Microbial interactions: from networks to models

Karoline Faust[1,2] and Jeroen Raes[1,2]

Abstract | Metagenomics and 16S pyrosequencing have enabled the study of ecosystem structure and dynamics to great depth and accuracy. Co-occurrence and correlation patterns found in these data sets are increasingly used for the prediction of species interactions in environments ranging from the oceans to the human microbiome. In addition, parallelized co-culture assays and combinatorial labelling experiments allow high-throughput discovery of cooperative and competitive relationships between species. In this Review, we describe how these techniques are opening the way towards global ecosystem network prediction and the development of ecosystem-wide dynamic models.

Mutualism
An interaction between two species in which each species derives a benefit. Also referred to as cooperation or symbiosis by some authors; however, mutualism is preferred here because 'symbiosis' can be used in a broader sense to include all ecological relationships, and 'cooperation' can be used to designate mutualism between single organisms rather than populations.

Niches
Defined by Hutchinson as the volume in which the growth rate of an organism is larger than or equal to 1, where the volume is an abstract space with axes that correspond to abiotic and biotic factors that affect the growth rate of the species.

[1]*Department of Structural Biology, Flemish Institute for Biotechnology (VIB), Belgium.*
[2]*Department of Applied Biological Sciences (DBIT), Vrije Universiteit Brussel, Belgium.*
Correspondence to J.R.
e-mail: jeroen.raes@vib-vub.be
doi:10.1038/nrmicro2832

Microorganisms (including viruses, bacteria, archaea and protists) do not exist in isolation but form complex ecological interaction webs. Interactions within these ecological webs can have a positive impact (that is, a win), a negative impact (that is, a loss) or no impact on the species involved. The possible combinations of win, loss and neutral outcomes for two interaction partners allow the classification of various interaction types[1] (FIG. 1). For instance, bacteria (of different taxonomic groups) may cooperate to build a biofilm, which confers antibiotic resistance to its members[2] — a win–win relationship that is known as mutualism. Other examples for mutualism are certain cases of cross-feeding (also known as syntrophy), in which two species exchange metabolic products to the benefit of both[3]. Classical loss–win interactions are predator–prey relationships (for example, ciliates feeding on bacteria) and host–parasite relationships (for example, between bacteria and their bacteriophages). Amensalism — in which one partner is harmed without any advantage to the other — occurs, for example, when metabolic by-products of a microbial species alter the environment to the detriment of other microorganisms (for example, lactobacilli lowering the pH of the surrounding environment). In commensalistic relationships, one partner benefits without helping or harming the other. Commensalism is often found in biodegradation, in which commensals cross-feed on compounds that are produced by other community members (for example, in cellulose degradation[4]). Finally, a famous example of competition between microorganisms (that is, a loss–loss relationship) dates back to the 1930s, when Gause[5] executed a series of co-culture experiments. He observed that for a number of species pairs, each species grew well in the absence of the other but, when co-cultured, one species (in this case, *Paramecium aurelia*) subdued the other (in this case, *Paramecium caudatum*). On the basis of these observations, Gause formulated his law of competitive exclusion, which states that two species with similar niches exclude each other.

Detecting and investigating these various types of interactions in microbial ecosystems is far from straightforward[6]. Novel approaches towards the reconstruction of ecosystem-wide association networks can open the way towards global models of ecosystem dynamics. Ultimately, such models will be able to predict the outcome of community alterations and the effects of perturbations, and they could, in the long run, help with the engineering of complex microbial communities.

Here, we review these various strategies, discuss their strengths and weaknesses and give an overview of the insights into microbial ecology that these studies have provided.

From abundance data to community models

Microbial communities can be monitored with a range of techniques, ranging from flow cytometry[7] via microarrays[8] to ribosomal RNA[9,10] and metagenomic sequencing[11,12]. The extraction of microbial abundances from the raw data is a complex, multiple-step procedure that relies on sequence clustering and reference databases (for example, REFS 12–14) to assign reads to known microbial taxa. A number of pipelines that facilitate this task have been published in recent years (see REF. 15 for an overview).

© 2012 Macmillan Publishers Limited. All rights reserved

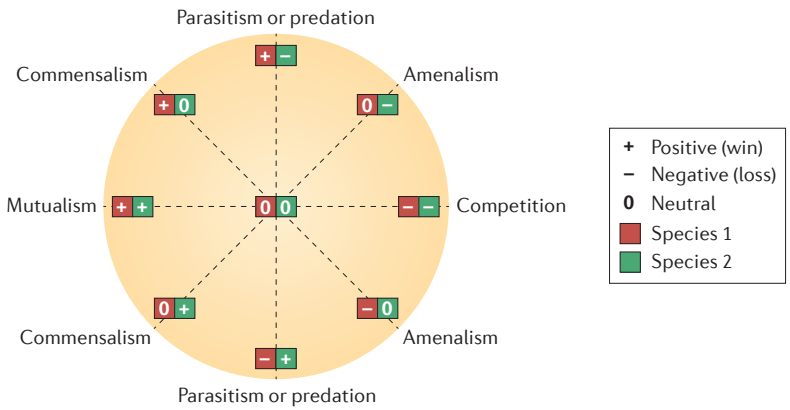


Figure 1 | **Summary of ecological interactions between members of different species.** The wheel display introduced by Lidicker[1] has been adapted to summarize all possible pairwise interactions. For each interaction partner, there are three possible outcomes: positive (+), negative (–) and neutral (0). For instance, in parasitism, the parasite benefits from the relationship (+), whereas the host is harmed (–); this relationship is thus represented by the symbol pair +–.

After abundance data have been obtained, it is possible to predict microbial relationships under the premise that strongly nonrandom distribution patterns are mostly due to ecological reasons (see BOX 1 on the debate about random processes versus ecological interactions in community assembly). This approach goes a long way back: in the 1970s, Jared Diamond suggested that 'checkerboard' species distributions (that is, distributions that are mutually excluding) are a consequence of competitive relationships (BOX 1). Predicting relationships from this principle is straightforward: when two species (or any taxonomically relevant units) co-occur or show a similar abundance pattern over multiple samples, a positive relationship is assumed; when they show mutual exclusion (as in the checkerboard pattern) or anticorrelation, a negative one is predicted. However, interpretation of the ecological relevance of these relationships is far from easy. For instance, a positive relationship can be due to cross-feeding, co-aggregation in biofilms, co-colonization, niche overlap or other reasons, whereas a negative relationship may result from amensalism, a prey–predator relationship, competition, and so on. In addition, relationships can be time-lagged (for example, if one species increases its abundance at a certain moment, another species might only disappear at a later time). Further (experimental) validation of such interactions is currently needed to disentangle the exact reasons behind the observed patterns.

The prediction of microbial association networks from presence–absence or abundance data is a problem that is known as network inference in computer science, and there is a long history of methods for its resolution[16,17]. Network inference techniques are widely applied in genomics[18,19] and are starting to be adopted in ecology as well[20]. Here, we distinguish between two groups of network inference methods: those that can predict relationships between two species (namely, pairwise relationships) and those that can predict more complex ones.

Network inference
The process of reconstructing the wiring diagram of a complex system from the behaviour of its components. For microbial communities, the goal of network inference is to predict ecological relationships between microorganisms from abundance data.

