Metagenomics meets time series analysis: Unraveling microbial community dynamics

Article *in* Current Opinion in Microbiology · May 2015

DOI: 10.1016/j.mib.2015.04.004 · Source: PubMed

5 authors, including:

Karoline Faust
KU Leuven
61 PUBLICATIONS **13,158** CITATIONS
SEE PROFILE

Leo Lahti
University of Turku
173 PUBLICATIONS **3,741** CITATIONS
SEE PROFILE

Didier Gonze
Université Libre de Bruxelles
116 PUBLICATIONS **3,646** CITATIONS
SEE PROFILE

Willem M De Vos
Wageningen University and University of Helsinki
1,764 PUBLICATIONS **115,127** CITATIONS
SEE PROFILE

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Metagenomics meets time series analysis: unraveling microbial community dynamics

Karoline Faust[1,2,3,9], Leo Lahti[4,5,9], Didier Gonze[6,7], Willem M de Vos[4,5,8] and Jeroen Raes[1,2,3]

The recent increase in the number of microbial time series studies offers new insights into the stability and dynamics of microbial communities, from the world's oceans to human microbiota. Dedicated time series analysis tools allow taking full advantage of these data. Such tools can reveal periodic patterns, help to build predictive models or, on the contrary, quantify irregularities that make community behavior unpredictable. Microbial communities can change abruptly in response to small perturbations, linked to changing conditions or the presence of multiple stable states. With sufficient samples or time points, such alternative states can be detected. In addition, temporal variation of microbial interactions can be captured with time-varying networks. Here, we apply these techniques on multiple longitudinal datasets to illustrate their potential for microbiome research.

Addresses
[1] Department of Microbiology and Immunology, Rega Institute KU Leuven, Leuven, Belgium
[2] VIB Center for the Biology of Disease, VIB, Belgium
[3] Laboratory of Microbiology, Vrije Universiteit Brussel (VUB), Brussels, Belgium
[4] Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands
[5] Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland
[6] Unité de Chronobiologie Théorique, Faculté des Sciences, Université Libre de Bruxelles (ULB), Brussels, Belgium
[7] Interuniversity Institute of Bioinformatics in Brussels (IB)^2, ULB-VUB, Brussels, Belgium
[8] Immunobiology Research Program, Department of Bacteriology and Immunology, Haartman Institute, University of Helsinki, Helsinki, Finland

Corresponding authors: Faust, Karoline (karoline.faust@vib-kuleuven.be) and Raes, Jeroen (jeroen.raes@vib-kuleuven.be)

[9] These authors contributed equally to this work.

Current Opinion in Microbiology 2015, **25**:56–66

This review comes from a themed issue on **Environmental microbiology**

Edited by **Nicole King** and **Susann Müller**

http://dx.doi.org/10.1016/j.mib.2015.04.004

Insights from microbial time series data

Recent improvements in high-throughput sequencing have led to a rise in longitudinal studies that record the temporal variation of microbial communities in a wide range of environments. These time series studies can offer unique ecological insights on community stability and response to perturbations that cannot be gained otherwise. Here, we provide an overview of these insights and discuss a range of methods (summarized in Table 1) for the analysis of longitudinal sequencing datasets.

In a recent meta-analysis of longitudinal studies, roughly half of the communities had time-decay curves with negative slopes, that is their community dissimilarities increased with time [1]. In addition, the temporal variability of microbial community diversity was found to be comparable across studies within the same environment but varied across them, being lowest in soil and brewery wastewater and highest in the human palm and the infant gut [1].

Temporal variation is not restricted to global diversity — long-term studies conducted for marine microbiota [2,3•,4] revealed, among other insights, strongly seasonal dynamics of individual community members. Some microbial communities go through a series of predictable states after colonization (primary succession), which occurs for instance during the formation of dental plaque, where oxygen-tolerant early colonizers prepare the ground for later oxygen-sensitive colonizers [5], and in soil and leaf surface communities (reviewed in [6]). In some cases, such as infant gut microbiota colonization [7], communities vary considerably in the initial stages of succession, but stabilize at similar states.

While large-scale studies such as the Human Microbiome and MetaHIT projects explored the phylogenetic and functional composition of the healthy human microbiota and its inter-individual variation [8,9], studies that investigate its temporal variation are still rare and either include many time points of a few subjects or a few time points of many subjects; thus large-scale (both in length and cohort size) longitudinal studies are needed. In one of the human microbial time series with the largest number of time points available to date, Caporaso *et al.* found considerable variability within body sites and only a small

Table 1

Summary of time series analysis techniques used in microbiome profiling studies. The table briefly explains each technique and the interpretation of its result, with references to specific implementations and microbial community studies applying these techniques. We do not mention the treatment of sequencing depth differences, which is relevant for all time series data generated with sequencing.

Name	Summary	Example implementation	Interpretation/goal	Pre-processing	Remarks	Reference
Time-decay	Log-linear model fitted to sample dissimilarities versus time between observations.	R package vegan, function vegdist (for sample-wise dissimilarities)	Rate of community change ('turnover').	Filtering of rare OTUs is recommended.	Sensitive to the time scale.	[1]
Detrending	Common techniques include regression/ differencing to remove linear trends	R package pracma, function detrend and package stats, function diff	Removal of trends that may hide underlying dynamics.	None.	Trends may be cyclic (see Auto-correlation) or non-linear.	[19••] (removal of autocorrelation)
Augmented-Dickey Fuller test	Stationarity of a time series is tested by fitting an autoregressive model.	R package tseries, function adf.test	Test whether microbial community has reached a stable state.	Filtering of rare OTUs is recommended.	Taxonomic resolution matters; strain-level dynamics may differ strongly from genus-level dynamics.	[19••]
Cross-correlation	Correlation between two time series computed as a function of the lag of one with respect to the other one.	R package stats, function ccf	Delayed response of a taxon to another taxon or environmental parameter.	Equidistant time points. Filtering of rare OTUs is recommended.	Lagged correlation does not establish causality. Time scale matters. Cross-correlations at larger lags are less reliable due to reduced overlap.	[19••]
Local similarity analysis (LSA)	Dynamic programming determines the lags between two time series that optimize their dot product [37].	eLSA: http://meta.usc.edu/softs/lsa/ [62,63] fastLSA: http://hallam.microbiology.ubc.ca/fastLSA/install/index.html [64]	(Delayed) response of a taxon to another taxon or environmental parameter.	Equidistant time points. Filtering of rare OTUs is recommended.	Lagged association does not establish causality. The result of the analysis is affected by the time scale.	[3•,38•,39]
Time-varying network inference	Various techniques, for example static network inference applied to time segments or time-varying dynamic Bayesian network (DBN) inference	A number of microbial association network inference methods is available [29,30,62,64,65]	Time evolution of association networks and stability of individual associations.	Filtering of rare OTUs is recommended. Time series of sufficient length needed to detect associations in segments.	Association does not establish causality.	[66] (fishery data)
Auto-correlation (Auto-correlogram)	The correlation of the time series to itself is plotted for all possible lags.	R package stats, function acf	Presence of periodical patterns, for example seasonality.	Equidistant time points.	Auto-correlations at larger lags are less reliable due to reduced overlap.	[2,3•]
Hurst exponent	Power law fitted to the lengths of time series segments versus the ranges of their cumulative deviations from the mean, rescaled by the standard deviation.	R package pracma, function hurstexp	Presence of persistent trends (Hurst exponent close to one) versus frequent fluctuations (Hurst exponent close to zero). A Hurst exponent of 1/2 indicates random walk.	Equidistant time points.	The Hurst exponent is sensitive to the sampling frequency and length of the time series.	[14••]

