals of Applied Probability* 24.2 (Apr. 2014), pp. 553–598. ISSN: 1050-5164. DOI: 10.1214/13-AAP923. URL: http://projecteuclid.org/euclid.aoap/1394465365 (visited on 10/31/2018).

[12] N. Fierer and R. B. Jackson. "The diversity and biogeography of soil bacterial communities". In: *Proceedings of the National Academy of Sciences* 103.3 (Jan. 17, 2006), pp. 626–631. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0507535103. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0507535103 (visited on 08/05/2019).

[13] Jonathan Friedman and Eric J. Alm. "Inferring Correlation Networks from Genomic Survey Data". In: *PLOS Computational Biology* 8.9 (Sept. 20, 2012), e1002687. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002687. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002687 (visited on 09/04/2018).

[14] Jed A. Fuhrman, Jacob A. Cram, and David M. Needham. "Marine microbial community dynamics and their ecological interpretation". In: *Nature Reviews Microbiology* 13.3 (Mar. 2015), pp. 133–146. ISSN: 1740-1526. DOI: 10.1038/nrmicro3417. URL: http://www.nature.com/nrmicro/journal/v13/n3/abs/nrmicro3417.html (visited on 03/03/2015).

[15] Pawel Gajer et al. "Temporal Dynamics of the Human Vaginal Microbiota". In: *Science Translational Medicine* 4.132 (May 2, 2012), 132ra52–132ra52. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.3003605. URL: http://stm.sciencemag.org/content/4/132/132ra52 (visited on 02/04/2016).

[16] Gregory Brian Gloor et al. "Compositional uncertainty should not be ignored in high-throughput sequencing data analysis". In: *Austrian Journal of Statistics* 45.4 (July 28, 2016), pp. 73–87. ISSN: 1026-597X. DOI: 10.17713/ajs.v45i4.122. URL: https://www.ajs.or.at/index.php/ajs/article/view/vol45-4-5 (visited on 07/11/2019).

[17] Gregory B. Gloor et al. "Microbiome Datasets Are Compositional: And This Is Not Optional". In: *Frontiers in Microbiology* 8 (2017). ISSN: 1664-302X. DOI: 10.3389/fmicb.2017.02224. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02224/full?report=reader (visited on 07/11/2019).

[18] Didier Gonze et al. "Microbial communities as dynamical systems". In: *Current Opinion in Microbiology* 44 (Aug. 1, 2018), pp. 41–49. ISSN: 1369-5274. DOI: 10.1016/j.mib.2018.07.004. URL: http://www.sciencedirect.com/science/article/pii/S1369527418300092 (visited on 07/24/2018).

[19] Shi Gu et al. "The Energy Landscape of Neurophysiological Activity Implicit in Brain Network Structure". In: *Scientific Reports* 8.1 (Feb. 6, 2018), p. 2507. ISSN: 2045-2322. DOI: 10.1038/s41598-018-20123-8. URL: https://www.nature.com/articles/s41598-018-20123-8 (visited on 11/19/2018).

[20] Martin Hartmann et al. "Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests". In: *The ISME Journal* 6.12 (Dec. 2012), pp. 2199–2218. ISSN: 1751-7370. DOI: 10.1038/ismej.2012.84. URL: https://www.nature.com/articles/ismej201284 (visited on 08/08/2019).

[21] Martin Hartmann et al. "V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences". In: *Journal of Microbiological Methods* 83.2 (Nov. 2010), pp. 250–253. ISSN: 1872-8359. DOI: 10.1016/j.mimet.2010.08.008.

[22] Ansel Hsiao et al. "Members of the human gut microbiota involved in recovery from Vibrio cholerae infection". In: *Nature* 515.7527 (Nov. 20, 2014), pp. 423–426. ISSN: 0028-0836. DOI: 10.1038/nature13738. URL: http://www.nature.com/nature/journal/v515/n7527/full/nature13738.html (visited on 02/11/2016).

[23] Sui Huang, Ingemar Ernberg, and Stuart Kauffman. "Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective". In: *Seminars in cell & developmental biology* 20.7 (Sept. 2009), pp. 869–876. ISSN: 1084-9521. DOI: 10.1016/j.semcdb.2009.07.003. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2754594/.

