White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/102537/

Version: Published Version

Article:
Chong, James Paul Jonathan, Widder, Stefanie, Allen, Rosalind J. et al. (2016) Challenges in microbial ecology: building predictive understanding of community function and dynamics. The ISME Journal. ISSN: 1751-7362

https://doi.org/10.1038/ismej.2016.45

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

OPEN

The ISME Journal (2016), 1–12
© 2016 International Society for Microbial Ecology All rights reserved 1751-7362/16
www.nature.com/ismej

MINI REVIEW

Challenges in microbial ecology: building predictive understanding of community function and dynamics

Stefanie Widder[1], Rosalind J Allen[2], Thomas Pfeiffer[3], Thomas P Curtis[4], Carsten Wiuf[5], William T Sloan[6], Otto X Cordero[7], Sam P Brown[8], Babak Momeni[9,10], Wenying Shou[10], Helen Kettle[11], Harry J Flint[12], Andreas F Haas[13], Béatrice Laroche[14], Jan-Ulrich Kreft[15], Paul B Rainey[3], Shiri Freilich[16], Stefan Schuster[17], Kim Milferstedt[18], Jan R van der Meer[19], Tobias Groβkopf[20], Jef Huisman[21], Andrew Free[22], Cristian Picioreanu[23], Christopher Quince[24], Isaac Klapper[25], Simon Labarthe[14], Barth F Smets[26], Harris Wang[27], Isaac Newton Institute Fellows[28] and Orkun S Soyer[20]

[1]*CUBE, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria;* [2]*SUPA, School of Physics and Astronomy, University of Edinburgh, Edinburgh, UK;* [3]*New Zealand Institute for Advanced Study, Massey University, Auckland, New Zealand;* [4]*School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne, UK;* [5]*Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark;* [6]*Infrastructure and Environment Research Division, School of Engineering, University of Glasgow, Glasgow, UK;* [7]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA;* [8]*Centre for Immunity, Infection and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh, UK;* [9]*Department of Biology, Boston College, Chestnut Hill, MA, USA;* [10]*Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;* [11]*Biomathematics and Statistics Scotland, Edinburgh, UK;* [12]*Rowett Institute of Nutrition and Health, University of Aberdeen, Aberdeen, UK;* [13]*Biology Department, San Diego State University, San Diego, CA, USA;* [14]*Département de Mathématiques Informatiques Appliquées, INRA, Jouy-en-Josas, France;* [15]*School of Biosciences, University of Birmingham, Birmingham, UK;* [16]*Newe Ya'ar Research Center, Agricultural Research Organization, Ramat Yishay, Israel;* [17]*Department of Bioinformatics, Friedrich-Schiller-University Jena, Jena, Germany;* [18]*INRA, UR0050, Laboratoire de Biotechnologie de l'Environnement, Narbonne, France;* [19]*Department of Fundamental Microbiology, Université de Lausanne, Lausanne, Switzerland;* [20]*School of Life Sciences, The University of Warwick, Coventry, UK;* [21]*Department of Aquatic Microbiology, University of Amsterdam, Amsterdam, The Netherlands;* [22]*Institute of Quantitative Biology, Biochemistry and Biotechnology, School of Biological Science, University of Edinburgh, Edinburgh, UK;* [23]*Department of Biotechnology, Delft University of Technology, Delft, The Netherlands;* [24]*Warwick Medical School, University of Warwick, Coventry, UK;* [25]*Department of Mathematics, Temple University, Philadelphia, PA, USA;* [26]*Department of Environmental Engineering, Technical University of Denmark, Kongens Lyngby, Denmark and* [27]*Department of Systems Biology, Columbia University, New York, NY, USA*

The importance of microbial communities (MCs) cannot be overstated. MCs underpin the biogeochemical cycles of the earth's soil, oceans and the atmosphere, and perform ecosystem functions that impact plants, animals and humans. Yet our ability to predict and manage the function of these highly complex, dynamically changing communities is limited. Building predictive models that link MC composition to function is a key emerging challenge in microbial ecology. Here, we argue that addressing this challenge requires close coordination of experimental data collection and method development with mathematical model building. We discuss specific examples where model–experiment integration has already resulted in important insights into MC function and structure. We also highlight key research questions that still demand better integration of experiments and models. We argue that such integration is needed to achieve significant progress in our understanding of MC dynamics and function, and we make specific practical suggestions as to how this could be achieved.
The ISME Journal advance online publication, 29 March 2016; doi:10.1038/ismej.2016.45

Correspondence: RJ Allen, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK.
E-mail: rallen2@staffmail.ed.ac.uk
or OS Soyer, School of Life Sciences, The University of Warwick, Coventry CV4 7AL, UK.
E-mail: O.Soyer@warwick.ac.uk
[28]Members of the Isaac Newton Institute Fellows are listed before References.
The manuscript resulted from a 4-month residential research program organised at the Isaac Newton Institute, Cambridge, UK. The aim of the research program was to bring theoretical and experimental scientists interested in microbial communities together through discussions, seminars and workshops. The manuscript summarizes the ideas that emerged from this program and was developed by the entire group, led by specific individuals. All listed authors contributed to the development of the ideas presented. Those listed individually made significant contributions to the writing of the manuscript.
Received 20 November 2015; revised 12 February 2016; accepted 22 February 2016

Introduction

Microbes exist in complex, highly diverse and highly dynamic communities (Flint *et al.*, 2007; Johnson *et al.*, 2015). These microbial communities (MCs) have crucial roles in global climate regulation, human health and industrial biotechnology. Understanding, predicting and controlling MCs thus holds the key to a wealth of potential applications, from smart waste treatment plants, through probiotic treatments of gut-related diseases, to cheese and wine making (Shong *et al.*, 2012; Johnson *et al.*, 2015; Wolfe and Dutton, 2015). The need to predict MC dynamics has stimulated the development of 'black box' mathematical models, in which microbial population dynamics is represented by global empirical functions that do not attempt to address the inner workings of the MC. This approach has proved useful in fields ranging from food science to climate modelling and wastewater treatment (Orhon and Artan, 1994; Baranyi and Tamplin, 2004). However, such models do not aim to provide mechanistic insight, and are necessarily limited by the data sets to which they are fitted.

High-throughput sequencing, proteomics and metabolomics now allow us to catalogue the diversity of MCs to an unprecedented level of detail. These data represent a relatively unbiased compositional snapshot of the species, genes, metabolites and activities that are present in a given MC. The key challenge now is to convert this empirical knowledge into fundamental insights and testable predictions, which can be used to improve MC function for useful purposes. Here, we argue that addressing this challenge will require the development of mathematical models with a basis in mechanistic understanding, integrated with controlled experiments (Figure 1). In our view, this integration between theory and experiments is a crucial 'missing link' in current microbial ecology. Making this link is a key step on the way to discovering possible design principles of MC community assembly and function.