Table 1 (*Continued*)

Name	Summary	Example implementation	Interpretation/goal	Pre-processing	Remarks	Reference
Lyapunov exponent	Speed of increase of small perturbations	R package tseriesChaos, function lyap_k; TISEAN [67]	A positive Lyapunov exponent is an indicator of chaos, which means that the community dynamics cannot be predicted in the long term.	Equidistant time points. In the presence of sharp spikes, fourth-root power trans-formation is recommended.	Time series need to be sufficiently long. Sampling frequency matters.	[47••]
Predictability analysis	Prediction accuracy of a non-linear and a linear model is compared for an increasing number of time points.	None	If the non-linear model outperforms the best-fitting linear model: non-linear community dynamics.	Equidistant time points.	Difference in performance may be due to overfitting, which has to be controlled.	[47••]
Bistability analysis	Bistability indicated by bimodal distribution and instability at intermediate abundances.	R package microbiome	Detection of alternative stable states in individual taxa.	Logarithmic abundances.	Confounding factors need to be controlled. Appropriate sampling frequency depends on system stability.	[52]
Early warning signs	Several generic indicators, including increased variance and auto-correlation.	earlywarnings tool box: http://www.early-warning-signals.org [68]	Anticipation of an imminent and sudden state shift.	Logarithmic abundances and equidistant time points.	Indicators hint at increased likelihood of a state shift but do not provide accurate predictions.	[56]

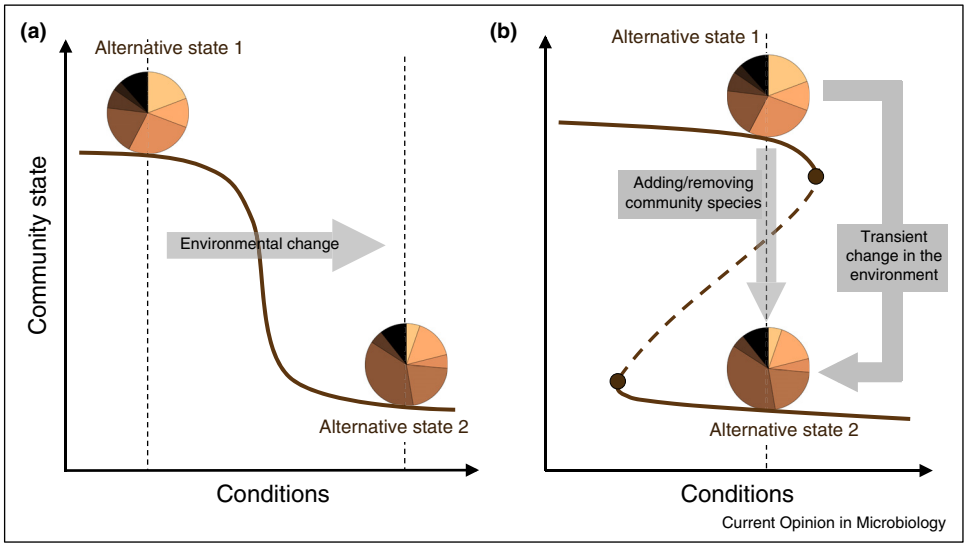
number of persistent core microbial taxa in two healthy subjects [10]. In contrast, Faith *et al.* reported a high persistence on strain level for deeply sequenced gut samples [11]. Despite daily variations, subject-specific differences in gut microbiota, discovered in the late 90s [12], are preserved over decades [13].

Hekstra and Leibler [14••] suggest that species interactions can be inferred either from biological replicates at a single time point or from a single system monitored over several time points. Since controlled replicates are not easily available for microbiome profiling studies, these results imply that combining information across multiple time series can improve the inference of interactions from observations. Moreover, replicates are crucial to distinguish stochastic fluctuations from underlying statistical laws governing ecosystem dynamics.

Microbial communities tend to evolve towards a stable composition. A change in the community state can be triggered by changes in external conditions (e.g. diet), by a direct modification of the microbial community (e.g. by antibiotic or probiotic treatment) [15], or by a transient perturbation that pushes the system into an alternative stable state (Figure 1). In the latter case, alternative stable states are not fully determined by environmental factors, but emerge from the complex non-linear interactions among community members. Such systems are referred to as bistable or multi-stable. Thus, perturbation studies help to probe community dynamics and resilience. A lake ecosystem undergoing a massive alteration was shown to recover its initial state after a few days, demonstrating resilience [16]. In contrast, less complete recovery has been reported for the gut microbiota after antibiotic treatment [17,18], pathogen invasion [19••] or small bowel transplantation [20], hinting at the existence of alternative stable states. Similar observations have been made for environmental changes such as substrate overload in a biogas reactor [21] or dietary interventions [22,23]. However, responses to perturbations can vary substantially between individuals [23]. Thus, incomplete recovery may reflect stochastic effects rather than alternative stable states. For instance, stochastic processes dominate groundwater communities after perturbation by nutrient injection, which the authors attribute to reduced competition and priority effects [24]. Priority effects, that is the exclusion of latecomers by strains already present, have been observed in many other ecosystems (e.g. in rock pool communities [25]) and may contribute to the variability observed in the early stages of the human gut community development.

These complex interactions between microorganisms among themselves and with their environment are key contributors to microbial dynamics and increasingly explored with network inference techniques.

Figure 1



A change in environmental conditions (dashed vertical line) can affect a microbial community (pie chart) such that its composition shifts into an alternative state **(a)**. An ecosystem can also exhibit two (or more) stable states under the same conditions **(b)**. A switch of the steady state composition can be induced either by a change in the composition (removal of species via antibiotics or introduction of new species) or by a transient change of the environmental conditions that pushes the system beyond the tipping points (small brown circles). In both cases, a small change in external conditions in the vicinity of a tipping point can trigger an abrupt community shift. Such a state switch may be preceded by early warning signs [55••].