[24] Omry Koren et al. "A Guide to Enterotypes across the Human Body: Meta-Analysis of Microbial Community Structures in Human Microbiome Datasets". In: *PLOS Computational Biology* 9.1 (Jan. 10, 2013), e1002863. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002863. URL: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002863 (visited on 04/27/2017).

[25] Zachary D. Kurtz et al. "Sparse and Compositionally Robust Inference of Microbial Ecological Networks". In: *PLOS Computational Biology* 11.5 (May 7, 2015), e1004226. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1004226. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004226 (visited on 04/30/2019).

[26] Ingrid A. van de Leemput et al. "Critical slowing down as early warning for the onset and termination of depression". In: *Proceedings of the National Academy of Sciences* 111.1 (Jan. 7, 2014), pp. 87–92. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1312114110. URL: http://www.pnas.org/content/111/1/87 (visited on 01/29/2016).

[27] Qin Li et al. "Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.10 (Mar. 8, 2016), pp. 2672–2677. ISSN: 1091-6490. DOI: 10.1073/pnas.1519210113.

[28] Rex R. Malmstrom et al. "Temporal dynamics of Prochlorococcus ecotypes in the Atlantic and Pacific oceans". In: *The ISME Journal* 4.10 (Oct. 2010), pp. 1252–1264. ISSN: 1751-7362. DOI: 10.1038/ismej.2010.60. URL: http://www.nature.com/ismej/journal/v4/n10/full/ismej201060a.html (visited on 07/13/2016).

[29] Lan Huong Nguyen and Susan Holmes. "Bayesian Unidimensional Scaling for visualizing uncertainty in high dimensional datasets with latent ordering of observations". In: *BMC Bioinformatics* 18.10 (Sept. 13, 2017), p. 394. ISSN: 1471-2105. DOI: 10.1186/s12859-017-1790-x. URL: https://doi.org/10.1186/s12859-017-1790-x (visited on 01/03/2019).

[30] Ben C. Nolting and Karen C. Abbott. "Balls, cups, and quasi-potentials: quantifying stability in stochastic systems". In: *Ecology* 97.4 (Apr. 1, 2016), pp. 850–864. ISSN: 1939-9170. DOI: 10.1890/15-1047.1. URL: http://onlinelibrary.wiley.com/doi/10.1890/15-1047.1/abstract (visited on 03/07/2018).

[31] Howard Ochman, Jeffrey G. Lawrence, and Eduardo A. Groisman. "Lateral gene transfer and the nature of bacterial innovation". In: *Nature* 405.6784 (May 18, 2000), pp. 299–304. ISSN: 0028-0836. DOI: 10.1038/35012500. URL: http://www.nature.com/nature/journal/v405/n6784/full/405299a0.html (visited on 09/01/2015).

[32] Martin F. Polz, Eric J. Alm, and William P. Hanage. "Horizontal Gene Transfer and the Evolution of Bacterial and Archaeal Population Structure". In: *Trends in genetics : TIG* 29.3 (Mar. 2013), pp. 170–175. ISSN: 0168-9525. DOI: 10.1016/j.tig.2012.12.006. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3760709/ (visited on 08/14/2017).

[33] Christian Quast et al. "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools". In: *Nucleic Acids Research* 41 (D1 Jan. 1, 2013), pp. D590–D596. ISSN: 0305-1048. DOI: 10.1093/nar/gks1219. URL: https://academic.oup.com/nar/article/41/D1/D590/1069277 (visited on 08/09/2019).

[34] Abbas H. Rizvi et al. "Single-cell topological RNA-Seq analysis reveals insights into cellular differentiation and development". In: *Nature biotechnology* 35.6 (June 2017), pp. 551–560. ISSN: 1087-0156. DOI: 10.1038/nbt.3854. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5569300/ (visited on 11/15/2017).

[35] Marten Scheffer et al. "Early-warning signals for critical transitions". In: *Nature* 461.7260 (Sept. 3, 2009), pp. 53–59. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature08227. URL: http://www.nature.com/doifinder/10.1038/nature08227 (visited on 02/09/2016).

[36] Marten Scheffer et al. "Generic Indicators of Ecological Resilience: Inferring the Chance of a Critical Transition". In: *Annual Review of Ecology, Evolution, and Systematics* 46.1 (2015), pp. 145–167. DOI: 10.1146/annurev-ecolsys-112414-054242. URL: http://dx.doi.org/10.1146/annurev-ecolsys-112414-054242 (visited on 11/19/2015).