MCs as complex, interacting dynamical systems

The dynamics of MCs are driven by a multitude of interactions between their constituent microbial populations, as well as by environmental and host factors, such as immunological processes in gut microbiota or nutrient limitation in plant MCs (Klitgord and Segrè, 2010; Vorholt, 2012; Coyte *et al.*, 2015). Species interactions within MCs can be metabolic, physical, regulatory and/or signalling based, and they can drive both temporal changes in MC composition and function, and spatial organisation. Phenomenologically, an interaction between two microbial populations can be defined as the dependence of one population's growth or survival on the abundance of the other population. These interactions can be negative or positive, implying growth inhibition or facilitation. Negative interactions can arise from competition for resources such as electron donors and acceptors, nutrients, light or

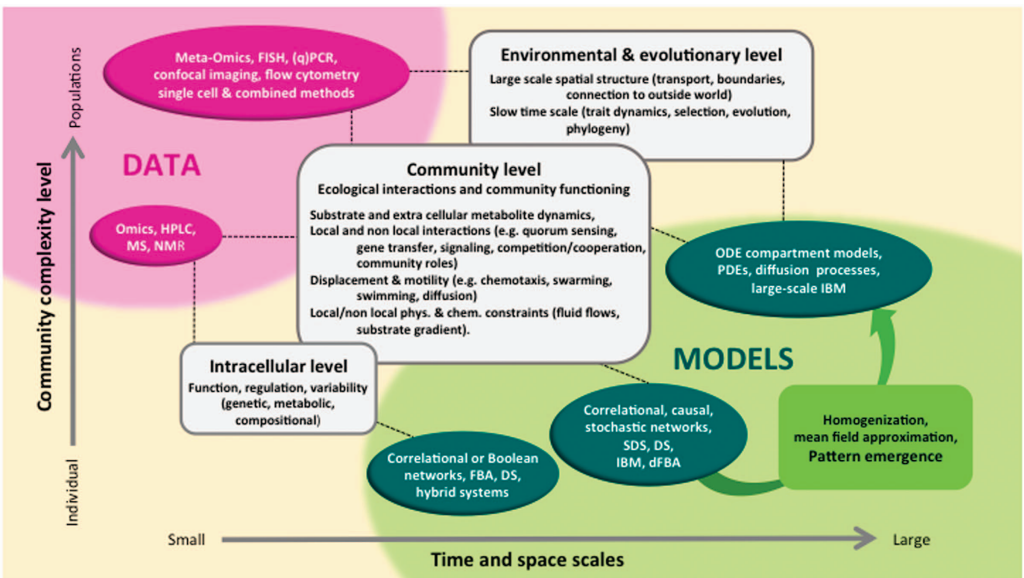

Figure 1 Linking MC research questions with data and modelling. Research areas, plotted according to their complexity and temporal or spatial scale, form the link between data of different forms (magenta) and modelling formalisms (dark green). Pattern emergence (light green), that is, collective behaviour obtained by up-scaling from individual description to population level, can be predicted by modelling and tested experimentally. Abbreviations: DS, dynamical systems of deterministic, mechanistic nature, implemented as discrete or continuous time models (difference equations, ODEs); (d) FBA, (dynamic) flux balance analysis; SDS, stochastic dynamical systems, such as Markov chains, random walks, birth–death processes; IBM, individual-based models; PDE, partial differential equations, these are deterministic structured population models (for example, according to space or traits); diffusion processes, probabilistic counterparts of ODEs and PDEs.

physical space (Hibbing *et al.*, 2010), or via direct microbe–microbe interactions, or secretion of toxins or other inhibitory compounds (Riley and Wertz, 2002). Negative interactions have been observed in evolution and competition experiments with single or several species (Passarge *et al.*, 2006; Brown *et al.*, 2008), and are also widespread among different species in natural populations, such as soil and marine bacteria (Vetsigian *et al.*, 2011; Cordero *et al.*, 2012b). Positive interactions between microbial populations can occur via several mechanisms. From a metabolic standpoint, cross-feeding of metabolic by-products, in which one population benefits from the excreted metabolites of another, is a key mediator of positive interactions (Sieber *et al.*, 2012); this also reduces the degree to which populations compete for resources (Kosaka *et al.*, 2008). Other mechanisms include production of 'public goods' such as iron-scavenging molecules, which can be used not only by the producer population but also by others (West *et al.*, 2006).

Identification and measurement of interactions within MCs, and accurate representation of these in theoretical models, is an important basis for building understanding, and thus presents a key challenge in microbial ecology. A second challenge is the converse: to use theoretical approaches to predict species interactions in MCs from proximal data such as taxon abundances. Both cases require direct integration of theory and data in specific ways. Although many advances have been made toward such integration (for examples, see Figures 2 and 3), a number of technical obstacles remain. In the following, we discuss key areas where integration theory and data can be highly productive. We then highlight a number of broader scientific questions, which we see as crucial for the longer-term development of models for MC structure, function and dynamics. Finally, we call for development of model systems where well-controlled experiments interrogating function–structure relation in communities can be more readily performed and information and methods among multiple laboratories can be shared.

Integrating theory and data to understand MC dynamics

Measuring population dynamics in MCs

A key objective of theoretical models for MC dynamics is to reproduce temporal trajectories of the populations within the community. However, state of the art high-throughput sequencing generates a snapshot of the *relative* abundances of taxa or genes within a MC; it does not provide *absolute* abundance information. Furthermore, there can be inherent biases even in this relative abundance data. These limitations mean that high-throughput sequencing on its own is not sufficient to track temporal and spatial population dynamics. Complementary experimental approaches such as quantitative PCR, flow cytometry, species-specific fluorescence *in situ* hybridisation or novel combinations of single-cell and functional-targeting methods with genomics (Berry *et al.*, 2015) do yield information on absolute abundances. Yet, this information is limited, because these approaches are usually targeted, that is, need *a priori* knowledge of which populations or functions are to be investigated. Although some of these approaches can be used in an untargeted manner, for example, by using general primers for fluorescence *in situ* hybridisation or by combining single-cell separation with genome sequencing, this involves technical challenges. Upscaling these absolute abundance measurements to

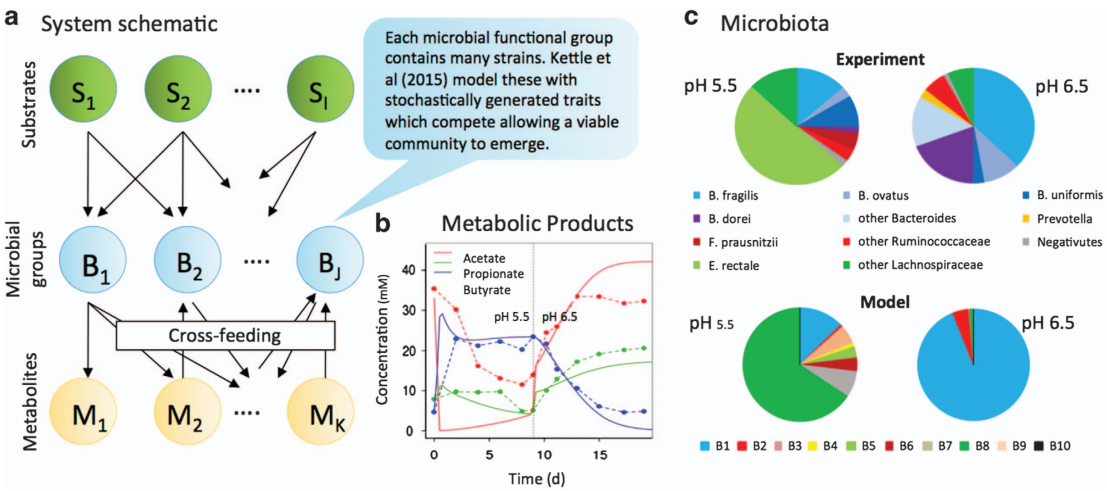

Figure 2 (**a**) The human large intestine and the rumen and caecum in herbivorous animals harbour dense MCs dominated by anaerobic microbes that cross-feed metabolites extensively. In recent models, these are approximated by a small number of functional groups (Muñoz-Tamayo *et al.*, 2010; Kettle *et al.*, 2015). (**b**) Comparison of this model to a fermenter experiment with a pH shift from 5.5. to 6.5 after 9 days for metabolic products (dashed lines are experiment data, solid lines are model results). (**c**) Comparison of temporal species dynamics between model predictions and data. The experimental data consists of phylogenetic groups (16S rRNA gene sequencing), simulations refer to functional groups; approximate correspondence between the two is indicated by colour coding (for example, *Lachnospiraceae* equivalent to B5 and B8, shown in green, *Bacteroides* belonging to B1 shown in blue and purple).