Network reconstruction from microbial time series

A number of methods are available to construct taxon co-occurrence networks from cross-sectional data (reviewed in [26]), ranging from correlation combined with permutation tests [27] and similarity assessment with the hypergeometric distribution [28] to approaches dealing with compositionality [29,30], indirect edges [31] and multiple factors influencing taxon abundances (multiple regression [29,32]).

These static network inference techniques can be applied to construct dynamic models. For instance, community dynamics is often mathematically described with the generalized Lotka–Volterra equations, which models the change in abundances as a function of taxon-specific growth rates and pair-wise interaction strengths. These parameters can be determined from time series data by (sparse) multiple regression ([33–36], also see [26]).

However, time series provide additional information ignored by such methods, namely the ordering and dependencies between the time points. These properties are exploited by dynamic approaches.

Dynamic network inference

Cross-correlation quantifies the similarity between shifted time series, as for instance in David *et al.*, where cross-correlations were calculated with varying lags from de-trended, clustered time series in order to detect associations between taxon abundances and host metadata [19••]. Interpolation, that is fitting a function to observed values to fill 'gaps' in a time series, allows selecting equidistant time points and lags smaller than the sampling interval, but can introduce bias, for instance when linear interpolation is applied to non-linear dynamics.

Local similarity analysis (LSA) employs dynamic programming to identify the lags between two time series that maximize their similarity score [37] and can therefore also detect associations between shifted time series. For instance, LSA was applied to predict interactions between bacteriophages and their hosts and between protist grazers and their prey from the San Pedro Ocean time series [38•,39].

Dynamic Bayesian network (DBN) techniques model the state of each variable as a function of the parent variables in the preceding time point. Thus, DBNs can detect dynamic dependencies, including cyclic ones, in time series [40]. Compared to standard Bayesian networks, their dynamic counterparts provide a more powerful modeling framework, albeit with increased computational cost, limited scalability [41], and difficulties in identifying the correct models when many alternative networks explain the data equally well.

Another group of dynamic network inference techniques is based on cross-prediction, which quantifies how well the future of one time series can be predicted from

another time series within the same system. These methods include Granger causality [42] and Sugihara's novel convergent cross mapping [43].

Time-varying networks

All techniques mentioned above infer a single network of species interactions from the entire time series. However, species interactions and hence the network structure may change over time. Time-varying network inference techniques aim at inferring such evolving network structures. One option is to build static networks for different, potentially overlapping, segments of a time series. A time-varying network constructed in this way from data reported in [19••] shows that in the human gut microbiota, taxon associations are not equally stable and vary in their strength over time (Figure 2, Supplementary Movie 1).

Non-stationary [44] and time-varying [45] DBNs can also be used to infer temporal changes in network structure [45]. An additional advantage of the time-varying DBNs is that they can also estimate the optimal number of time windows.

Lessons from microbial time series analysis

Time series analysis is used to detect regularities such as trends and periodicity, but also the opposite: irregularities that challenge prediction and abrupt state shifts potentially preceded by early warning signs.

Trends, periodicity and predictability

Microbial abundance may increase or decrease monotonically by natural growth or in response to environmental variables such as temperature. The persistence of such trends can be quantified with the Hurst exponent, which indicates trend-following ($H > 0.5$), random-walk ($H \sim 0.5$), or highly fluctuating ($H < 0.5$) patterns [46]. Autocorrelograms visualize auto-correlations in time-series across all possible time lags and can highlight repetitive patterns and seasonality, as for instance in marine microbial communities [2,3] (see also Figure 3).

Linear trends and seasonality may mask other underlying signals, however, and are hence often removed by 'detrending' techniques. The Augmented Dickey Fuller test, which tests for the absence of such trends, was recently applied to quantify gut microbiota stability [19••].

Predictability analysis compares the ability of different models to predict future behavior of a time series based on earlier observations. Benincà and co-workers demonstrated, for instance, that a non-linear model significantly outperformed the best-fitting linear model in predicting temporal dynamics of phytoplankton communities [47••]. Rapid decrease in predictability with increasing prediction time hints at chaotic systems that are sensitive to small changes in initial conditions and perturbations, and hence not predictable in the long term. Positive (maximum) Lyapunov exponents for selected taxa from the Western English Channel time series [3] in Figure 3 hint at chaos underneath seasonal variation. This is in line with Dakos *et al.*, who suggested that chaotic community dynamics is coupled with periodic forcing by the seasons [48]. However, distinguishing chaos from stochastic variation is challenging, and Lyapunov estimation is more robust when derived from explicit dynamical models rather than noisy observations; Benincà *et al.* calculated Lyapunov exponents both from the observational data and a model fitted to these data to test for chaotic dynamics.