[37] M. Scheffer et al. "Anticipating Critical Transitions". In: *Science* 338.6105 (Oct. 19, 2012), pp. 344–348. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1225244. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1225244 (visited on 01/24/2015).

[38] Patrick D. Schloss et al. "Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities". In: *Applied and Environmental Microbiology* 75.23 (Dec. 1, 2009), pp. 7537–7541. ISSN: 0099-2240, 1098-5336. DOI: 10.1128/AEM.01541-09. URL: https://aem.asm.org/content/75/23/7537 (visited on 08/08/2019).

[39] Justin D. Silverman et al. "A phylogenetic transform enhances analysis of compositional microbiota data". In: *eLife* 6 (Feb. 15, 2017), e21887. ISSN: 2050-084X. DOI: 10.7554/eLife.21887. URL: https://elifesciences.org/articles/21887 (visited on 07/12/2017).

[40] Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. "Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition". In: *Eurographics Symposium on Point-Based Graphics* (2007), p. 11.

[41] Richard R. Stein et al. "Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota". In: *PLoS Comput Biol* 9.12 (Dec. 12, 2013), e1003388. DOI: 10.1371/journal.pcbi.1003388. URL: http://dx.doi.org/10.1371/journal.pcbi.1003388 (visited on 12/18/2014).

[42] Curtis A. Suttle. "Marine viruses — major players in the global ecosystem". In: *Nature Reviews Microbiology* 5.10 (Oct. 2007), pp. 801–812. ISSN: 1740-1526. DOI: 10.1038/nrmicro1750. URL: http://www.nature.com/nrmicro/journal/v5/n10/full/nrmicro1750.html#B2 (visited on 03/23/2017).

[43] Peter J. Turnbaugh et al. "The human microbiome project". In: *Nature* 449.7164 (Oct. 18, 2007), pp. 804–810. ISSN: 1476-4687. DOI: 10.1038/nature06244.

[44] C. H. Waddington. *The Strategy Of The Genes*. 1957. URL: http://archive.org/details/in.ernet.dli.2015.547782 (visited on 08/05/2019).

[45] Jesse R. Zaneveld, Ryan McMinds, and Rebecca Vega Thurber. "Stress and stability: applying the Anna Karenina principle to animal microbiomes". In: *Nature Microbiology* 2.9 (Sept. 2017), p. 17121. ISSN: 2058-5276. DOI: 10.1038/nmicrobiol.2017.121. URL: https://www.nature.com/articles/nmicrobiol2017121 (visited on 11/27/2018).

[46] J. X. Zhou et al. "Quasi-potential landscape in complex multi-stable systems". In: *Journal of The Royal Society Interface* 9.77 (Dec. 7, 2012), pp. 3539–3553. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2012.0434. URL: http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2012.0434 (visited on 10/11/2018).