Regression
Prediction of a relationship between a dependent variable (here, the abundance of a target species) and independent variables (the abundance (or abundances) of one or more independent source species, environmental traits and possibly a noise term), which are termed factors here.

Pairwise relationships: similarity-based network inference. Although there are many subtleties and pitfalls (see below), generally speaking, similarity-based network inference assesses the co-occurrence and/or mutual exclusion pattern of two species over multiple samples using a measure that quantifies the similarity of two species distributions. In a second step, the significance of the similarity score is assessed (usually with a null distribution generated by permuting the data set in question). After assessing all possible combinations of species in a given abundance data set, all significant pairwise relationships are then combined to construct a network (FIG. 2).

Complex relationships: regression- and rule-based networks. Pairwise relationships do not capture more complex forms of ecological interactions, in which one species depends on (or is influenced by) multiple other species.

To infer these types of interactions, other techniques must be applied. The first approach is regression: the abundance of one species is predicted from the combined abundances of other organisms (FIG. 2). Although the method is simple and often used, the meaning and interpretation of regression results are sometimes difficult, especially if factors are chosen using automatic feature selection methods (such as sparse regression). Even though the method often automatically selects sets of species that can mathematically predict the abundance of another species, this does not always mean that there is a biological reason for this. A second approach towards complex relationship inference is association rule mining[21]. In short, this approach consists of enumerating all logical rules supported by a presence–absence data set to find significant rules. For example, a rule could say 'in the presence of species A and the absence of species B, species C is also present'. Rules are mined by enumerating all possible sets of taxa up to a certain size and generating all possible rules for each set. In the second (and more challenging) step, a combination of filters is applied to correct for multiple testing and to retain only significant rules.

A complex relationship that is inferred through multiple regression or association rule mining can be represented in the resulting network as an edge that connects more than two nodes in a directed way to point from the independent taxa to the dependent taxon. Networks with such edges are formally known as directed hypergraphs. FIGURE 3a displays a network inferred from a similarity-based approach, in which pairwise relationships are represented by edges connecting two nodes, whereas FIG. 3b gives an example of a directed hypergraph that results from association rule mining in a global microbial presence–absence data set[22] and that visualizes complex relationships with hyper-edges connecting up to three nodes. This said, more developments will be needed to design and to apply targeted multivariate approaches truly to disentangle complex relationships as well as to visualize them.

© 2012 Macmillan Publishers Limited. All rights reserved

Box 1 | **Community assembly: ecological interactions versus random processes**

The extent to which ecological interactions between species shape their abundances remains the subject of debate. In his seminal paper on the distribution of bird species across tropical islands, Jared Diamond[102] derived community assembly rules from the birds' presence–absence patterns. In brief, he suggested that the presence of one bird species will prevent other bird species that are adapted to the same niche from settling on the same island, allowing competitive exclusion to be deduced from 'checkerboard'-like presence–absence patterns.

This study sparked a decades-long discussion about the importance of such assembly rules in the formation and structure of communities. Connor and Simberloff[103] attacked Diamond's rules as being partly redundant and criticized that Diamond's patterns did not withstand significance testing in simulations. Discussion of whether species presence–absence or abundance patterns are due to chance or due to ecological processes has advanced the development of null models[53]. A null model can be described as a set of rules for the simulation of data that are expected under the null hypothesis. In this context, it is of interest to mention Hubbell's unified theory of biogeography and biodiversity (from here on abbreviated as 'the neutral theory')[83]. According to the neutral theory, observed species distributions can be well explained by random processes of birth, death and immigration from a 'metapopulation'. Because it does not model the impact of ecological interactions on species abundances, Hubbell's neutral theory can serve as a null model against which the impact of ecological interactions can be tested. The neutral theory has been confirmed in some ecosystems[104] but contradicted in other studies[105]. This ambiguity suggests that both assembly rules and random processes can determine species abundance distributions, possibly to different, ecosystem-specific extents.

The niche-based analysis (using Diamond's rules) and the neutral-theory-based analysis (using Hubbell's model) of species abundances have both been applied to microbial distribution data[106,107]. For instance, the number of perfect checkerboard patterns is significantly higher than random expectation in a collection of microbial data sets spanning samples from soil, sediment, freshwater and marine communities[106] (supporting Diamond's assembly rules); however, checkerboards can be caused both by competition and by different niche preferences or history[108].

In REF. 107, Hubbell's model is converted into a continuous form to deal with large microbial populations, and different growth rates are introduced in the model (thus relaxing the neutrality assumption). In this extended form, the neutral theory is able accurately to describe microbial abundances measured in lakes and in the lower respiratory tract but not in human faeces (which is possibly explained by the presence of distinct community configurations in the gut[36]).

In summary, both ecological and random processes seem to have a role in microbial communities and ideally should both be taken into account.

Network inference and environmental traits. Network inference can incorporate measurements of environmental traits by treating them as additional 'species' in similarity-based approaches[23,24] or as additional factors in regression- or logical-rule-based approaches. This allows predicting relationships between species and environmental traits. For instance, strong relationships between microorganisms and concentrations of nutrients such as nitrites and nitrates could pinpoint species that are involved in global biogeochemical cycling[25,26].

From static networks to dynamic models. Network inference produces a set of significant dependencies (that are either simple or complex) from multiple measurements (that are transversal, longitudinal or both) of a microbial community. These dependencies can then help to build dynamic models of the microbial community. A dynamic model consists of a set of differential or difference equations or Boolean functions[27] that describe the change of community member abundances over time. Dynamic modelling has a long tradition in population ecology[28], but few studies have attempted to build dynamic models for microbial communities that consist of multiple species. Mounier and colleagues[29] modelled cheese fermentation community interactions with generalized Lotka–Volterra equations (gLV equations) (BOX 2). The model predicted negative interactions among three yeast species that were afterwards confirmed by co-culture experiments. Likewise, Hoffmann and colleagues[30] tested several approaches to model marine phage abundances. From the best-fitting model (which is a modified gLV equation), the authors concluded that the dynamics of phage–host communities alternates between long periods of very low abundance and brief periods of bloom. Phage abundances are predicted to rise quickly following the blooms of their specific hosts (in agreement with the 'kill the winner' model[31]), after which phage numbers rapidly decrease at first and then more slowly, such that they survive at very low numbers until the next bloom of their host. This cycling through different dominating hosts that are decimated by their respective phages has recently been confirmed at the strain level[32].

Lotka–Volterra equations
Equations that describe the dynamics of a prey–predator system. In their generalized form, the Lotka–Volterra equations can model the dynamics of more than two species and describe relationships other than prey–predator.

Hypergeometric distribution
This distribution underlies Fisher's exact test, which is commonly used to infer networks from presence–absence data. Given the occurrences of two taxa across the samples, the test assesses the significance of the number of observed co-presences.