Alternative stable states and early warning signs

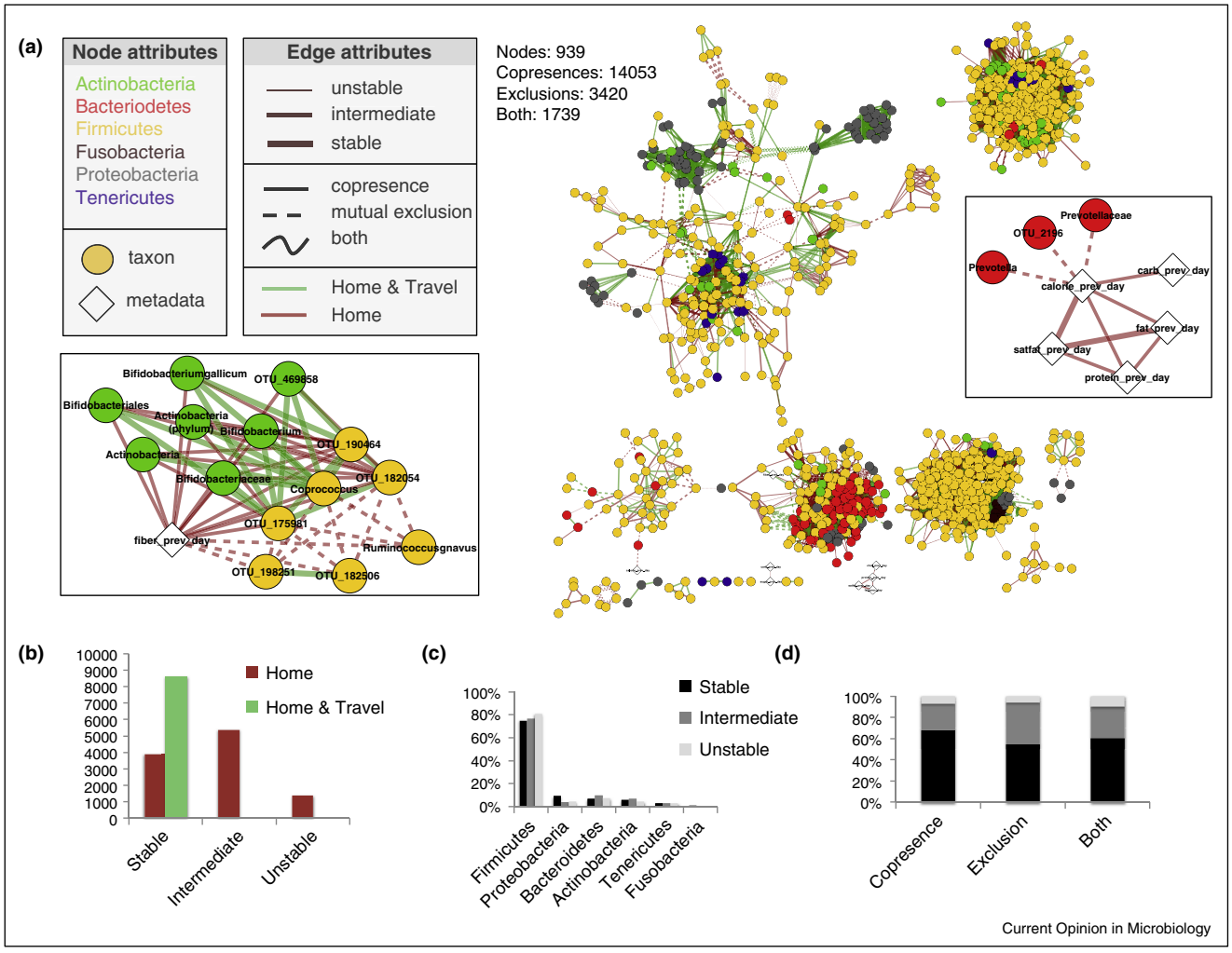
Arumugam and colleagues suggested that the human gut microbiota preferentially assembles into three configurations of community composition known as enterotypes [49]. Since enterotypes have been linked to dietary habits [22,50], they might represent alternative states driven by external factors (Figure 1a) rather than multi-stability (Figure 1b). Though the exact nature of enterotypes is being debated [51], members of the enterotype drivers such as Prevotella and Bacteroides, as well as many other gut taxa, have been suggested to present independently varying, bistable 'tipping elements' of the gut microbiota, where the bistability is reflected by bimodal distributions with peaks at low and at high abundances ([52], Supplementary Figure 3), coupled with reduced stability at the intermediate abundance range. We also detected a bistable genus in the vaginal microbiota [53••], whose states of low and high abundance are visible across time in a single subject as well as across subjects (Figure 4). While sample clustering based on community similarity is frequently applied to discover alternative states in microbial communities (e.g. [49,53••,54]), sample-wise clustering is not sufficient for establishing multi-stability, as the clusters could be associated with unknown environmental factors.

Even a small perturbation may trigger an abrupt shift to a new stable state in a system that is close to a bifurcation or 'tipping' point (Figure 1). The theory of early warning signs claims that such abrupt state switches can be predicted from the time series directly. For instance, the increasing recovery time from small perturbations, as the system reaches the tipping point, can be quantified [55••]. Such early warning signs have been reported for instance in lake ecosystems [56].

Discussion

The measurement technique, spatial and temporal frequency of the sampling as well as the availability of replicates can all strongly affect the results of a time series analysis. Whereas frequent sampling is crucial to capture the full richness of community dynamics [57•], the ideal sampling frequency depends on the system