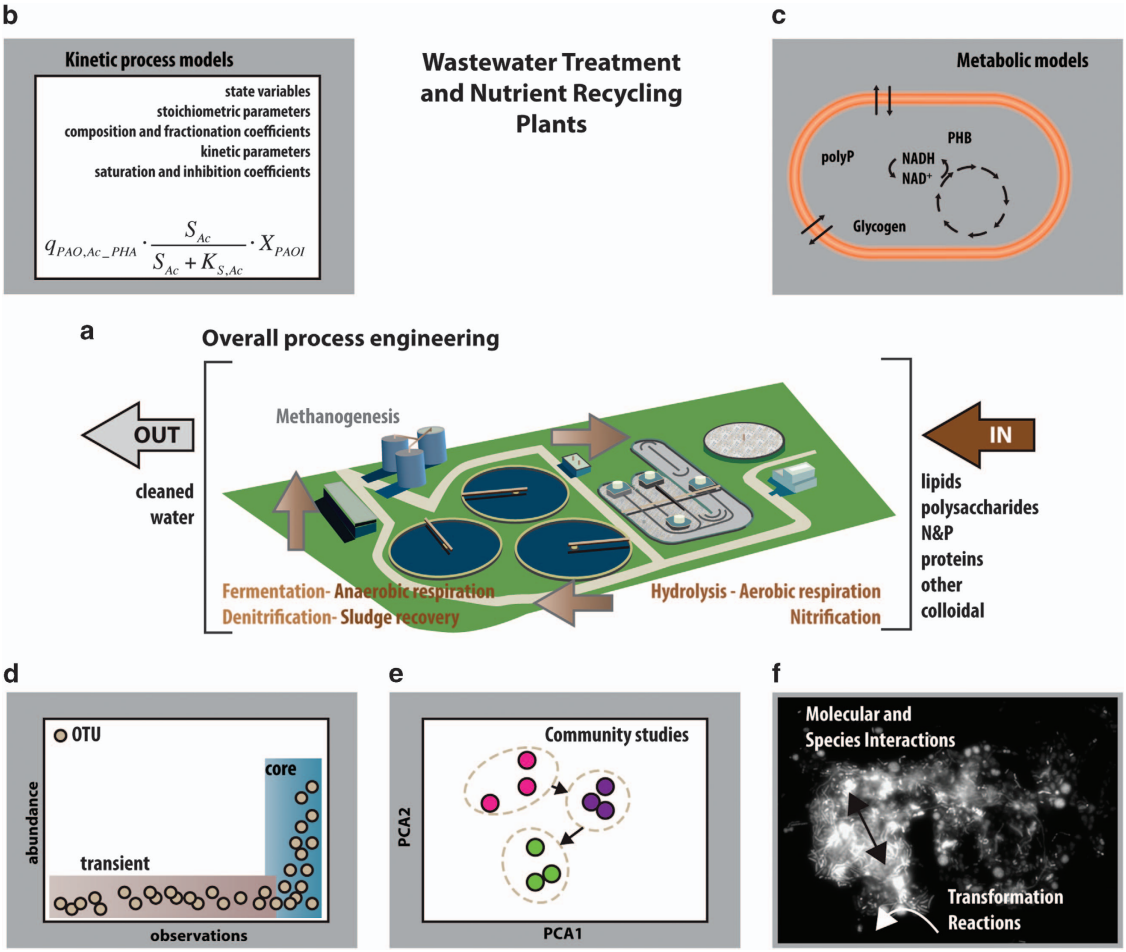

Figure 3 MC analysis and predictive modelling in wastewater treatment and nutrient recycling facilities (WWTP). (**a**) Carbon and nutrients are eliminated from wastewater by (micro-)biological activities before discharging the cleaned water. WWTP processes are optimised through controlling MC conditions (for example, anaerobic, aerobic) and partial recycling of the MC biomass. (**b**) Overall system behaviour and MC biomass are predicted using growth kinetic and dynamic models. (**c**) Functional metabolic models are derived from pure culture physiology or metagenomic data. (**d**) Although the WWTP system is largely engineered, the MCs form mostly spontaneous, consisting of 'core' assemblages and 'passenger' groups as a function of constant wastewater input, after (Saunders *et al*., 2016). (**e**) The assemblage process depends on taxonomic relatedness and ecological interactions, and can lead to a drift of MCs over time. (**f**) MCs in WWTP occur mostly in flocs and or biofilms (picture courtesy of J van der Meer, University of Lausanne) that display immense microdiversity. Both, interactions and spatial organisation are factors that generate niches, gradients and foster co-existence.

the community level is an important goal. We note that combining high-throughput sequencing with quantitative PCR analysis to measure total bacterial abundance should, in principle, allow absolute taxon abundances to be computed. Yet, we are not aware of studies that have used this approach to date.

We also note that the potential of sequencing approaches to predict aggregate community function is currently limited, because neither metagenomic data nor 16S RNA gene sequence data are a perfect predictor of metabolic function for an individual species. This is an area where new sequencing and bioinformatics approaches can result in significant improvement.

Inferring species interactions from proximal data

Direct measurement of interactions between species within a community (for example, for metabolic interactions, by tracking radioactively or isotopically labelled compounds) is an excellent way to collect the basic information needed for model building, but it is necessarily targeted to specific types of interaction. To obtain community-wide information on microbial interactions where direct evidence is lacking or restricted, one can use statistical inference based on correlations between taxon abundances from high-throughput sequence data (Fuhrman, 2009; Freilich *et al*., 2010; Faust *et al*., 2012). The resulting co-occurrence interaction networks can be used to make ecological predictions, such as the existence of metabolic dependencies, the fragility of communities to environmental change or the likely identity of keystone species that stabilise the community (Faust *et al*., 2012; Berry and Widder, 2014; Widder *et al*., 2014). Although correlational approaches can point to previously unknown interactions, they require confirmation by direct evidence

(Friedman and Alm, 2012), because correlations in taxon abundances can be confounded by biases in 16S data (Zhou *et al.*, 2011) and can also arise from indirect species interactions. Examples of such indirect interactions include apparent competition by shared pathogens or predators, or the immunological response of a shared host (Brown *et al.*, 2008; McNally *et al.*, 2014). As these indirect interactions typically occur on larger spatial scales than direct interactions, development of spatially resolved methods for interrogating MCs can be highly beneficial. Existing methods such as fluorescence *in situ* hybridisation provide proximity data among species, but are invasive. Recent advances in non-invasive spatially resolved methods, such as in-line digital microscopy, for gathering population dynamics data are thus of great interest (Frentz *et al.*, 2010).