Microbial network inference — state of the art

Microbial association networks have been inferred for a range of communities, from soil[33,34] and ocean communities[24,35] to human body communities[12,36,37]. In addition, global networks have been independently constructed from different data sets (namely, Greengenes rRNA database and PubMed abstracts)[22,38]. In these studies, the networks were all built with similarity-based techniques, mainly using either Pearson or Spearman correlations for abundance data[12,33,34,36] and the hypergeometric distribution for presence–absence data[22,38]. Another popular similarity-based network inference methodology is local similarity analysis (LSA)[39], which can detect similarity between shifted abundance profiles and is therefore frequently used to build association networks from time series data[23,24,40]. To set thresholds on similarity scores, most authors use a permutation test, as described in FIG. 2. In an interesting alternative approach, Pearson correlation thresholds are determined using random matrix theory[34].

© 2012 Macmillan Publishers Limited. All rights reserved

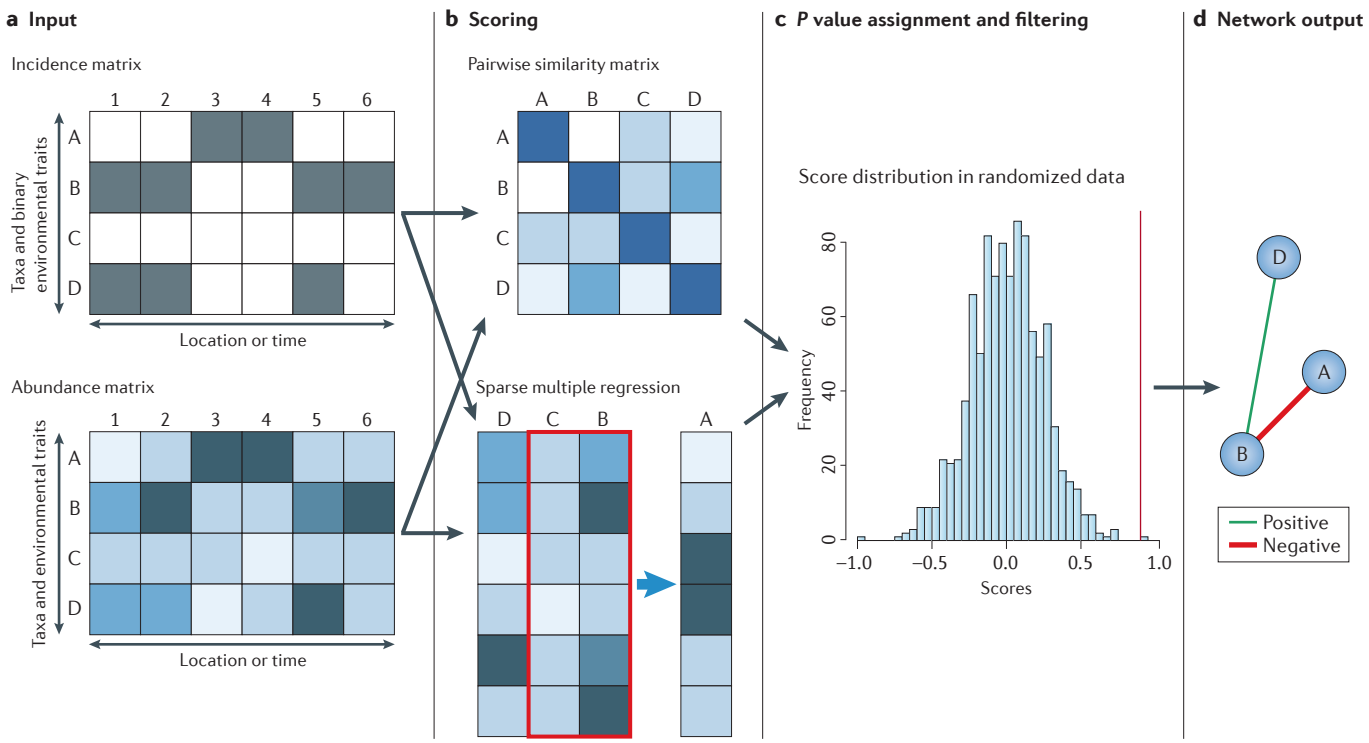


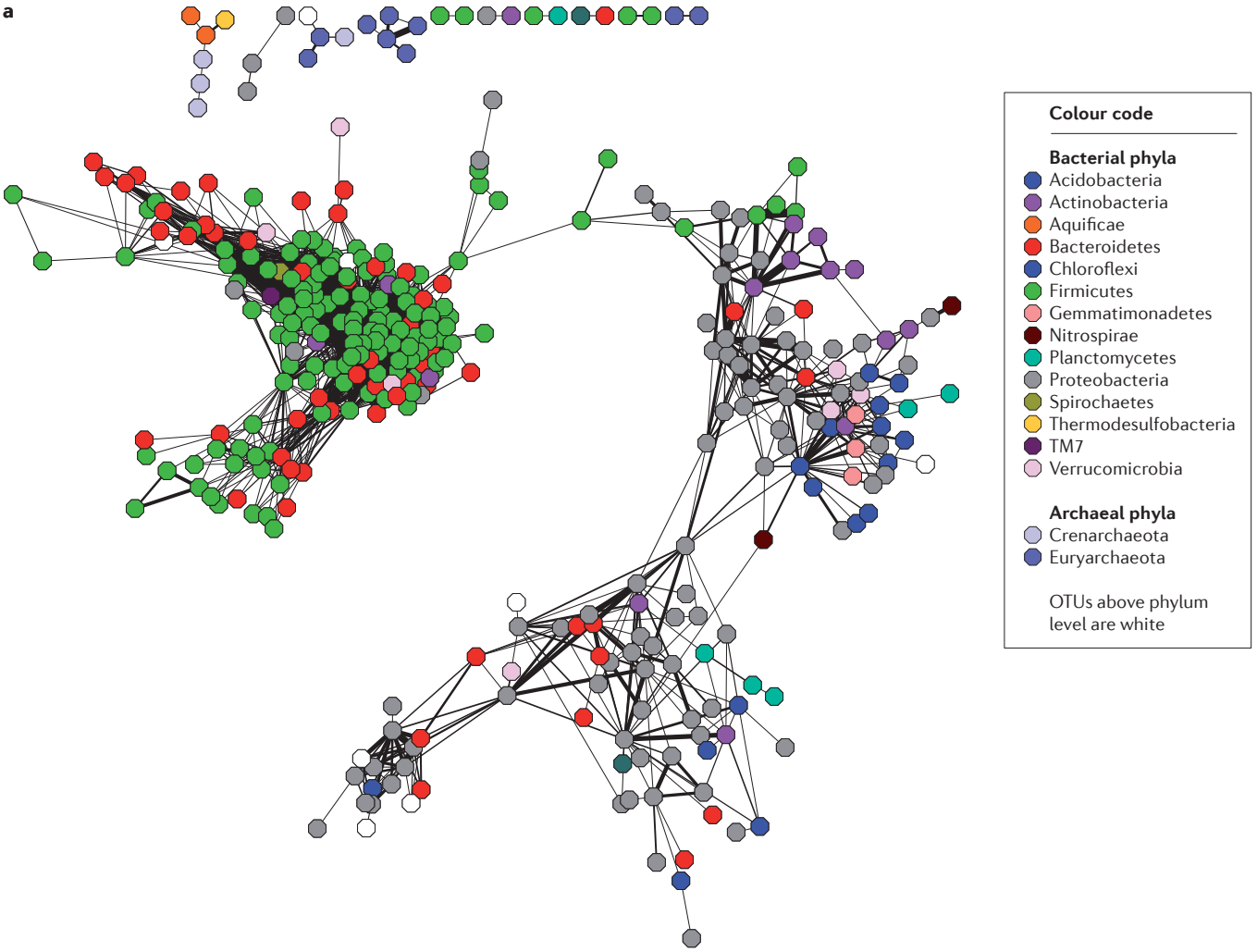
Figure 2 | **Principle of similarity- and regression-based network inference.** The goal of network inference is to identify combinations of microorganisms that show significant co-presence or mutual exclusion patterns across samples and to combine them into a network. **a** | Network inference starts from an incidence or an abundance matrix, both of which store observations across different samples, locations or time points. **b** | Pairwise scores between taxa are then computed using a suitable similarity or distance measure. A range of such measures are used in the literature (for example, Pearson, Spearman, hypergeometric distribution and the Jaccard index). In contrast to similarity-based approaches, multiple regression can detect relationships that involve more than two taxa. To reduce over-fitting, sparse multiple regression is usually carried out — that is, the source taxa subset that best predicts the target taxon's abundance is selected. In addition, the regression model is cross-validated: that is, after regression coefficients have been identified with a training data set, the model's prediction accuracy is quantified on a test data set. **c** | In the next step, a random score distribution is generated by repeating the scoring step a large number of times (often 1,000 times or more). The random score distribution computes the *P* value (that is, the probability of obtaining a score by chance that is equal to or better than the observed score) to measure the significance of the predicted relationship. The *P* value is usually adjusted for multiple testing with procedures such as Bonferroni or Benjamini–Hochberg[54]. **d** | Taxon pairs with *P* values below the threshold are visualized as a network, where nodes represent taxa and edges represent the significant relationships between them. The edge thickness can reflect the strength of the relationship.