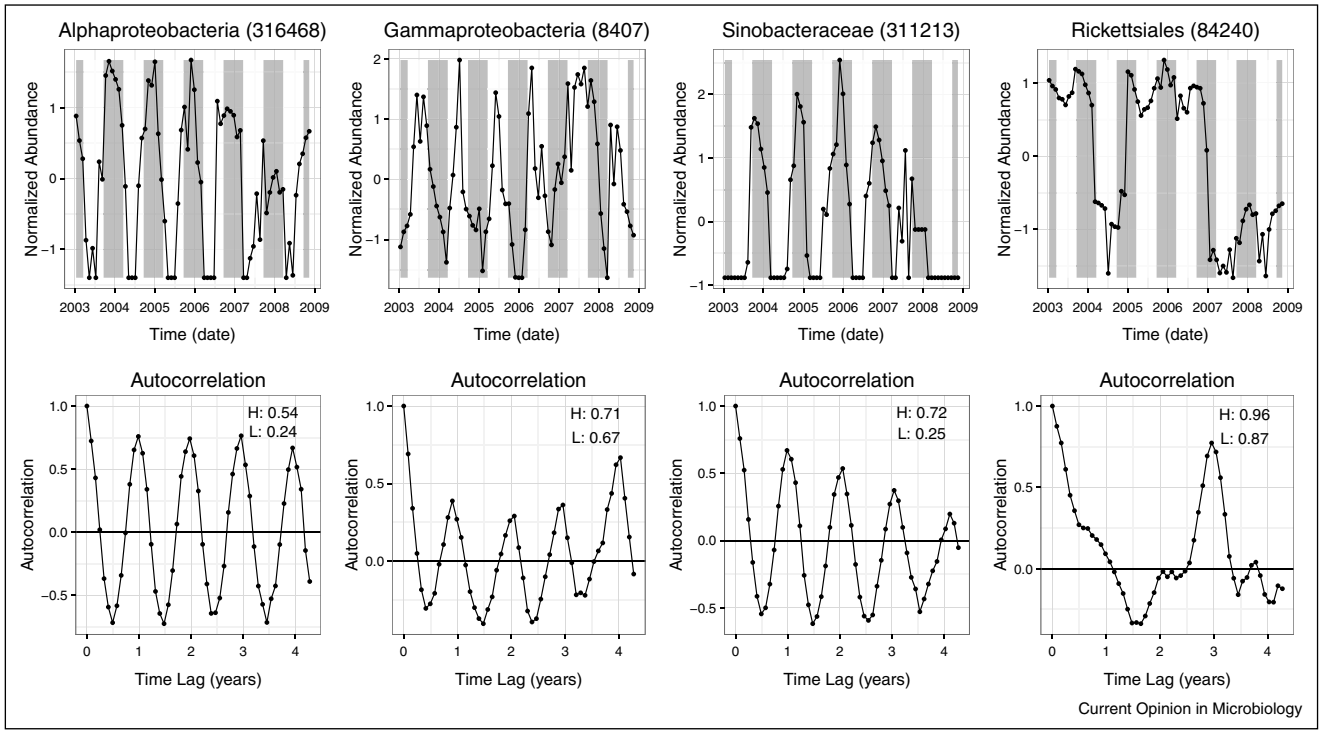
Figure 2



Time-varying association network between gut microbiota members (data from [19••], individual A) shows considerable variation in the community network over time. **(a)** The clustered network summarizes 31 time-window-specific networks, constructed with CoNet [29]. Edges represent global associations between organisms as well as those specific to certain time-periods. For instance, there are edges specific to home periods (in red) and edges present in both travel and home periods (in green; stable edges). The boxed subnetworks show the first neighbors of the fiber and calorie intake metadata nodes (available for home periods). Prevotella and Ruminococcus OTUs are inversely correlated with calorie and fiber intake, respectively, whereas Bifidobacterium and Coprococcus OTUs (OTU-190464 is a Lachnospiraceae member of unknown genus) are correlated with fiber intake, as reported in [19••]. Supplementary Figure 1 summarizes positive and negative class-level relationships, respectively. **(b)** All edges occurring in both home and travel periods are stable. Overall, 65% of interactions can be categorized as stable across time, with 28% and 7% of associations being intermediate and unstable, respectively. Remarkably, one third of the unstable edges occurs in only one window, namely in the home-coming period. **(c)** Phylum-level node composition changes slightly between stability categories, with Proteobacteria engaging more in time-independent, stable interactions, while Firmicutes having more intermediate and unstable associations over time. **(d)** Stable edges contribute a higher percentage to co-presences than to mutual exclusions or to mixed interactions (where edges change their interaction type across windows). Stability-stratified edge percentages, phyla and interaction types were computed prior to network clustering.

characteristics, rate of change, and the study hypotheses. In ocean microbiomes driven by seasonal patterns, typical sampling intervals range from weeks to months [48], whereas, for instance, in vaginal microbiomes more regular sampling frequencies counted in days have been used [53••]. Different sampling frequencies can be used to quantify different, complementary properties of a system and can even change the associations inferred from the time series data, as demonstrated for SAR11 members, which were highly correlated on a daily, but not on a monthly scale [58]. While seasonality typically characterizes the strongest signal at broader sampling intervals in marine communities, denser sampling can reveal chaotic fluctuations [48].

Figure 3



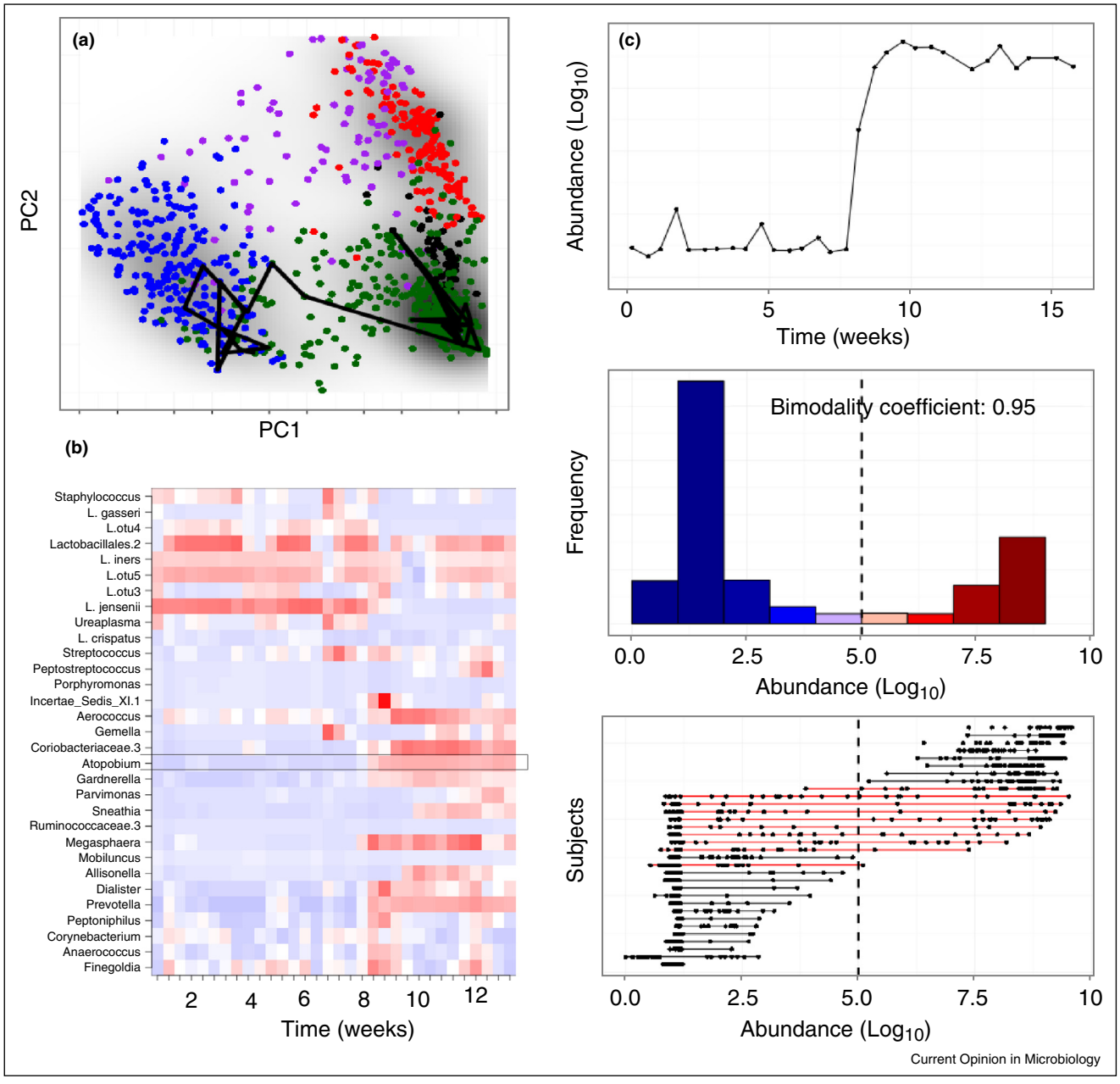
Standardized abundance profiles from the Western English Channel microbiota time series [3•] spanning six years (upper panel, light and gray bars representing summer and winter). As reported in Gilbert *et al.*, the seasonality of specific taxa is visible in their abundance profiles and autocorrelograms (lower panel), which present the correlation of the time series with itself for different lags. The stronger the seasonal pattern, the more closely the autocorrelogram approaches a periodic function. Rickettsiales are an interesting exception: their dynamics is a combination of seasonal fluctuation and a three-year periodicity (three-year peak in the autocorrelogram). Insets in the autocorrelograms quantify the extent and regularity of the fluctuations with the Hurst (H) and the (maximal) Lyapunov exponent (L), estimated from interpolated and standardized time series. Hurst and Lyapunov exponent point out Rickettsiales as the most persistent and irregular and Alphaproteobacteria as the least persistent and most regular taxon.