Predicting species interactions using stoichiometric models

Many of the important interactions in MCs are mediated by metabolite exchange; thus computational approaches modelling metabolic fluxes between organisms can be used to predict interactions. These approaches are based on knowledge of the genome, which is used to build a 'stoichiometric model' for the complete set of metabolic reactions associated with a given organism (Schilling *et al.*, 1999). Classical flux balance analysis (FBA) (Orth *et al.*, 2010) can then be used to predict the organism's metabolic fluxes. Up-scaling this approach to construct community-wide stoichiometric models or performing detailed metabolic studies of individual participants within a community (Freilich *et al.*, 2011) allows for insights into metabolic interactions occurring within a MC (Stolyar *et al.*, 2007). For example, metabolic studies have shed light on the succession of primary and secondary degraders that ensures communal accessibility of complex carbohydrates in the mammalian gut (Muñoz-Tamayo *et al.*, 2010; Eilam *et al.*, 2014; Kettle *et al.*, 2015) (Figure 2). Moreover, dynamical FBA can also form a basis for dynamical models of MCs (Mahadevan *et al.*, 2002). In these models, microbial growth is coupled to a dynamically changing chemical environment, which is in turn influenced by metabolism, as computed by FBA. Recently, such dynamical models have also been extended to include spatial resolution of community structure (Harcombe *et al.*, 2014). These developments are very promising, but it is important to note that the application of FBA to MCs is challenging. In particular, standardised methods are needed to generate reliable stoichiometric models of the large number of species involved in MCs and to be able to integrate models so that they can be simulated under a single, shared environment.

Kinetic models for community dynamics

Kinetic models, such as the widely used Monod equation, aim to predict microbial growth rate, given the concentrations of essential nutrients and/or inhibitory chemicals, and species-dependent parameters such as the maximal growth rate (Figure 3). Extending this approach to the community level is attractive because it is conceptually simple, computationally tractable and provides dynamical predictions. However, its success depends on identification of the most important microbial species to include in the model, and the interactions between them. Moreover, the reliability of the approach depends crucially on the underlying kinetic growth model and its parameters—thus large-scale measurement of kinetic growth parameters for a variety of microbial species would be extremely useful. More fundamentally, existing kinetic growth models are often '*ad hoc*' and may miss aspects of the growth kinetics that are important in community function, such as product inhibition. To address this problem, new kinetic models are being developed that include features like condition-dependent changes in the maximal growth rate (Bonachela *et al.*, 2011; Desmond-Le Quemener and Bouchez, 2014) and the thermodynamics of microbial metabolism (Jin *et al.*, 2013). In particular, the inclusion of thermodynamics in kinetic growth models has contributed to improved modelling of MCs, for example in anaerobic waste treatment (Gonzalez-Cabaleiro *et al.*, 2013).

Long-term challenges for developing an understanding of structure–function relation and dynamics in MCs

The need to include evolutionary processes

Traditionally, evolution has been ignored in ecological models, as it is assumed to occur only on long timescales. While this assumption might hold for animal and plant communities, ecological and evolutionary timescales can coincide in some MCs, because of their short generation times, large population sizes and high rates of gene transfer. Thus, models that address MC dynamics should take into account both ecology and evolution, for example, see (Post and Palkovacs, 2009). The intermixing of ecological and evolutionary dynamics is strikingly demonstrated by long-term growth experiments with *Escherichia coli* in glucose-limited chemostats. Here, genetic diversification produces two genotypes, which cross-feed; that is, it leads to a new ecological interaction. Functional diversification at the genome level is maintained by the novel ecological interaction and vice versa (Little *et al.*, 2008). As ecological interactions evolve in MCs on the same timescales as the species themselves evolve (Cordero *et al.*, 2012a; Hillesland *et al.*, 2014), the development of modelling frameworks that include evolution of species' traits and interactions should have an important role in microbial ecology. From a modelling perspective, simplified 'toy-models' have been developed, which include ecology and evolution, for example, in (Pfeiffer *et al.*, 2001). The challenge is to make these

more realistic (for example, including more aspects of cellular metabolism). From an experimental perspective, attempts are being made to follow the evolution of selected species, for example, in natural soil MCs (Gomez and Buckling, 2013); the next step is to track more species simultaneously, and to extend to other environments, such as gut communities (for example, using axenic animal systems). An alternative approach is to use defined synthetic communities, constructed from known species, in which it is possible to track the ecological and evolutionary dynamics of individual species within the community (see for example, Mee and Wang, 2012; Bodenhausen *et al*., 2014; Faith *et al*., 2014).

Social evolution and bacterial strategies

An important challenge in model development is to correctly account for the complex social organisation that can occur amongst microbes and, for experimentalists, to test how frequently social interactions occur in natural settings. For example, secretion of 'public goods' such as toxins, enzymes, metabolic cofactors or signalling molecules can lead to intricate evolutionary dynamics (Leggett *et al*., 2014). Game theory provides a way to model complex social behaviours in mixed MCs. Here, different microbial behaviours are abstracted as simple 'strategies'. As an example, some microbial populations produce extracellular enzymes such as cellulases that digest recalcitrant nutrients in the environment. 'Cheater' cells do not produce these enzymes, but nevertheless benefit from their release by 'cooperative' donor cells. Game theory maps this scenario on to a number of classic 'games', such as Prisoner's Dilemma, the hawk-dove game or the mutual benefit game, depending on the values of the kinetic parameters (Gore *et al*., 2009). Extending such game theoretical models to include MC spatial structure remains a challenging area, although some important advances have been made using lattice-based simulations and continuum partial differential equation approaches (Reichenbach *et al*., 2007, 2008).

Community assembly and historical contingency

Community assembly, or the mechanism by which a community forms, is a widely studied topic in macro-ecology, but has been relatively little addressed for MCs (Woodcock *et al*., 2007). For some MCs it is known, however, that historical contingency—the order in which different species arrive in the community—can have a strong impact on community composition; examples include oral communities (Teles *et al*., 2012) and gut microbiota (David *et al*., 2014). Importantly, different patterns of species arrival can result in different long-term interaction networks within the community (Vannette and Fukami 2014). These historical contingency effects are likely to have important consequences for the engineering and control of MCs in the environment, agriculture and medicine. Future work should systematically investigate these effects experimentally for complex MCs and, concurrently, integrate them into population dynamic models. Moreover revisiting suitably designed, older experiments with new methods may also contribute to understanding temporal processes of community assembly in MCs.

The importance of spatial structure

Another important feature of MCs is their complex spatial structure (Figure 4). Indeed, densely packed aggregates, which may be free-floating or in the form of surface-attached biofilms, are believed to be the predominant mode of life for many microbes in the natural environment, for example, MCs on marine snow particles (Kiorboe *et al*., 2003; Elias and Banin, 2012), and are crucial in processes such as wastewater treatment (Figure 3). Within these microbial aggregates, driving factors for spatial organisation include (i) metabolite gradients caused by consumption/production, diffusion and advection, (ii) gradients of abiotic factors such as light or temperature, (iii) physical adhesion and (iv) motility. A number of well-established methods exist for modelling spatial structure development within aggregated MCs (especially biofilms). Continuum spatial models predict how microbial biomass density and chemical concentrations change in space and time and typically include diffusion, advection and mechanical forces (Klapper and Dockery, 2010). At a more detailed level, individual-based models track the location and fate of individual microbial cells within the community, taking into account a plethora of features such as spatially resolved metabolite concentrations or electrochemical interactions with a surface (for example, an electrode). The use, and further development, of such spatially explicit models is important because spatial organisation of aggregated MCs is likely to have a drastic effect on their structure and function (Momeni *et al*., 2013). For example, biofilm infections are notoriously more resistant to antibiotic treatment than well-mixed planktonic cultures (Davies, 2003). Such improvements in modelling spatially structured MCs must go hand-in-hand with improvements in experimental interrogation of spatial structure within natural settings, and in particular development of non-invasive measurement methods.