Association networks that were built for marine and lake ecosystems not only captured links between microorganisms within and between different domains but also between microorganisms and seasonally variable environmental traits, such as concentrations of nitrites, nitrates and phosphates[23,24]. Apart from predicting individual relationships between microorganisms, the structure of association networks also gives insights into the organization of microbial communities (see TABLE 1 for an overview). For instance, microbial networks, like protein interaction and human social networks, are scale-free[22,24,33,34]. This implies the presence of many taxa with only a few links and a few highly connected (hub) taxa. It is as of yet unclear whether these hub taxa are ecologically more relevant than less-connected taxa. In addition, the networks can be partitioned into clusters (which are densely interconnected nodes), and these are interpreted as groups of taxa with overlapping niches[22,33,38]. Likewise, phylogeny shapes these networks: co-occurring taxa are more closely related than would be expected to occur at random[22,33]. Interestingly, network properties can change when an important environmental parameter is altered, as has been demonstrated for soil communities at ambient and elevated carbon dioxide levels[34].

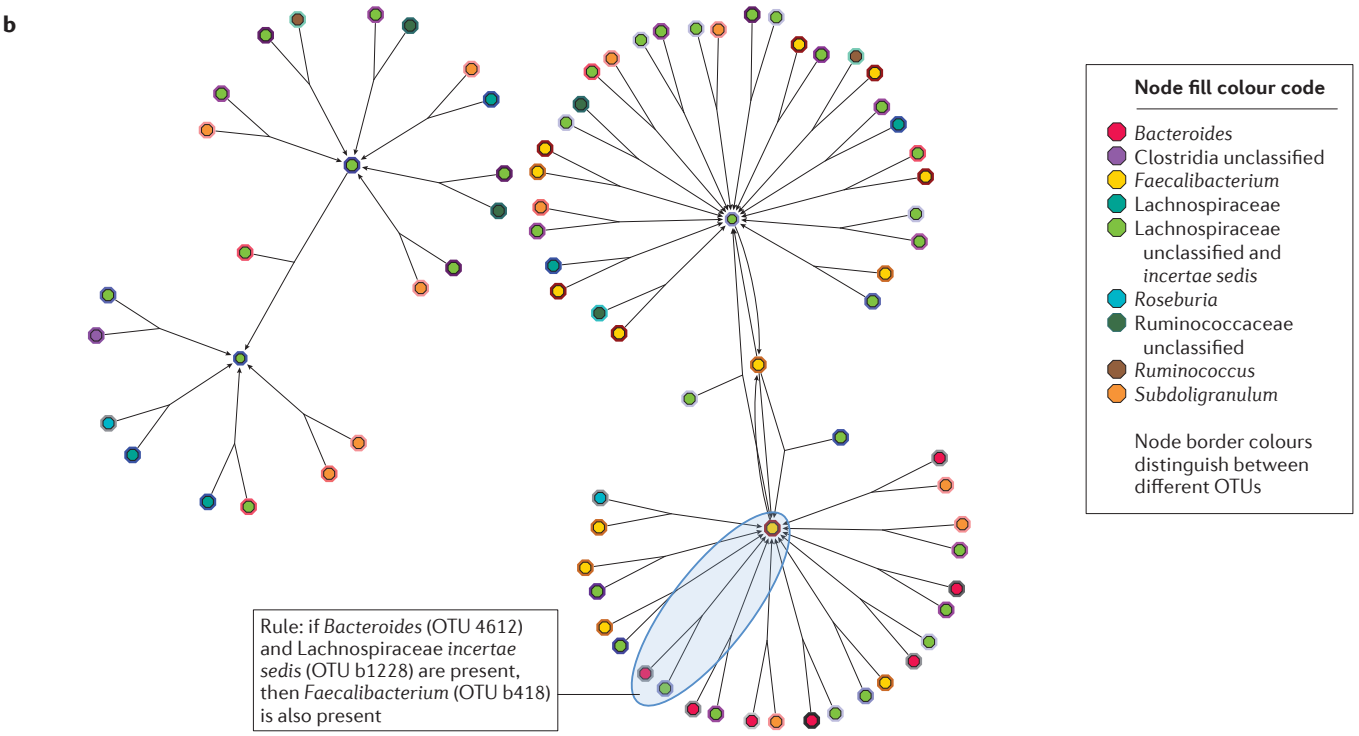
With the recent expansion of the field towards studying host-associated communities, it will not take long for ecological approaches, as discussed here, to intersect with clinical research. Networks for healthy human gut communities have already been inferred. For instance, a strain-level network that had been built from 124 illumina-sequenced stool samples places Bacteroidetes and Firmicutes in different clusters[12]. Association networks were also used to identify the members of alternative gut communities (which are termed enterotypes)[36]. Recently, a microbial association network covering 18 human body sites (including the oral cavity, airways, skin, gut and vagina) has been constructed from 16S sequence

© 2012 Macmillan Publishers Limited. All rights reserved

a



b



© 2012 Macmillan Publishers Limited. All rights reserved

◀ Figure 3 | **Examples for the prediction of pairwise versus complex relationships.** Pairwise (**a**) and complex relationships (**b**) were inferred from a global microbial operational taxonomic unit (OTU) presence–absence data set[22]. **a** | Each node represents an OTU, and each edge represents a significant pairwise association between them. Significant relationships were detected with the hypergeometric distribution (the *P* values of which were adjusted for multiple testing). The edge thickness increases with significance. **b** | This network summarizes association rules mined with the a priori algorithm[21,109] and filtered with the multiple testing correction suggested in REF. 110. The text box provides an example for such a rule. As the data set is extremely sparse, rules are restricted to positive associations involving up to three OTUs. Each node in network **b** represents an OTU, whereas each edge corresponds to a rule. In contrast to network **a**, an edge can connect three OTUs if they are all involved in the same rule. For ease of interpretation, the same OTU (with the same node fill and border colour) may occur multiple times in network **b**.