In general, increased sampling frequencies can provide increased resolution on the system dynamics, but there are limits due to costs and, in host microbiota studies, ethical issues. Besides sampling frequency, the regularity of sampling is an important factor for many analysis techniques, such as autocorrelation. Interpolation can provide estimates for specific time points when regular sampling intervals are not available, but can also be misleading if it relies on inaccurate modeling assumptions.

Although many standard approaches for longitudinal analysis require long time series with short and regular sampling intervals, the currently available metagenomic time series tend to be short (few time points), gapped (missing time points), sparse (zero-rich) and noisy, necessitating preprocessing steps such as filtering, standardizing, interpolation and detrending to make time points equidistant and comparable. Small sample sizes and low signal-to-noise ratios combined with heavy multiple testing from simultaneous profiling of up to thousands of taxonomic units poses challenges for statistical analyses. These problems with microbiome time series are particularly challenging in human studies, where recruitment and regular sampling of study participants under specific interventions and over long periods of time can be difficult and expensive. Another challenge to the analysis of microbial time series is the interplay of population and environmental dynamics, especially in air, river and ocean currents [58]. For instance, the importance of hydrological parameters and upstream events was recently demonstrated for river communities [59]. In general, dynamic environments increase sample heterogeneity, which can only be addressed by combining longitudinal with cross-sectional sampling.

While simulations with varying time series lengths and noise levels may help to determine the required number of time points for particular analysis tasks, the analyses could also benefit from improved statistical techniques to integrate information across multiple time series [60]. A recent study demonstrated how pooling data from short time series across many individuals helps to quantify state stability in large cohorts [52]. Moreover, distinguishing

Figure 4



PCA visualization **(a)** summarizing the vaginal microbial community compositions encountered in a cohort of 32 healthy women (data from [53••]). The visualization shows the trajectory (black line) of an individual (subject 10) through community composition space. The community types identified by Gajer *et al.* are indicated by red (I), black (II), green (III), purple (IVA) and blue (IVB); the gray background shade indicates the density of data points (in samples of the entire cohort). The trajectory highlights an abrupt shift from community type III to the community type IVA in subject 10. The heatmap **(b)** visualizes the abundance of the most abundant vaginal OTUs in individual 10 across time (horizontal axis). Blue and red indicate low and high abundance, respectively, with respect to mean abundance of the indicated OTU across all samples. The Atopobium abundance variation across time is highlighted by a black frame and seen to switch from low to high abundance in the 8th week, after the onset of menses. **(c)** Time series of Atopobium abundances for subject 10 across the sampling period (upper panel). The abundance histogram (middle panel) indicates two distinct states of low (blue) and high (red) abundance, confirmed by a high bimodality coefficient of 0.95 (implemented in the microbiome R package). The illustration of Atopobium abundances (lower panel; black dots) in all 32 subjects (horizontal lines) further indicates that low and high abundance states are divided by an unstable state, or a tipping point beyond which the system shifts to an alternative stable state (dashed line). This indicates instability at the intermediate abundance range.

between complex system dynamics such as chaos and stochastic variation, or 'noise', can be challenging [61], but can be to some extent addressed by systematic model comparisons [47••,61].

Whereas tests of community stability [19••] and early warning signs help to understand community dynamics, abrupt state shifts may also occur in response to unpredictable external perturbations. Uncontrolled environmental factors (confounding factors) can also hamper dynamic network inference, when a relationship is inferred between two taxa that respond to changing environmental factors, but do not interact. Thus, time series alone are not sufficient to distinguish correlation from causality.

The investigation of the impact of network structure on state transitions is still in its infancy [55••]. An interesting future line of research is to explore whether time-varying networks have 'early warning' properties that can predict such transitions.

Despite the challenges, time series analysis techniques already provide a rich set of tools to gain insights into temporal patterns, help to understand system dynamics and responses to perturbation, and to construct predictive models. We hope that this review will help to apply these powerful techniques in microbiology and metagenomics, where longitudinal time series and associated modeling challenges are now being encountered at an accelerating pace.

Acknowledgements

We would like to thank Lawrence A. David and Jack A. Gilbert for sending us their data sets, as well as Gipsi Lima-Mendez, Flora Vincent and Nicolas Henry for helpful discussions. In addition, we thank our reviewers for their insightful comments.

K. Faust and J. Raes are supported by the Research Foundation Flanders (FWO), the Flemish agency for Innovation by Science and Technology (IWT), the EU-FP7 grant METACARDIS HEALTH-F4-2012-305312, by KU Leuven and the Rega Institute. L. Lahti and W. de Vos benefit from grants of the Academy of Finland (256950 and 141140, respectively), the European Research Council (ERC250172) and the Alfred Kordelin foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.mib.2015.04.004.

References and recommended reading

Papers of particular interest, published recently, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Shade A, Gregory Caporaso J, Handelsman J, Knight R, Fierer N: **A meta-analysis of changes in bacterial and archaeal communities with time**. *ISME J* 2013, **7**:1493-1506.

2. Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naee S: **Annually reoccurring bacterial communities are predictable from ocean conditions**. *Proc Natl Acad Sci U S A* 2006, **103**:13104-13109.

3. • Gilbert JA, Steele JA, Caporaso JG, Steinbrueck L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I *et al.*: **Defining seasonal marine microbial community dynamics**. *ISME J* 2012, **6**:298-308.
This study demonstrates plankton seasonality in one of the longest marine time series available, consisting of monthly samples spanning six years. It also provides an example to what extent a rare taxon can dominate the community in a bloom.

4. Giovannoni SJ, Vergin KL: **Seasonality in ocean microbial communities**. *Science* 2012, **335**:671-676.

5. Kolenbrander PE, Andersen RN, Blehert DS, Egland PG, Foster JS Jr: **RJP: Communication among oral bacteria**. *Microbiol Mol Biol Rev* 2002, **66**:486-505.

6. Fierer N, Nemergut D, Knight R, Craine JM: **Changes through time: integrating microorganisms into the study of succession**. *Res Microbiol* 2010, **161**:635-642.

7. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO: **Development of the human infant intestinal microbiota**. *PLoS Biol* 2007, **5**:1556-1573.

8. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS *et al.*: **Structure, function and diversity of the healthy human microbiome**. *Nature* 2012, **486**:207-214.

9. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T *et al.*: **A human gut microbial gene catalogue established by metagenomic sequencing**. *Nature* 2009, **464**:59-65.

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

11. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL *et al.*: **The long-term stability of the human gut microbiota**. *Science* 2013, **341**:1237439.

12. Zoetendal EG, Akkermans ADL, de Vos WM: **Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria**. *Appl Environ Microbiol* 1998, **64**:3854-3859.

13. Rajilić-Stojanović M, Heilig HG, Tims S, Zoetendal EG, de Vos WM: **Long-term monitoring of the human intestinal microbiota composition**. *Environ Microbiol* 2013, **15**:1146-1159.

14. •• Hekstra DR, Leibler S: **Contingency and statistical laws in replicate microbial closed ecosystems**. *Cell* 2012, **149**:1164-1173.
In this ground-breaking study, the dynamics of three interacting species is monitored with an array of closed ecosystems, permitting the authors to compare changes in species abundance across repetitions and across time. Remarkably, they found that replicate measurements at a single time point can be used to infer species interactions that are also observed in long-term time series.

15. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA: **The application of ecological theory toward an understanding of the human microbiome**. *Science* 2012:336.

16. Shade A, Read JS, Youngblut ND, Fierer N, Knight R, Kratz TK, Lottig NR, Roden EE, Stanley EH, Stombaugh J *et al.*: **Lake microbial communities are resilient after a whole-ecosystem disturbance**. *ISME J* 2012, **6**:2153-2167.

17. Dethlefsen L, Relman DA: **Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation**. *Proc Natl Acad Sci U S A* 2011, **108**:4554-4561.

18. Pérez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, Otto W, Rojo D, Bargiela R, Bergen M *et al.*: **Gut microbiota disturbance during antibiotic therapy: a multi-omic approach**. *Gut* 2013, **62**:1591-1601.

19. •• David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, Erdman SE, Alm EJ: **Host lifestyle affects human microbiota on daily timescales**. *Genome Biol* 2014, **15**:R89.

This study offers one of the longest, densely sampled metagenomic time series available to date, including two subjects sampled almost daily for a year. The special interest of this time series is the presence of perturbations, in the form of travel and diet change for subject A and food poisoning for subject B. In both subjects, the gut microbiota approaches its original state upon return to home (subject A) or recovery from enteric infection (subject B). This study also pioneers the application of a number of time series analysis techniques to microbiota, such as the Augmented Dickey Fuller test to quantify community stability.

20. Hartman AL, Lough DM, Barupal DK, Fiehn O, Fishbein T, Zasloff M, Eisen JA: **Human gut microbiome adopts an alternative state following small bowel transplantation**. *Proc Natl Acad Sci U S A* 2009, **106**:17187-17192.

21. Koch C, Fetzer I, Schmidt T, Harms H, Müller S: **Monitoring functions in managed microbial systems by cytometric bar coding**. *Environ Sci Technol* 2013, **47**:1753-1760.

22. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA *et al.*: **Diet rapidly and reproducibly alters the human gut microbiome**. *Nature* 2013, **505**:559-563.

23. Salonen A, Lahti L, Salojärvi J, Holtrop G, Korpela K, Duncan SH, Date P, Farquharson F, Johnstone AM, Lobley GE *et al.*: **Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men**. *ISME J* 2014, **8**:2218-2230.

24. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Nostrand JDV, Yang Y, He Z, Wu L, Stahle DA *et al.*: **Stochasticity, succession, and environmental perturbations in a fluidic ecosystem**. *Proc Natl Acad Sci U S A* 2013, **111**:E836-E845.

25. Andersson M.G.I., Berga M, Lindström ES, Langenheder S: **The spatial structure of bacterial communities is influenced by historical environmental conditions**. *Ecology* 2014, **95**:1134-1140.

26. Faust K, Raes J: **Microbial interactions: from networks to models**. *Nat Rev Microbiol* 2012, **10**:538-550.

27. Barberán A, Bates ST, Casamayor EO, Fierer N: **Using network analysis to explore co-occurrence patterns in soil microbial communities**. *ISME J* 2012, **6**:343-351.

28. Chaffron S, Rehrauer H, Pernthaler J, Von Mering C: **A global network of coexisting microbes from environmental and whole-genome sequence data**. *Genome Res* 2010, **20**:947-959.

29. Faust S, Izard J, Segata N, Gevers D, Raes J, Huttenhower C: **Microbial co-occurrence relationships in the human microbiome**. *PLoS Comput Biol* 2012, **8**:e1002606.

30. Friedman J, Alm EJ: **Inferring correlation networks from genomic survey data**. *PLoS Comput Biol* 2012, **8**:e1002687.

31. van den Bergh MR, Biesbroek G, Rossen JWA, de Steenhuijsen Piters WAA, Bosch AATM, van Gils EJM, Wang X, Boonacker CWB, Veenhoven RH, Bruin JP *et al.*: **Associations between pathogens in the upper respiratory tract of young children: interplay between viruses and bacteria**. *PLOS ONE* 2012, **7**:e47711.

32. Trosvik P, Stenseth NC, Rudi K: **Convergent temporal dynamics of the human infant gut microbiota**. *ISME J* 2010, **4**:151-158.

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

34. Marino S, Baxter NT, Huffnagle GB, Petrosino JF, Schloss PD: **Mathematical modeling of primary succession of murine intestinal microbiota**. *Proc Natl Acad Sci U S A* 2013, **111**:439-444.

35. Mounier J, Monnet C, Vallaeys T, Arditi R, Sarthou A-S, Hélias A, Irlinger F: **Microbial interactions within a cheese microbial community**. *Appl Environ Microbiol* 2008, **74**:172-181.

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

37. Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F: **Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors**. *Bioinformatics* 2006, **22**:2532-2538.

38. • Chow C-ET, Kim DY, Sachdeva R, Caron DA, Fuhrman JA: **Top-down controls on bacterial community structure: microbial network analysis of bacteria T4-like viruses and protists**. *ISME J* 2014, **8**:816-829.

In this study, ecological interactions between microorganisms from the San Pedro Ocean time series are predicted with the LSA network inference algorithm. The study suggests that viral–bacterial interactions are on average more specific than protistan–bacterial interactions.

39. Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow C-ET, Sachdeva R, Jones AC, Schwalbach MS *et al.*: **Marine bacterial, archaeal and protistan association networks reveal ecological linkages**. *ISME J* 2011, **5**:1414-1425.