Figures

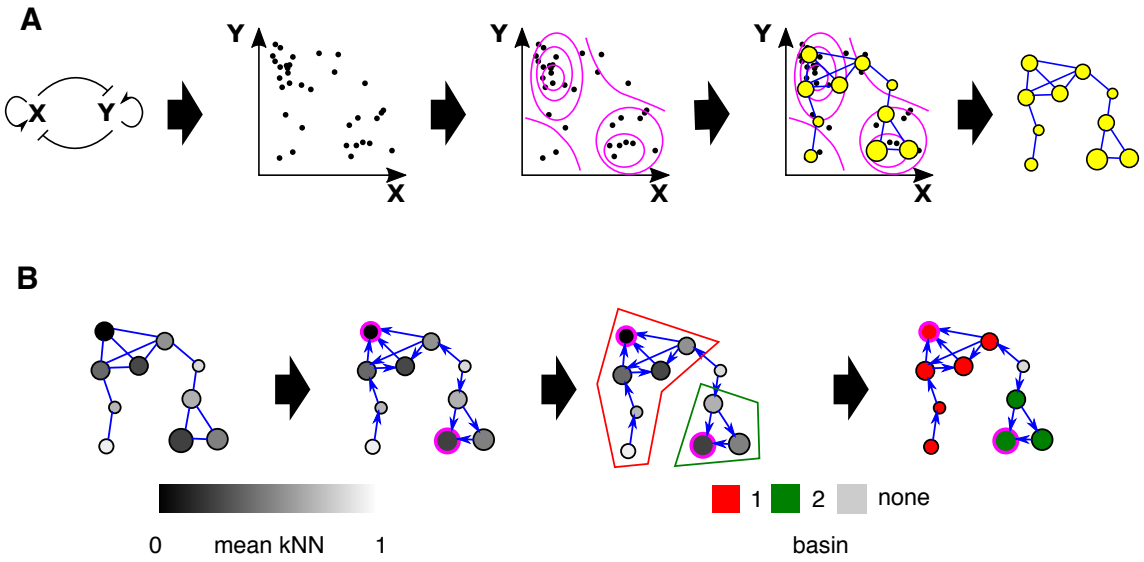

Figure 1: Using Mapper to characterize the microbial phase space. A. Using the Mapper algorithm to infer the potential landscape of a toy ecosystem. The mutually antagonistic interaction between species X and Y leads to denser sampling of the phase space where either X or Y is abundant and the other is rare than in other regions; configurations in which X and Y are similar in density are unstable, as small uncertainties in numerical advantage will eventually lead to the dominance of one species over the other. This probability density is analogous to an inverse of the potential landscape. Mapper infers a 'skeleton' of density from the data represented as a point cloud. This representation preserves major features of the landscape such as the two densely-sampled clusters separated by a sparsely-sampled region. B. Identification of local minima and metastable states in the Mapper graph shown in A. Data density for each vertex is estimated by the mean kNN density (see Methods) for samples associated with that vertex. The graph is converted to a directed graph, with each edge pointing in the direction of increasing kNN density. A local minimum, highlighted in pink, is defined as a vertex that has lower kNN than all its neighbors. Finally, the state associated with a local minimum is defined as the set of vertices that have uniquely shortest directed graph distance to that minimum. Non-minima vertices with equal graph distances to multiple local minima are unassociated with any state (grey).