A call for the development of model MCs and well-controlled experiments on them

In microbiology, fundamental understanding of microbial physiology and metabolism has been acquired by studying a set of standard microbial model organisms, for example, *Saccharomyces cerevisiae*, *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas aeruginosa*, under well-defined conditions.

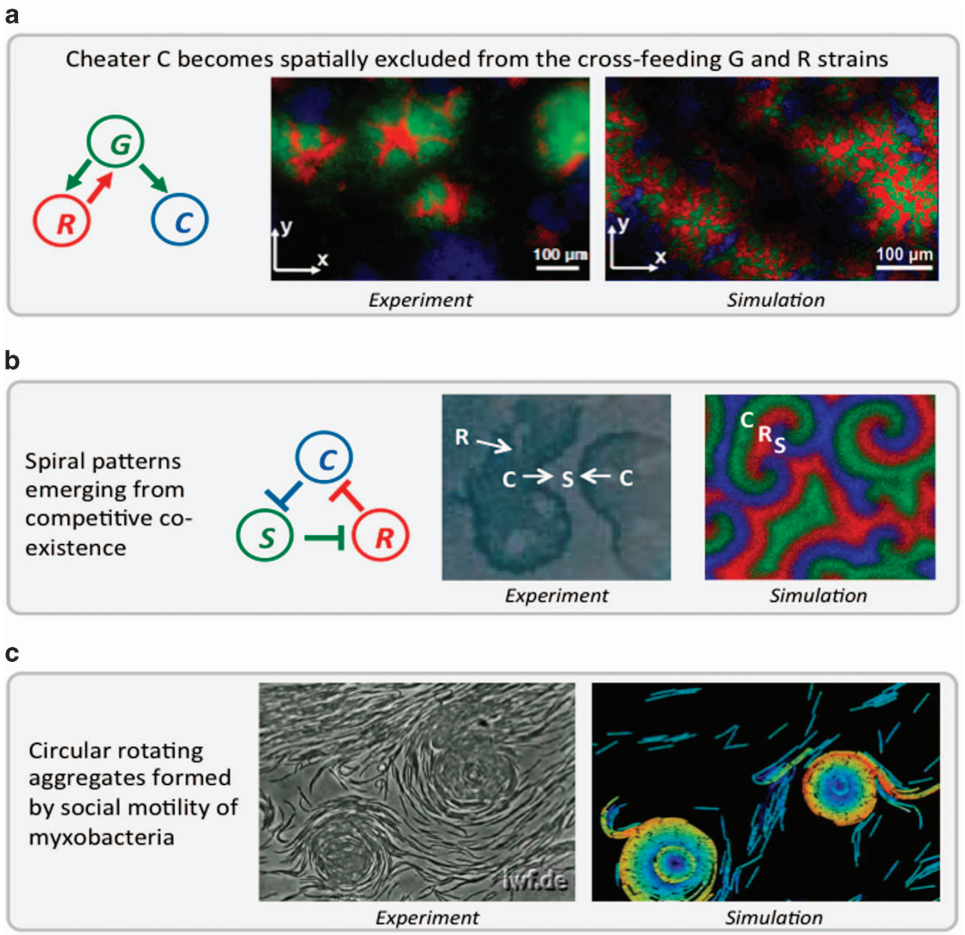

Figure 4 Spatial patterns can emerge from local metabolic and mechanical interactions between individual microbes. Such self-organised patterns facilitate, for example, cooperation (**a**), competitive co-existence (**b**) or formation of fruiting bodies (**c**). (**a**) Cross-feeding between G (green) and R (red) strains facilitates exclusion of cheater C (blue) over time. Cross-feeding was engineered in yeast strains and visualised by fluorescence tagging, adapted from (Momeni *et al.*, 2013). (**b**) Spiral patterns emerge from chasing in space: Colicin producer C kills sensitive S. S outcompetes resistant R, which outcompetes C in turn. Spatial structure facilitates dynamic co-existence. Experimental results on agar plates adapted from (Kerr *et al.*, 2002) and simulation of spatial system from (Szczesny *et al.*, 2014). (**c**) Experimental observation (Reichenbach *et al.*, 1968) and simulation (Janulevicius *et al.*, 2015) of circular aggregates formed by social motility of myxobacteria as an intermediate step in the development of fruiting bodies.

In microbial ecology, in contrast, data collection has mainly focused on characterisation of environmental samples from a multitude of different settings. We argue here that much can be gained by focusing efforts on a more limited set of well-defined 'model' MCs (Denef *et al.*, 2010; Groβkopf and Soyer, 2014; Estrela *et al.*, 2015), including both synthetic communities constructed from known microbial species (Bodenhausen *et al.*, 2014; Faith *et al.*, 2014; Bai *et al.*, 2015), and microcosm communities made from environmental samples (Foster and Bell, 2012; Pagaling *et al.*, 2014; Wolfe and Dutton, 2015). Such an approach offers many advantages. From a general point of view, focusing on a limited set of model systems would make results much more transferrable between studies, allowing multiple groups to work synergistically, and therefore speeding up progress toward mechanistic understanding. Lab experiments with well-defined MCs under controlled conditions, for example, in chemostats, should allow hypothesis testing via control of key external parameters such as substrate concentration, system size or temperature, and allow for multiple replicate experiments (Figure 5).

From a more specific point of view, synthetic ecological communities provide a way to limit the system to a manageable number of microbial constituents, making it simpler to analyse and model—it may even be possible to measure exhaustively kinetic and interaction parameters for an entire community, as input for theoretical models. Indeed, *de-novo* assembly of low diversity MCs (Mee and Wang, 2012) provides control of microbial interactions, non-linear effects because of adding traits or community members and evolutionary changes (Celiker and Gore, 2014; Fiegna *et al.*, 2014).

Synthetic communities can also be used to test the role of particular ecological mechanisms, such as metabolic interactions, spatial heterogeneity (van Gestel *et al.*, 2014) or induced cell death