40. Friedman N, Murphy K, Russell S: **Learning the structure of dynamic probabilistic networks**. *Proc. Uncertainty in Artificial Intelligence (UAI)*. Edited by Publishers MK; 1998:139-147.

41. Kim S, Imoto S, Miyano S: **Inferring gene networks from time series microarray data using dynamic Bayesian networks**. *Brief Bioinform* 2003, **4**:228-235.

42. Granger CWJ: **Investigating causal relations by econometric models and cross-spectral methods**. *Econometrica* 1969, **37**:424-438.

43. Sugihara G, May R, Ye H, Hsieh C-h, Deyle E, Fogarty M, Munch S: **Detecting causality in complex ecosystems**. *Science* 2012, **338**:496-500.

44. Robinson J, Hartemink A: **Non-stationary dynamic Bayesian networks**. *Neural Information Processing Systems (NIPS)*. 2008:1372-1379.

45. Song L, Kolar M, Xing E: **Time-varying dynamic Bayesian networks**. *Neural Information Processing Systems (NIPS)*. 2009:1732-1739.

46. Mandelbrot BB: *The (Mis)Behavior of Markets, A Fractal View of Risk, Ruin and Reward*. 2004.

47. •• Benincà E, Huisman J, Heerkloss R, Jöhnk KD, Branco P, van Nes EH, Scheffer M, Ellner SP: **Chaos in a long-term experiment with a plankton community**. *Nature* 2008, **451**:822-825.

This is the first demonstration of chaos in a long-term mesocosm experiment with a marine microbial food web. The mesocosm was sampled twice weekly for over six years. Besides demonstrating chaos, the article features a predictability analysis that compares the forecasting accuracy of a linear with that of a non-linear model.

48. Dakos V, Benincà E, van Nes EH, Philippart CJM, Scheffer M, Huisman J: **Interannual variability in species composition explained as seasonally entrained chaos**. *Proc R Soc B* 2009, **276**:2871-2880.

49. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM *et al.*: **Enterotypes of the human gut microbiome**. *Nature* 2011, **473**:174-180.

50. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R *et al.*: **Linking long-term dietary patterns with gut microbial enterotypes**. *Science* 2011, **334**:105-108.

51. Knights D, Ward TL, McKinlay CE, Miller H, Gonzalez A, McDonald D, Knight R: **Rethinking enterotypes**. *Cell Host Microbe* 2014, **16**:433-437.

52. Lahti L, Salojärvi J, Salonen A, Scheffer M, de Vos WM: **Tipping elements in the human intestinal ecosystem**. *Nat Commun* 2014, **5**:4344.

53. •• Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, Koenig SSK, Fu L, Ma ZS, Zhou X *et al.*: **Temporal dynamics of the human vaginal microbiota**. *Sci Transl Med* 2012, **4**:132ra152.

This landmark study, which monitors the vaginal microbiota dynamics in 32 healthy women sampled twice weekly for 16 weeks, is one of a few metagenomic time series studies that combines a relatively large number of subjects with a long and dense sampling period. The study demonstrates that some women frequently switch between community states, whereas others have more stable vaginal microbiota. Interestingly, state transitions are not always linked to menstruation or sexual activity. The article pioneers the visualization of a microbiota's trajectory in community space as well as the computation of microbial community state transition probabilities.

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

55. •• Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, van de Koppel J, van de Leemput IA, Levin SA, van Nes EH *et al.*: **Anticipating critical transitions**. *Science* 2012, **338**:344-348.
An excellent overview bridging the gap between the analysis of network structure and alternative stable states. The article describes, with examples from population ecology, financial market and climatology, how fundamental architectural features in ecological networks could create unstable tipping points where a small perturbation can trigger a drastic state shift. The article discusses how to detect early warning signals to anticipate such critical transitions based on time series analysis.

56. Wang R, Dearing JA, Langdon PG, Zhang E, Yang X, Dakos V, Scheffer M: **Flickering gives early warning signals of a critical transition to a eutrophic lake state**. *Nature* 2012, **492**:419-422.

57. • Gerber GK: **The dynamic microbiome**. *FEBS Lett* 2014, **588**:4131-4139.
A comprehensive review on human microbiome time series studies, highlighting pitfalls in time series analysis and presenting cross-prediction methods and the application of the generalized Lotka–Volterra model in more depth.

58. Fuhrman JA, Cram JA, Needham DM: **Marine microbial community dynamics and their ecological interpretation**. *Nat Rev Microbiol* 2015, **13**:133-146.

59. Widder S, Besemer K, Singer GA, Ceola S, Bertuzzo E, Quince C, Sloan WT, Rinaldo A, Battin TJ: **Fluvial network organization imprints on microbial co-occurrence networks**. *Proc Natl Acad Sci U S A* 2014, **111**:12799-12804.

60. Hsieh C-H, Anderson C, Sugihara G: **Extending nonlinear analysis to short ecological time series**. *Am Nat* 2008, **171**:71-80.

61. Hsieh C-H, Glaser SM, Lucas AJ, Sugihara G: **Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean**. *Nature* 2005, **435**:336-339.

62. Xia LC, Ai D, Cram J, Fuhrman JA, Sun F: **Efficient statistical significance approximation for local association analysis of high-throughput time series data**. *Bioinformatics* 2013, **29**:230-237.

63. Xia LC, Steele JA, Cram JA, Cardon ZG, Simmons SL, Vallino JJ, Fuhrman JA, Sun F: **Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates**. *BMC Syst Biol* 2011, **5**:S15.

64. Durno WE, Hanson NW, Konwar KM, Hallam SJ: **Expanding the boundaries of local similarity analysis**. *BMC Genomics* 2013, **14**:S3.

65. Deng Y, Jiang Y, Yang Y, He Z, Luo F, Zhou J: **Molecular ecological network analyses**. *BMC Bioinformatics* 2012, **13**:113.

66. Trifonova N, Duplisea D, Kenny A, Tucker A: **A Spatio-temporal Bayesian Network Approach for Revealing Functional Ecological Networks in Fisheries**. *Advances in Intelligent Data Analysis XIII*. **Lecture Notes in Computer Science**. 2014:298-308.

67. Hegger R, Kantz H, Schreiber T: **Practical implementation of nonlinear time series methods: the TISEAN package**. *Chaos* 1999, **9**:413-435.

68. Dakos V, Carpenter SR, Brock WA, Ellison AM, Guttal V, Ives AR, Kéfi S, Livina V, Seekell DA, van Nes EH *et al.*: **Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data**. *PLOS ONE* 2012, **7**:e41010.