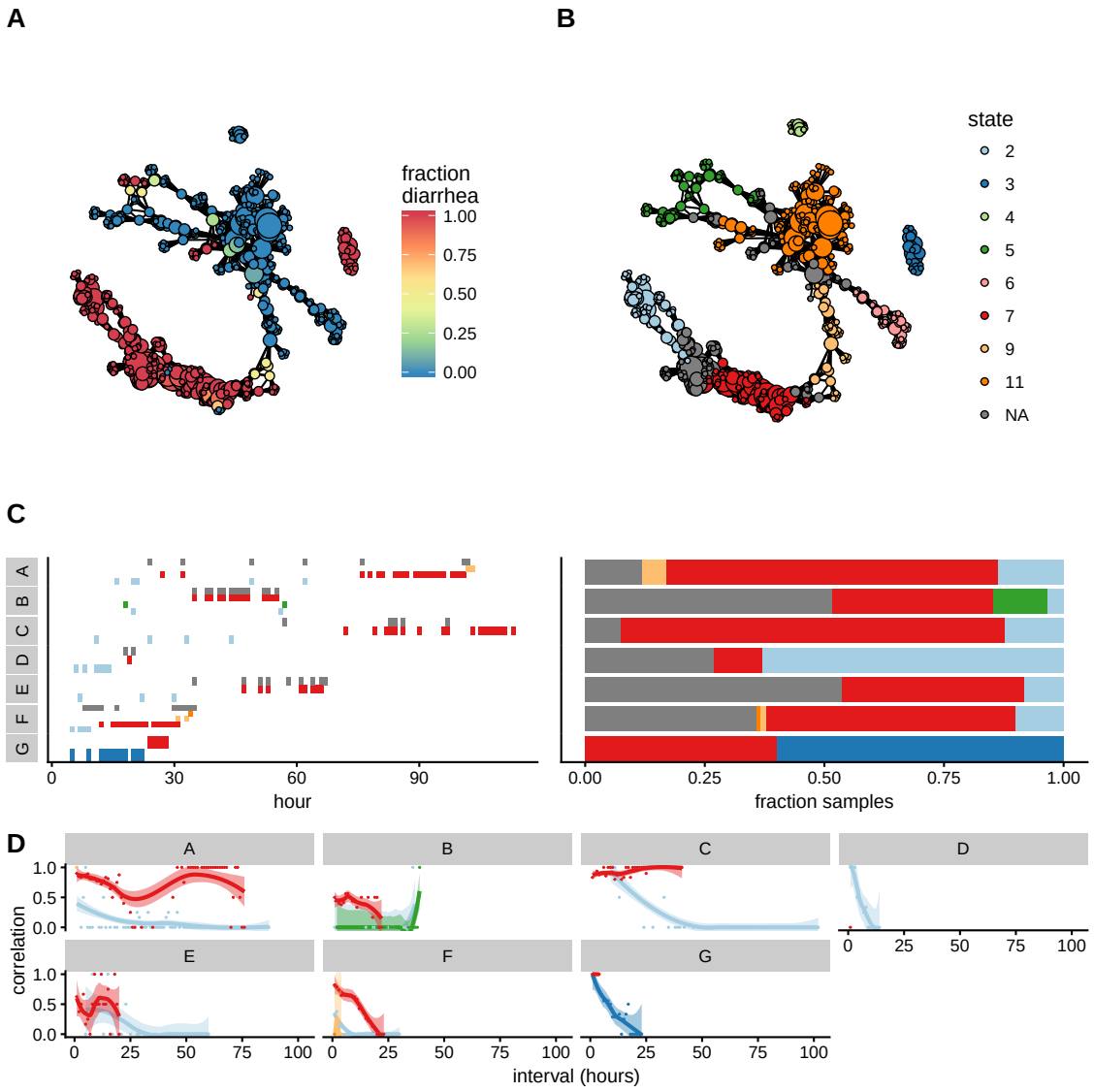

Figure 2: The phase space of the cholera gut microbiome. A. Mapper representation of the combined cholera data reveals disease- and healthy-associated neighborhoods of the phase space. Color: fraction of samples in each vertex associated with diarrhea. Connected components of the Mapper graph representing only one sample are not shown. Disjoint regions of phase space are represented as separate connected components. B. Partitioning of the phase space into metastable states. Vertices unassigned to any state are colored in grey. C. Left: progression of subject compositions during the diarrhea phase by state, showing persistence of states over time. Y axis and color indicate state index, with color indexing as in B. Where a sample was associated with multiple states, all were included. Right: frequency of samples associated with each states during the diarrhea phase for each subject with colors as in B. D. Temporal correlation function for the diarrhea phase of each subject. Lines: smoothed empirical mean; ribbons: standard error of the mean. Values outside the range of $0 \leq y \leq 1$ omitted.

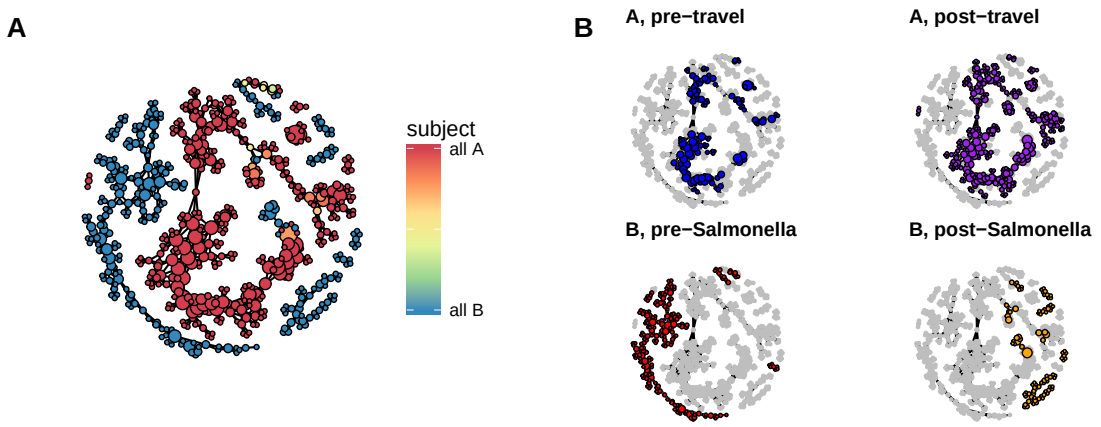

Figure 3: The phase space of two healthy adult male gut microbiomes. A. Mapper representation of the combined daily time series of two healthy adult human gut microbiomes. Connected components of the Mapper graph representing only one sample are not shown. B. Regions of phase space occupied by each subject before after perturbation.