data collected and processed by the Human Microbiome Consortium[41] for a healthy cohort of 242 individuals. By combining several similarity measures with sparse regression in an ensemble approach, microbial relationships have been inferred within and between body sites[42]. Cross-body-site associations mostly occurred between microorganisms in related body sites (for example, between different vaginal sites), a finding that is in agreement with the discovery that body sites represent different microbial niches[43]. Alternative community configurations that are known for different stages of dental plaque formation[44] and that are also described for the gut and vagina[36,45] were indirectly captured as strong mutual exclusions. Furthermore, the comparison of phylogenetic and functional similarity of associated microorganisms showed that closely related pairs co-occurred in related habitats (in agreement with REF. 22), whereas exclusive relationships tended to occur in more distantly related pairs.

Pitfalls in microbial network inference

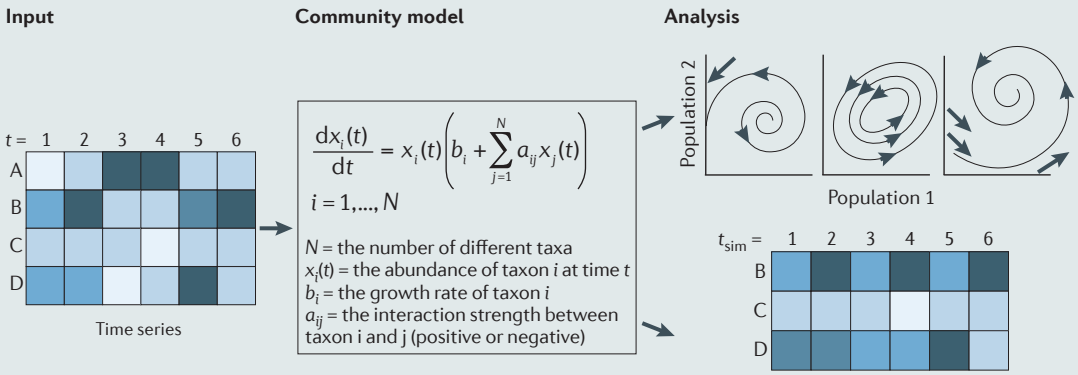
Network inference is a versatile tool for predicting relationships that has several advantages. It is generic (that is, networks can be built for genes, pathways and species), it can integrate different data types (such as species abundances and environmental traits), it can accommodate hierarchical data (that is, when controlling for parent–child relationships, networks can be built for several taxonomic ranks simultaneously, allowing detection of relationships that occur between higher taxonomic units), and it can identify community properties that are encoded in the network structure (such as niches, hub species and alternative communities).

However, network inference from microbial abundance data faces multiple pitfalls that are often ignored but that can seriously affect the end results, such as normalization, similarity measure biases, the choice of appropriate null models and multiple testing issues. Furthermore, interpretation of such networks is not straightforward, and conclusions should be drawn with great care.

Pitfalls in network construction. In the ideal case, samples of equal volume are collected, and all microorganisms therein are counted and analysed with standardized and uniform procedures. For count data thus obtained, no bias is introduced, and consequently no normalization is required. In most cases, however, a whole processing pipeline leads from the samples to the abundance data, and this can introduce biases at each step along the line[6,46–48]. One important effect of current sampling and sample characterization pipelines is that unequal

Box 2 | **Studying the dynamics of microbial communities with generalized Lotka–Volterra equations**

The generalized Lotka–Volterra (gLV) equations (see the figure) are an example of a dynamic model of a microbial community. As an input, they require the growth rates and interaction strengths of the different community members, which can either be obtained experimentally or can be estimated from time series data using network inference techniques. The equations in this simplified form have drawbacks — for example, they do not account for random processes and miss an upper bound for total species abundances — but they are a good starting point for more sophisticated community models. After a model has been defined and its parameters have been determined, it can be used to study community stability and to carry out simulations to predict community behaviour in altered conditions. *t*, time; t_{sim}, represents taxon units simulated after 1–6 time units. The top panel in the 'Analysis' part of the figure is adapted, with permission, from REF. 73 © (1973) Annual Reviews.



Top: identification of unstable and stable states using stability analysis; bottom: simulations of community dynamics with various initial conditions

Operational taxonomic unit (OTU). Refers to bacterial and archaeal taxonomic groups that are derived by sequence clustering and that are thus specific to the samples analysed.

© 2012 Macmillan Publishers Limited. All rights reserved

Table 1 | **Properties of microbial association networks and their interpretation**

Network property	Application to microbial networks
Node degree distribution	The distribution of the number of interactions each node has in a network. In random networks, the node degree distribution follows a Poisson distribution[111]. However, for most biological networks, the degree distribution is better described by a power law distribution ('as seen in scale-free' networks)[112]. Although the node degree distributions of microbial networks are not always fit by a power law[24], they are clearly far from being random[22,24,34]
Hub nodes	Scale-free networks have many nodes with few links and a few highly connected nodes that are termed hubs. They are therefore supposed to be robust towards random node removal but sensitive to the removal of hub nodes[113]. In REF. 24, the hub nodes have been linked to the ecological concept of the keystone species[64]
Modularity	Networks can be divided into clusters either manually or by using a network cluster algorithm (for example, REF. 114). More modular networks have a higher number of within-cluster edges than between-cluster edges compared to random expectation[115]. Clusters have been interpreted as niches in REFS 22,38
Average shortest path length (AL)	AL is calculated as the average number of steps in the shortest paths between each node to each other node in a network. Networks with a small AL are also known as small-world networks[116]. Microbial networks have a mostly small AL. A small AL has been interpreted to increase the speed of the network's response to perturbations[34]

amounts of abundance-yielding material are obtained per sample (for example, the amount of DNA extracted, the amount sequenced and the percentage of high-quality reads). For this reason, measurements are usually converted into relative abundances through normalization or downsampling. However, computing correlations between the components of compositions (such as the relative abundances of taxa across samples) can strongly distort results — an effect that is known as compositionality bias[49] — yielding artefactual correlations. For example, artefactual correlations may occur between non-correlated, low-abundance community members in the presence of highly abundant members. To assess similarity in compositional data, the choice of distance metric can help; Aitchison[49] suggested log-ratio-based measures, but these necessitate the treatment of zeros to avoid negative infinities. Alternatively, dedicated normalization strategies can mitigate this effect as well[42].

Another feature of microbial abundance data is the presence of a large percentage of zeros (that is, there is no observation) in the abundance matrix. This is problematic because of their ambiguous interpretation: a zero in a data set obtained from 16S or metagenomic sequencing can either mean that the taxon is indeed absent from the sample or, more likely, that its abundance is below the detection level. This data sparsity problem is most prominent for microorganisms in the 'rare biosphere'[50], which often have to be excluded from the analysis altogether because they are absent in too many samples. Besides data reduction, some of these problems can be alleviated by using the right distance measure.