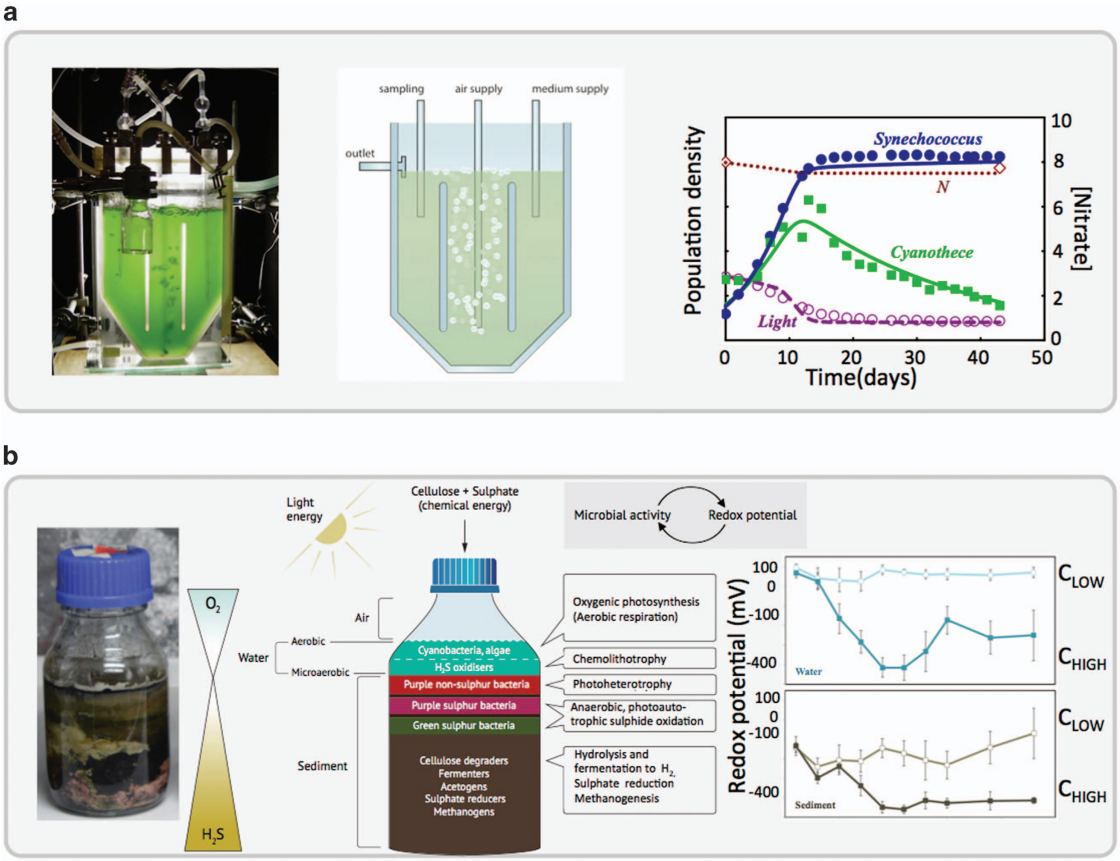

Figure 5 (**a**) (Left) Chemostat competition study between marine, nitrogen-fixing *Cyanothece sp.* and a non-nitrogen-fixing *Synechococcus* species (source: Department of Aquatic Microbiology, University of Amsterdam). (Middle) Schematic drawing of a chemostat. (Right) At high nitrate levels, the nitrogen-fixer (*Cyanothece*) is competitively excluded by the non-nitrogen-fixer (*Synechococcus*). Symbols are measurements; lines are model predictions (after Agawin *et al.*, 2007). (**b**) (Left) Study in Winogradsky column microcosms. (Middle) Schematic of the vertically layered structure of a mature Winogradsky column. Principal microbial types are found in different layers, their ecological activities and the associated core chemical reactions are illustrated. As a result opposing gradients of sulphide and oxygen develop. (Right) Microbial activity leads to a transient drop in redox potential in the overlying water, and a long-term drop in the sediment, at high levels of added cellulose ('high C'). Low levels of added cellulose ('low C') induce only a short-term reduction in redox potential.

(Asally *et al.*, 2012). As a pathway toward more systematic development of synthetic model MCs, we advocate building on the success of early microbiological studies in collecting data on specific model species under standardised culture conditions. In particular, we suggest that obtaining detailed physiological data (for example, kinetic growth parameters under different conditions) on a collection of ecologically relevant microbial model organisms, would greatly facilitate the community-wide construction of synthetic MCs. Although synthetic communities do not reproduce the full diversity and complexity of natural MCs, many of the principles of community organisation and dynamics, which we learn from them should be transferable to more complex MCs.

As a stepping stone from simplified synthetic systems to natural MCs, and to address questions concerning contingency, functional redundancy, species diversity and variability and the nature of interactions in highly diverse, complex communities (Foster and Bell, 2012), microcosms provide an excellent platform (Figure 5). A microcosm consists of an environmental sample that is cultured in the lab under well-defined conditions. The microcosm often retains much of the compositional and functional diversity of the seed community, and may also have spatial structure, allowing, for example, for nutrient cycles and redox gradients (Pagaling *et al.*, 2014). As microcosm experiments can be replicated, sampled and perturbed under well-controlled conditions, they provide an excellent bridge between the simplicity of synthetic model ecosystems and the complexity of MCs in the natural environment (Pagaling *et al.*, 2014; Wolfe and Dutton, 2015). We also advocate performing experiments on a limited set of standardized microcosm communities; for example, the Winogradsky column, in which an aquatic sediment-water sample develops under lab conditions into a self-sustaining, vertically stratified, nutrient-cycling community.

Conclusions

Despite impressive advances in our knowledge of the species composition of MCs, we are still far from achieving the level of fundamental understanding of the dynamics and function of MCs that is needed to predict and control MC behaviour. Here, we have argued that the key to achieving this level of predictive understanding is the integrative development of mathematical models with experimental data collection and method development. There are considerable challenges associated with both experimental method development and data collection, and mathematical model building in the study of MCs. These challenges are intertwined, such that tackling them effectively requires an integrated approach. To advance toward this goal, we advocate increased interaction between empirical and theoretical scientists, as well as the development of well-defined model MCs that can act as test-beds for the integrative development of experimental and modelling approaches.

How can this best be achieved in practice? Although many of the points that we make here are addressed to individual scientists, interactions between empiricists and theoreticians should be facilitated by the continuation of community-wide activities like the recent 4-month programme 'Understanding Microbial Communities' held at the Isaac Newton for Mathematical Sciences at Cambridge University. Moreover, some of the developments, which we call for here, such as the extensive characterisation of specific model systems, can best be done by groups of scientists working together rather than by individual groups. Specific funding of such community-directed research activities, in parallel with individual research efforts, would be a very welcome development.

Conflict of Interest

Dr Patrick B Warren holds equity (>$10k) in Unilever PLC. All other authors declare no conflict of interest.

Acknowledgements

We thank the Isaac Newton Institute of Mathematical Sciences for hosting the programme 'Understanding microbial communities: structure, function dynamics', which made this paper possible. This programme was partially funded by the Biotechnology and Biological Sciences Research Council and the US Army Research Office under grant number W911NF-14-1-0445. We thank all participants of the programme for inspiring discussions, and Aglika Gungova and Stela Ilieva for assistance with the figures.

Isaac Newton Institute Fellows

Steven D Allison[29], James Chong[30], Marco Cosentino Lagomarsino[31], Ottavio A Croze[32], Jérôme Hamelin[33], Jérôme Harmand[33], Rebecca Hoyle[34], Terence T Hwa[35], Qusheng Jin[36], David R Johnson[37], Víctor de Lorenzo[38], Mauro Mobilia[39], Barry Murphy[40], François Peaudecerf[41], James I Prosser[42], Robert A Quinn[14], Markus Ralser[43,44], Alison G Smith[45], Jean-Philippe Steyer[33], Neil Swainston[46], Corina E Tarnita[47], Eric Trably[33], Patrick B Warren[40], Paul Wilmes[48]

[29]Department of Ecology and Evolutionary Biology, University of California Irvine, USA, [30]Department of Biology, University of York, UK; [31]UMR7238 CNRS—Laboratoire de Biologie Computationnelle et Quantitative, Université Pierre et Marie Curie, France; [32]Cavendish Laboratory, University of Cambridge, UK; [33]INRA, UR0050, Laboratoire de Biotechnologie de l'Environnement, France; [34]Mathematical Sciences, University of Southampton, UK; [35]Division of Biological Sciences, University of California San Diego, USA; [36]Department of Geological Sciences, University of Oregon, USA; [37]Swiss Federal Institute of Aquatic Science and Technology (Eawag), Department of Environmental Microbiology, Switzerland; [38]Systems Biology Program, Centro Nacional de Biotecnología, Spain; [39]Department of Applied Mathematics, School of Mathematics, University of Leeds, UK; [40]Unilever R&D Port Sunlight, UK, [41]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK; [42]Institute of Biological and Environmental Sciences, University of Aberdeen, UK; [43]Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, UK; [44]The Francis Crick Institute, Mill Hill Laboratory, UK; [45]Department of Plant Sciences, University of Cambridge, UK; [46]School of Computer Science, University of Manchester, UK; [47]Department of Ecology and Evolutionary Biology, Princeton University, USA; [48]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg.