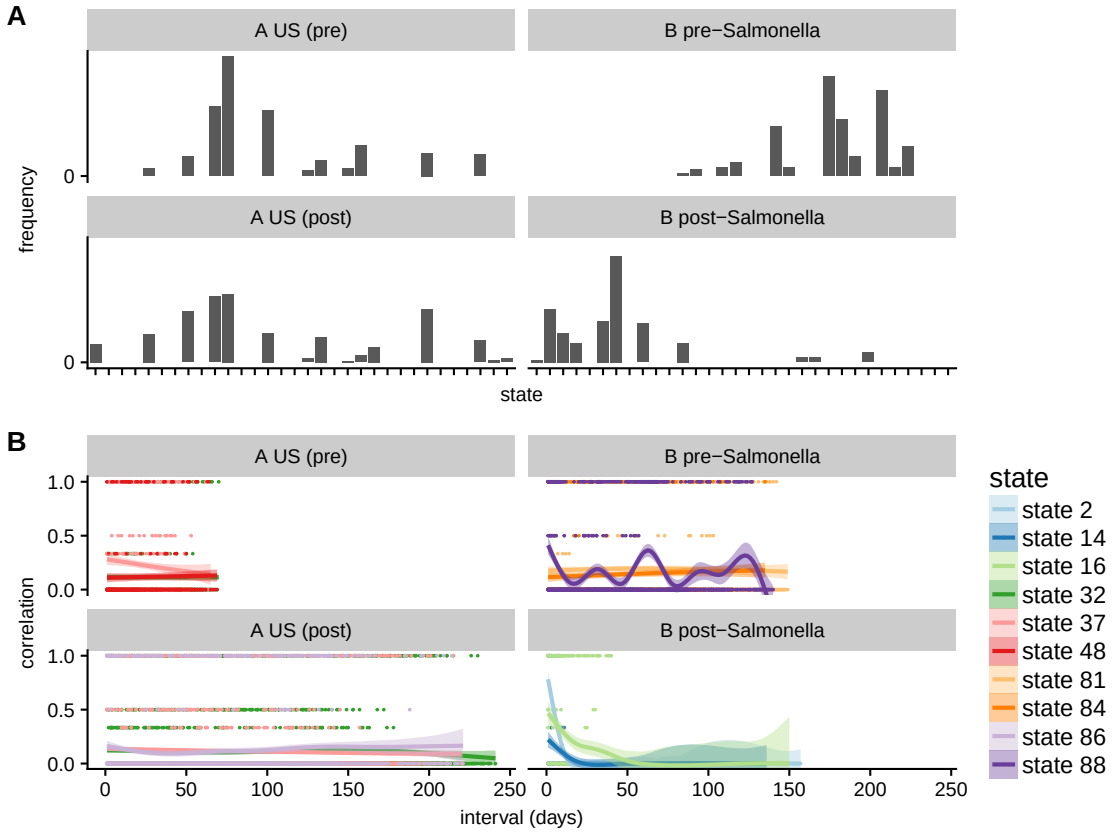

Figure 4: States and dynamics of two healthy adult male gut microbiomes. A. Frequency of states for healthy periods before and after perturbation. X axis: state index. Y axis: frequency of samples. B. Temporal correlation functions for the three most probable states during each event in the 'healthy' phases of each subject. Lines: smoothed empirical mean; ribbons: standard error of the mean.