When choosing a measure, several criteria have to be taken into account, such as its robustness to noise and to outliers, the assumption it makes about the relationship between the two taxa being compared (for example, a Pearson correlation assumes it to be linear) and its sensitivity to data sparsity and compositionality issues. For instance, the Bray–Curtis dissimilarity does not take into account matching zeros in abundance profiles (as do correlation measures) and is robust to the compositionality bias, but from our experience it can miss genuine relationships that would be captured by other measures. In this context, a recently published measure based on mutual information might be worth exploring[51]. FIGURE 4 shows that the choice of the measure has a large impact on the resulting network. An extensive comparison of various measures, as was recently done for the comparison of microbial community composition[52], is still lacking in the context of microbial network inference.

As explained in FIG. 2, the significance of a predicted association is assessed by computing a score distribution from random data, which in turn is generated from a null model (BOX 1). However, the 'best' randomization procedure for abundance data has yet to be determined; for instance, Gotelli[53] lists no less than nine ways of data shuffling, each of which preserves a different combination of data properties. Thus, to select a suitable null model, it is necessary to know which data properties have to be preserved and which can be relaxed, a choice that ultimately depends on the sampling strategy and the ecosystem (or ecosystems) that are under study. Because the null model determines the threshold, its choice has a huge impact on the result and therefore has to be made with care.

The more taxon pairs that are compared for significant (anti)correlation, the more likely it is that some will be significant by chance alone. Multiple testing correction is therefore an essential feature of every network inference method. Usually, *P* values are adjusted using the Benjamini–Hochberg procedure[54] (for example, REFS 22,38,39), which controls the expected proportion of wrongly rejected null hypotheses (that is, the number of false-positive associations).

The multiple testing problem is even more pronounced in regression- and logical-rule-based approaches, which allow unravelling relationships between more than two taxa. For this reason and to avoid over-fitting (that is, an overly accurate fit to the data that captures noise rather than a relevant relationship), the number of independent taxa that are considered for predicting the

© 2012 Macmillan Publishers Limited. All rights reserved

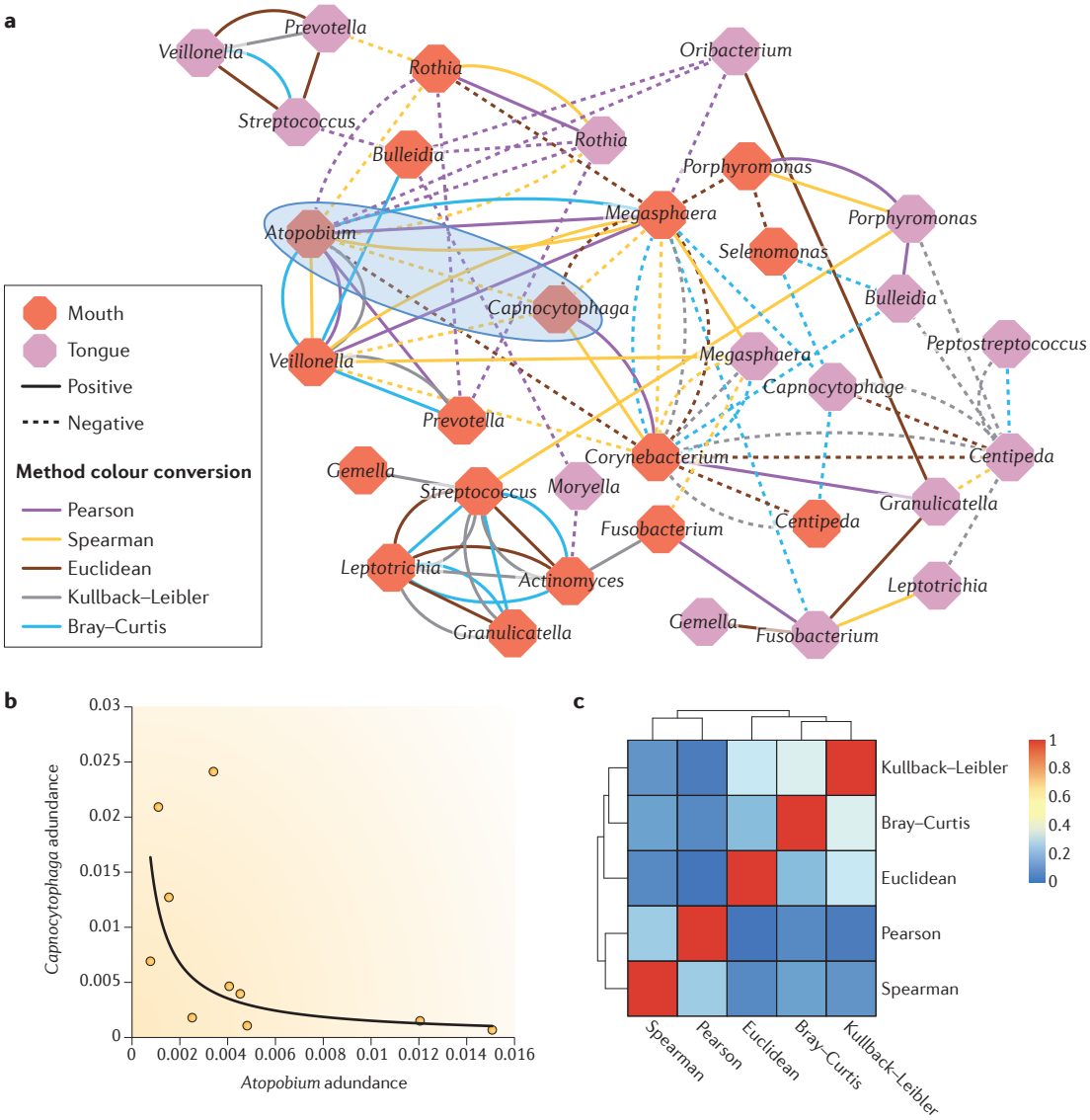


Figure 4 | **Impact of the similarity measure on the network inference result.** The network inferred from the oral cavity subset of a human microbial data set[43] illustrates the impact of the measure. The top ten positive and negative edges of five different measures are displayed together in a multi-graph (**a**). The measures rank some edges consistently high but disagree on others. The scatter plot (**b**) illustrates the nonlinear relationship between the abundances of *Atopobium* and *Capnocytophaga*, which was detected by a Spearman correlation but not by a Pearson correlation. When clustering the measures on the basis of their edge overlap using Jaccard's index (**c**), two groups emerge: the first group spans correlations and the second group consists of the dissimilarities.

profile of a dependent taxon needs to be restrained (for example, by sparse regression).

Another pitfall in network construction is the faulty prediction of a relationship between two species because both are affected by a third one (for example, a positive link between two microbial taxa that are prey to a third). Tests for conditional independence (such as partial correlation or data-processing inequality) are commonly applied to remove these indirect links (for example, REF. 55).