Glossary: Modelling terms

Population dynamic models. These models deal with the long- and short-time changes in size and composition of biological populations, as well as the biological and environmental factors affecting these changes.

Stoichiometric models. These describe the set of metabolites taking part in cellular metabolism and their reactions based on chemical stoichiometry at steady state. The framework allows graph theoretical investigation of the cellular metabolism and linear optimisation of cellular fluxes. The latter approach includes FBA that optimises to predefined constraints such as biomass yield for given uptake rates and dynamic FBA, which adds a dynamical framework.

Individual-based models. These represent individual organisms (here microbes) as 'agents', whose behaviour depends on a predefined set of rules. The global dynamics of the system are a

consequence of the interplay between the local interactions of all agents in the system.

Causal and correlational networks. Networks (graphs) are general tools for representation and analysis of systems with a distinct interaction structure. Nodes represent the constituents of the system, edges between them their interactions that can be based on causal evidence or on statistical inference. Studies on network topology and dynamics can predict community properties and behaviour.

Kinetic growth models. These models seek to describe the growth rate of a microbial population as a function of the concentrations of one or more chemicals (nutrients/inhibitors). To represent MCs, several coupled equations are used, each describing a different population within the community. An example of a kinetic growth model is the classical Monod model.

Well-mixed and spatially resolved approaches. In 'well-mixed models', for example, ODEs, the spatial positions of the model constituents are not resolved; they are assumed to be homogenously distributed in space. If spatial structure is important for the dynamics of a system, for example, if its behaviour is controlled by motility or by chemical diffusion processes, then 'spatially resolved' approaches such as continuum spatial models or spatially resolved individual-based models should be considered. These models represent space explicitly.

Evolutionary game theory. Game theory considers strategic decision making in a group of competitors, resulting in conflict and cooperation. The application of game theory to evolving biological populations is called evolutionary game theory. Questions focus on evolving strategies (change of strategies and its frequency), such as the rock–paper–scissors dynamics in bacteriocin-producing MCs (Figure 4).

References

Agawin NSR, Rabouille S, Veldhuis MJW, Servatius L, Hol S, van Overzee HMJ *et al.* (2007). Competition and facilitation between unicellular nitrogen-fixing cyanobacteria and non-nitrogen-fixing phytoplankton species. *Limnol Oceanogr* **52**: 2233–2248.

Asally M, Kittisopikul M, Rue P, Du YJ, Hu ZX, Cagatay T *et al.* (2012). Localized cell death focuses mechanical forces during 3D patterning in a biofilm. *Proc Natl Acad Sci USA* **109**: 18891–18896.

Bai Y, Muller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M *et al.* (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**: 364–369.

Baranyi J, Tamplin ML. (2004). ComBase: a common database on microbial responses to food environments. *J Food Prot* **67**: 1967–1971.

Berry D, Mader E, Lee TK, Woebken D, Wang Y, Zhu D *et al.* (2015). Tracking heavy water (D2O) incorporation for identifying and sorting active microbial cells. *Proc Natl Acad Sci USA* **112**: E194–E203.

Berry D, Widder S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front Microbiol* **5**: 219.

Bodenhausen N, Bortfeld-Miller M, Ackermann M, Vorholt JA. (2014). A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genet* **10**: e1004283.

Bonachela JA, Raghib M, Levin SA. (2011). Dynamic model of flexible phytoplankton nutrient uptake. *Proc Natl Acad Sci USA* **108**: 20633–20638.

Brown SP, Le Chat L, Taddei F. (2008). Evolution of virulence: triggering host inflammation allows invading pathogens to exclude competitors. *Ecol Lett* **11**: 44–51.

Celiker H, Gore J. (2014). Clustering in community structure across replicate ecosystems following a long-term bacterial evolution experiment. *Nat Commun* **5**: 4643.

Cordero OX, Ventouras L-A, DeLong EF, Polz MF. (2012a). Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc Natl Acad Sci USA* **109**: 20059–20064.

Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F *et al.* (2012b). Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* **337**: 1228–1231.

Coyte KZ, Schluter J, Foster KR. (2015). The ecology of the microbiome: networks, competition, and stability. *Science* **350**: 663–666.

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE *et al.* (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**: 559–563.

Davies D. (2003). Understanding biofilm resistance to antibacterial agents. *Nat Rev Drug Disc* **2**: 114–122.

Denef VJ, Mueller RS, Banfield JF. (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J* **4**: 599–610.

Desmond-Le Quemener E, Bouchez T. (2014). A thermodynamic theory of microbial growth. *ISME J* **8**: 1747–1751.

Eilam O, Zarecki R, Oberhardt M, Ursell LK, Kupiec M, Knight R *et al.* (2014). Glycan degradation (GlyDeR) analysis predicts mammalian gut microbiota abundance and host diet-specific adaptations. *mBio* **5**.

Elias S, Banin E. (2012). Multi-species biofilms: living with friendly neighbors. *FEMS Microbiol Rev* **36**: 990–1004.

Estrela S, Whiteley M, Brown SP. (2015). The demographic determinants of human microbiome health. *Trends Microbiol* **23**: 134–141.

Faith JJ, Ahern PP, Ridaura VK, Cheng J, Gordon JI. (2014). Identifying gut microbe-host phenotype relationships using combinatorial communities in gnotobiotic mice. *Sci Transl Med* **6**: 220ra211.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J *et al.* (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* **8**: e1002606.

Fiegna F, Moreno-Letelier A, Bell T, Barraclough TG. (2014). Evolution of species interactions determines microbial community productivity in new environments. *ISME J* **9**: 1235–1245.

Flint HJ, Duncan SH, Scott KP, Louis P. (2007). Interactions and competition within the microbial community of the human colon: links between diet and health. *Environ Mcrobiol* **9**: 1101–1111.

Foster KR, Bell T. (2012). Competition, not cooperation, dominates interactions among culturable microbial species. *Curr Biol* **22**: 1845–1850.

Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, Ruppin E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res* **38**: 3857–3868.

Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M *et al.* (2011). Competitive and cooperative metabolic interactions in bacterial communities. *Nat Commun* **2**: 589.

Frentz Z, Kuehn S, Hekstra D, Leibler S. (2010). Microbial population dynamics by digital in-line holographic microscopy. *Rev Sci Instrum* **81**: 084301.

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

Fuhrman JA. (2009). Microbial community structure and its functional implications. *Nature* **459**: 193–199.

Gomez P, Buckling A. (2013). Real-time microbial adaptive diversification in soil. *Ecol Lett* **16**: 650–655.

Gonzalez-Cabaleiro R, Lema JM, Rodríguez J, Kleerebezem R. (2013). Linking thermodynamics and kinetics to assess pathway reversibility in anaerobic bioprocesses. *Energ Environ Sci* **6**: 3780–3789.

Gore J, Youk H, van Oudenaarden A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459**: 253–256.