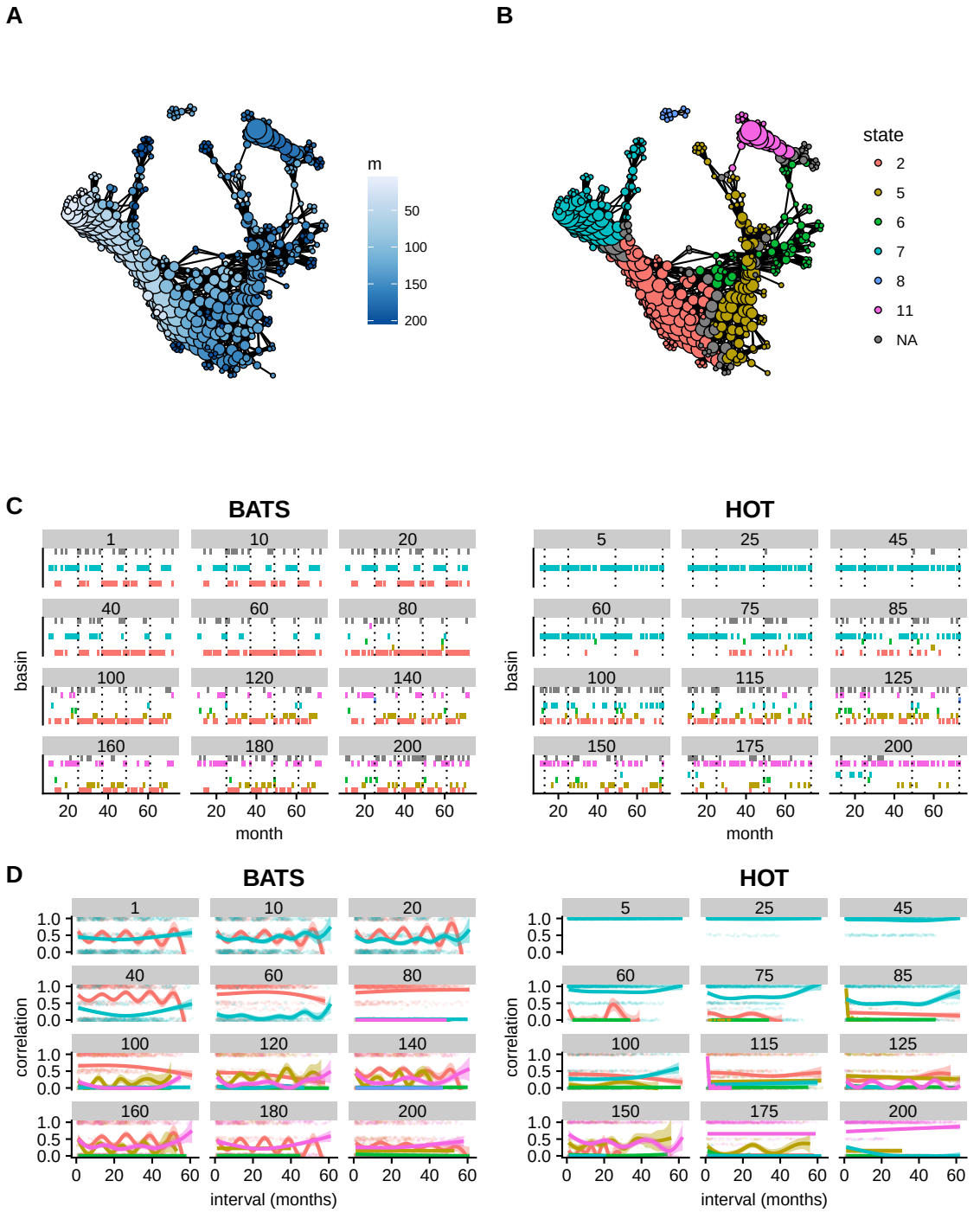

Figure 5: The landscape of *Prochlorococcus* communities. The combined phase space of two *Prochlorococcus* communities inhabiting the Atlantic and Pacific Oceans, respectively. Connected components of the Mapper graph representing only one sample are not shown. A. Vertices colored by mean depth in meters of represented samples. B. Partitioning of the phase space into states. C. Successions of states for each site-depth fraction combination. Dotted lines indicate samples during January. Colors indicate states as in B. D. Temporal correlation functions for each state per site-depth fraction combination.

Tables

Data set	# intervals for (rank(PCo1), rank(PCo2))	% overlap	# bins
Cholera	$(15, 15)$	70	10
Two healthy adult males	$(30, 30)$	50	10
Prochlorococcus	$(20, 20)$	60	10

Table 1: Hyperparameters used to generate the Mapper representation of each data set.

Additional Files

Supplementary information

Supplementary figures showing the results of the data rarefaction test. Supplementary figure showing the temperature gradients across the *Prochlorococcus* phase space.