In addition to identifying indirect links, partial correlations can predict directed edges[56] (that is, asymmetric relationships), which would allow distinguishing unidirectional ecological interactions (such as amensalism and commensalism) from bidirectional ecological interactions (such as competition and mutualism). However, as the number of samples is typically smaller than the number of taxa in each sample (this is known as the 'small *n* large *p*' problem), classical partial correlations cannot be computed (see REF. 57 for an alternative). The small *n* large *p* situation also affects the calculation of other measures, such as correlation[58].

Pitfalls in network interpretation. When interpreting the inferred network, it is important to keep in mind that most network inference approaches cannot distinguish between true ecological interactions and other nonrandom processes (for example, cross-feeding versus niche overlap). To differentiate between some of

© 2012 Macmillan Publishers Limited. All rights reserved

these possibilities, metabolic data can be considered. The analysis of the metabolic activities or potential (for example, from (meta)genomics data or activity profiling) can reveal complementary pathways, indicating syntrophy as well as redundant specialized pathways, which can be a telltale sign of competition[38,59]. In addition, because taxon assignment on lower taxonomic levels is not always accurate, a link between two closely related taxa may be entirely due to mutual (that is, false) cross-assignment. However, a relationship between niche similarity, co-presence and phylogenetic similarity does exist[11,22]. To disentangle ecological signals and artefacts that occur because of inaccurate assignment, some authors (for example, REFS 36,38) only consider links above a certain taxonomic rank (for example, above family).

After a network has been obtained, various network properties (TABLE 1) can be computed with publicly available network analysis tool suites (for example, REFS 60–62). However, it is important to keep in mind that network properties might not be stable under changing network inference parameters. Furthermore, the network representation simplifies the system to such a degree that network properties can be misleading[63] if they are not interpreted with care. An example is the interpretation of hub species (that is, species that are associated with many other species) as keystone species, the removal of which has a large impact on the community structure[24]. This concept, which was originally applied to top predators[64], is extended here to any species for which removal is assumed to alter the community substantially. However, the effect of the removal of a species with few connections can be enhanced by trophic cascades, as in the case of top predators. Vice versa, the effect of a hub species deletion may be small if the links are mostly due to niche overlap. A dynamic model is therefore better suited to studying the impact of species removal on the system.

As the field is rather new, the choice of the similarity measure, the null model or the threshold is often made without further justification. In the absence of a benchmark data set, it is not clear whether one measure or null model is indeed superior to another. A comprehensive data set of known microbial interactions is needed to compare alternative microbial network inference approaches.

Experimental discovery of species interactions

Gause's co-culture experiments[5], which demonstrated competition between microbial species, are today a standard procedure for experimentally investigating microbial interactions. Co-culture experiments create an artificial community in a controlled environment and thus provide ideal conditions to test ecological concepts concerning community stability and dynamics that cannot easily be measured in macro-ecological systems. For instance, an important debate in population ecology concerned the possibility of chaos in population dynamics[28,65]. Becks and colleagues[66] could demonstrate chaotic behaviour in a three-species system in the chemostat. In this system, one species was a predator with a preference for one of two prey species. In the absence of the predator, the preferred prey species out-competed the second prey species, but the latter did better with increased predator abundance. The abundances of these three species, which were monitored over several weeks, fluctuated irregularly and did not reach a stable state. However, the system dynamics depended on the flow of the medium through the chemostat: at higher dilution rates, the system reached a stable state, whereas at lower rates, regular oscillations followed. In another interesting co-culture experiment, Harcombe[67] demonstrated that mutualism can be evolved experimentally. He co-cultivated a *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain together with an *Escherichia coli* strain that was unable to synthesize methionine. The two strains could not be grown together on acetate or glucose plates, but on lactose plates, cooperative (and costly) methionine excretion evolved in the *S.* Typhimurium strain, which benefited from the lactose degradation by-products that had been excreted by the *E. coli* mutant. Interestingly, the percentage of methionine-excreting *S.* Typhimurium bacteria was much higher in a structured environment (that is, the lactose plate) than in an unstructured environment (that is, the lactose liquid), presumably because the exchange of metabolites was more efficient in the plate than in the liquid.

Despite their versatility and amenability to a multitude of ecological questions, 'classic' co-culture experiments routinely involve only a small number of species. Just as high-throughput meta-omics-based profiling has boosted the field of computational microbial network inference, co-culture experiments are scaling up using novel technologies as well. However, several hurdles have to be taken towards achieving this goal.

The first hurdle to overcome for any high-throughput approach towards co-culturing is the 'uncultivable' microorganisms, which make up the majority of most communities. Cultivating uncultivable microorganisms is a question of finding the right conditions. Simulating their natural environment (which is called *in situ* cultivation) can substantially extend the portion of cultivable microorganisms[68]. Recently, high-throughput approaches towards *in situ* cultivation, such as the isolation chip[69], have been developed.

A second challenge, especially if more than two species are to be co-cultured, is identifying the community members and, ideally, quantifying their abundance during co-culturing to improve our understanding of the dynamics of the interactions. From such quantitative data, interaction strengths could be estimated, enabling the development of fully fledged dynamic community models (BOX 2).

Valm and colleagues[70] have developed a combinatorial labelling strategy based on fluorescence, which allowed simultaneously identifying 15 different genera in a human dental plaque sample. They currently use 8 fluorophores that can be combined pairwise to form 28 different colours. This combinatorial generation of colours makes the approach scalable: with 15 fluorophores, 105 binary combinations and 455 ternary combinations can be generated, which is enough to label complex microbial communities differentially.

Chaos
A type of dynamic behaviour that is characterized by irregular oscillations and a sensitivity to small differences in initial conditions, so that for two similar sets of start conditions, the system may behave entirely differently after some time has elapsed.

Stable state
A region in (multi-dimensional) space in which the system remains and to which the system returns after a small perturbation. The stable state may be a point (also referred to as stable equilibrium or stable steady state), a limit cycle (at which the system oscillates) or may have other shapes (for example, the strange attractors of chaotic systems).

© 2012 Macmillan Publishers Limited. All rights reserved

A third difficulty lies in the generation of a large number of artificial microbial communities that would allow analysing different species combinations in parallel.

This difficulty could be overcome with the microdroplet technology developed by Park and colleagues[71]. Using this technology, the authors could encapsulate mixtures of cells coming from different species in droplets and pack ~1,400 droplets into a small chip. As a proof of concept, they simulated a case of obligate mutualism with two auxotrophs, which only survived when injected together into a droplet. Park and colleagues noted that their co-culture device can even quantify the strength of the mutualistic relationship in terms of growth rate. Artificial microbial communities can also be spatially structured. For instance, in REF. 72, two mutualistic species are grown separately in the shell and core of fibres that are designed for co-culturing.

If these technological improvements in cultivation, labelling and parallelization are put together, a 'co-culture chip' may soon be available for high-throughput co-culturing and species-removal experiments. Because the number of potential interactions to test is still much too high for the average microbial community, high-throughput co-culturing needs to go hand in hand with network inference or other approaches to pinpoint candidates. Vice versa, high-throughput co-culturing could provide the benchmark data that are needed for the evaluation and improvement of network inference approaches. Network inference combined with co-culture experiments can deliver the growth rates and interaction strengths that are needed for mathematical simulations of the community.