Groβkopf T, Soyer OS. (2014). Synthetic microbial communities. *Curr Opin Microbiol* **18**: 72–77.

Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH *et al.* (2014). Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep* **7**: 1104–1115.

Hibbing ME, Fuqua C, Parsek MR, Peterson SB. (2010). Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* **8**: 15–25.

Hillesland KL, Lim SJ, Flowers JJ, Turkarslan S, Pinel N, Zane GM *et al.* (2014). Erosion of functional independence early in the evolution of a microbial mutualism. *Proc Natl Acad Sci USA* **111**: 14822–14827.

Janulevicius A, Van Loosdrecht MCM, Picioreanu C. (2015). Short-range guiding can result in the formation of circular aggregates in myxobacteria populations. *PLoS Comput Biol* **11**: e1004213.

Jin Q, Roden EE, Giska JR. (2013). Geomicrobial kinetics: extrapolating laboratory studies to natural environments. *Geomicrobiol J* **30**: 173–185.

Johnson DR, Helbling DE, Lee TK, Park J, Fenner K, Kohler H-PE *et al.* (2015). Association of biodiversity with the rates of micropollutant biotransformations among full-scale wastewater treatment plant communities. *Appl Environ Microb* **81**: 666–675.

Kerr B, Riley MA, Feldman MW, Bohannan BJM. (2002). Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**: 171–174.

Kettle H, Louis P, Holtrop G, Duncan SH, Flint HJ. (2015). Modelling the emergent dynamics and major metabolites of the human colonic microbiota. *Environ Microbiol* **17**: 1615–1630.

Kiorboe T, Tang K, Grossart H-P, Ploug H. (2003). Dynamics of microbial communities on marine snow aggregates: colonization, growth, detachment, and grazing mortality of attached bacteria. *Appl Environ Microb* **69**: 3036–3047.

Klapper I, Dockery J. (2010). Mathematical description of microbial biofilms. *Siam Rev* **52**: 221–265.

Klitgord N, Segrè D. (2010). Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol* **6**: e1001002.

Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T, Watanabe K. (2008). The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota. *Genome Res* **18**: 442–448.

Leggett HC, Brown SP, Reece SE. (2014). War and peace: social interactions in infections. *Philos Trans R Soc Lond B Biol Sci* **369**: 20130365.

Little AE, Robinson CJ, Peterson SB, Raffa KF, Handelsman J. (2008). Rules of engagement: interspecies interactions that regulate microbial communities. *Annu Rev Microbiol* **62**: 375–401.

Mahadevan R, Edwards JS, Doyle R, Francis J. (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* **83**: 1331–1340.

McNally L, Viana M, Brown SP. (2014). Cooperative secretions facilitate host range expansion in bacteria. *Nat Commun* **5**: 4594.

Mee MT, Wang HH. (2012). Engineering ecosystems and synthetic ecologies. *Mol Biosyst* **8**: 2470–2483.

Momeni B, Waite AJ, Shou W. (2013). Spatial self-organization favors heterotypic cooperation over cheating. *Elife* **2**: e00960.

Muñoz-Tamayo R, Laroche B, Walter E, Dorè J, Leclerc M. (2010). Mathematical modelling of carbohydrate degradation by human colonic microbiota. *J Theor Biol* **266**: 189–201.

Orhon D, Artan N. (1994). *Modelling of Activated Sludge Systems*. Technomic Publishing: Lancaster, PA, USA.

Orth JD, Thiele I, Palsson B. (2010). What is flux balance analysis? *Nat Biotechnol* **28**: 245–248.

Pagaling E, Strathdee F, Spears BM, Cates ME, Allen RJ, Free A. (2014). Community history affects the predictability of microbial ecosystem development. *ISME J* **8**: 19–30.

Passarge J, Hol S, Escher M, Huisman J. (2006). Competition for nutrients and light: stable coexistence, alternative stable states, or competitive exclusion? *Ecol Monographs* **76**: 57–72.

Pfeiffer T, Schuster S, Bonhoeffer S. (2001). Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**: 504–507.

Post DM, Palkovacs EP. (2009). Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play. *Philos Trans R Soc Lond B Biol Sci* **364**: 1629–1640.

Reichenbach H, Heunert HH, Kuczka H. (1968). Archangium violaceum (Myxobacteriales)—Schwarmentwicklung und Bildung von Protocysten. *Institut für den Wissenschaftlichen Film*.

Reichenbach T, Mobilia M, Frey E. (2007). Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. *Nature* **448**: 1046–1049.

Reichenbach T, Mobilia M, Frey E. (2008). Self-organization of mobile populations in cyclic competition. *J Theor Biol* **254**: 368–383.

Riley MA, Wertz JE. (2002). Bacteriocins: evolution, ecology, and application. *Annu Rev Microbiol* **56**: 117–137.

Saunders AM, Albertsen M, Vollertsen J, Nielsen PH. (2016). The activated sludge ecosystem contains a core community of abundant organisms. *ISME J* **10**: 11–20.

Schilling CH, Schuster S, Palsson BO, Heinrich R. (1999). Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog* **15**: 296–303.

Shong J, Diaz MRJ, Collins CH. (2012). Towards synthetic microbial consortia for bioprocessing. *Curr Opin Biotech* **23**: 798–802.

Sieber JR, McInerney MJ, Gunsalus RP. (2012). Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol* **66**: 429–452.

Stolyar S, Van Dien Steve, Hillesland KL, Pinel N, Lie TJ, Leigh JA *et al.* (2007). Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol* **3**: 92.

Szczesny B, Mobilia M, Rucklidge AM. (2014). Characterization of spiraling patterns in spatial rock-paper-scissors games. *Phys Rev E* **90**: 032704.

Teles FR, Teles RP, Uzel NG, Song XQ, Torresyap G, Socransky SS *et al.* (2012). Early microbial succession in redeveloping dental biofilms in periodontal health and disease. *J Periodontal Res* **47**: 95–104.

van Gestel J, Weissing FJ, Kuipers OP, Kovacs AT. (2014). Density of founder cells affects spatial pattern formation and cooperation in Bacillus subtilis biofilms. *ISME J* **8**: 2069–2079.

Vannette RL, Fukami T. (2014). Historical contingency in species interactions: towards niche-based predictions. *Ecol Lett* **17**: 115–124.

Vetsigian K, Jajoo R, Kishony R. (2011). Structure and evolution of Streptomyces interaction networks in soil and in silico. *PLoS Biol* **9**: e1001184.

Vorholt JA. (2012). Microbial life in the phyllosphere. *Nat Rev Microbiol* **10**: 828–840.

West SA, Griffin AS, Gardner A, Diggle SP. (2006). Social evolution theory for microorganisms. *Nat Rev Microbiol* **4**: 597–607.

Widder S, Besemer K, Singer GA, Ceola S, Bertuzzo E, Quince C *et al.* (2014). Fluvial network organization imprints on microbial co-occurrence networks. *Proc Natl Acad Sci USA* **111**: 12799–12804.

Wolfe BE, Dutton RJ. (2015). Fermented foods as experimentally tractable microbial ecosystems. *Cell* **161**: 49–55.

Woodcock S, van der Gast CJ, Bell T, Lunn M, Curtis TP, Head IM *et al.* (2007). Neutral assembly of bacterial communities. *FEMS Microbiol Ecol* **62**: 171–180.

Zhou JZ, Wu LY, Deng Y, Zhi XY, Jiang YH, Tu QC *et al.* (2011). Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J* **5**: 1303–1313.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/