Dynamic modelling of microbial interactions

Inferred networks can be considered to be static models of microbial communities, effectively providing a 'snapshot' of the community status at a given time. However, numerous phenomena (such as stability, perturbation and succession) can only be studied if the model describes how microbial populations change over time, which will require the use of dynamic models of microbial communities.

The stability and development of macro-ecological communities has been intensively studied with the aid of dynamic models[28,65,73,74]. The recent availability of metagenomics time series data[8,23,24,35,75] makes it worthwhile to consider these modelling techniques for microbial communities. Here, we briefly describe three potential key areas in which dynamic models will be able to provide important insights.

Community stability. Thanks to their rapid growth and short generation times, microorganisms lend themselves perfectly for studies on community stability and response to perturbation. Mathematical models of microbial communities can systematically assess the impact of species removal in order to pinpoint keystone species, to identify crucial thresholds beyond which a perturbation will irreversibly alter or destroy the community or to study chaotic behaviour. This is particularly relevant in, for example, the gut microbiome, in which there is substantial interest in investigating the effect of invading species or modulating the microbiota using pre- and probiotics.

Alternative stable states. Recently, alternative microbial communities were identified in the vaginal and gut ecosystems (that is, the enterotypes)[36,45]. The exact nature and underlying drivers of these alternative communities are of great interest. The possibility that ecosystems could support alternative stable states was first suggested by Lewontin[76] and has since then been discussed and tested in several ecosystems (reviewed in REF. 77).

Connell and Sousa[78] formulated three criteria that natural communities must fulfil to be considered as alternative stable states: each alternative state must be truly stable (at least until all community members have been replaced by their offspring); switches should take place only on perturbation; and alternative stable states must occur in the same environment. The first two criteria are difficult to validate in macro-ecological systems, but they are feasible for microbial communities. The difficulty lies with the third criterion: instead of representing different stable states, the communities could be driven by subtle differences in the environment that escaped detection. A telltale sign for a true multiple-stable-state system would be if, in repeated experiments, the community switch only occurs after perturbation and in the absence of any other changes. With a dynamic model of the community, the number of stable states in the system can be computed.

Further studies will need to be carried out to assess whether the alternative community configurations in the vaginal and gut ecosystems[36,45] might represent true alternative stable states. So far, induced enterotype switching has not been successful[79].

Microbial succession. Microbial succession is known to occur in, for instance, dental plaque biofilms (which form in several stages[44]), in plankton communities that vary with the seasons[23,24] and in the infant gut (which is sterile at birth[8,80]). Network inference can indirectly capture succession in the form of strong mutual exclusions but cannot distinguish between succession and other relationships (such as competition).

Many factors (for example, dispersal and random processes) shape microbial succession[81]. Mathematical models can help to study their relative importance by simulating the dynamics of communities with varying initial conditions and by comparing deterministic and stochastic models. Models can also investigate whether the end result of succession depends on the initial species present, their abundance and the order of settlement. For instance, in a data set that features microbial abundances in infant intestinal microbiota monitored over a year[8], Trosvik and colleagues[82] showed that the same stable-state phylum-level abundances were reached regardless of initial conditions but that the initial conditions had an impact on the time span needed to reach the stable state. Finally, a special challenge for models on succession is to capture the impact of evolution, as microorganisms are known to evolve novel relationships rapidly[67].

Succession
The orderly and predictable manner by which communities change over time following the colonization of a new environment.

© 2012 Macmillan Publishers Limited. All rights reserved

Challenges of dynamic microbial community modelling. Modelling microbial communities that consist of multiple species is still in its infancy and faces numerous challenges. First of all, models of real-world microbial communities have to handle a large number of species (or groups of species) and also have to take into account that important community members (such as bacteriophages) might not have been monitored at all. In addition, to be realistic, models have to take into account random processes[83] and to relax the assumption that interaction strengths between species are constant over time (because novel relationships can rapidly evolve). Also, the appropriate level at which to model the community has to be selected depending on the questions in mind (for example, at the species level versus at the pathway level). In some systems — for example, when microorganisms cooperate in biofilms — game theoretical modelling approaches may be worth exploring[84].

The most important challenge will be to integrate metabolic data into dynamic models. Static metabolic models have been already developed using flux balance analysis (FBA)[85,86]. However, FBA assumes that the community is at a stable steady state, which is likely to be an unrealistic assumption[87], and it requires high-quality metabolic reconstructions, which are hard to come by in a fully automated way (the most advanced automated metabolic reconstruction procedure still contains manual steps[88]). Although dynamic extensions of FBA have been made[89], the exploration of other modelling techniques that require less detail may be worthwhile.

Future directions in microbial modelling

The main application of microbial community modelling is the manipulation of microbial communities to enhance the abundances of beneficial species and functions and to suppress harmful ones. This constitutes an ecosystem-wide extension of synthetic biology (to 'synthetic ecology'): instead of engineering a single species, the goal is to engineer a whole community[90,91] by species removal or addition, manipulation of species abundances or by genetically engineering community members[92,93].

Manipulating microbial communities is relevant for waste water treatment[94], food production[29] and for the prevention and treatment of diseases such as caries[95], inflammatory bowel disease[96] and obesity[97,98]. For inflammatory bowel disease and obesity, the gut ecosystem is a prime target for modelling. However, few attempts have been made so far to develop models for the gut microbiota[99]. A microbial community model could be applied to simulate the treatment with pre- and probiotics, to find out whether the enterotype has an impact on the response to the treatment and to predict the duration that is necessary to achieve a stable shift towards a healthier state. A gut community model could be validated and refined with the new experimental systems that are being developed: for instance, with *in vitro* gut models[100] or gnotobiotic mice[101]. However, realistic models that could predict the effects of pre- and probiotics are still a long way off, because our knowledge of the interactions between the gut microbiota, the host and the metabolic processes in the gut is still incomplete.

Even if all of this knowledge were available, microbial community engineering would face daunting challenges. Microbial communities are highly complex, nonlinear, evolving systems that can be chaotic and therefore unpredictable. The long list of failed attempts to manipulate macro-ecological systems to control pests or invasive species should warn us that engineering ecosystems is a difficult endeavour.

However, the quick growth and short generation times in microbial ecosystems permit rapid and thorough model testing. Thanks to technological advances and constant methodological innovation, a bright future lies ahead for microbial ecological modelling, and many more fundamental insights can be expected in the coming years.

1. Lidicker, W. Z. A. Clarification of interactions in ecological systems. *Bioscience* **29**, 475–477 (1979).
2. Rodríguez-Martínez, J. M. & Pascual, A. Antimicrobial resistance in bacterial biofilms. *Rev. Med. Microbiol.* **17**, 65–75 (2006).
3. Woyke, T. *et al.* Symbiotic insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955 (2006).
4. Leschine, S. B. Cellulose degradation in anaerobic environments. *Annu. Rev. Microbiol.* **49**, 399–426 (1995).
5. Gause, G. F. *The Struggle for Existence* (Williams & Wilkins, 1934).
6. Raes, J., Foerstner, K. U. & Bork, P. Get the most out of your metagenome: computational analysis of environmental sequence data. *Curr. Opin. Microbiol.* **10**, 1–9 (2007).
7. Dubelaar, G. B. J. & Jonker, R. R. Flow cytometry as a tool for the study of phytoplankton. *Sci. Mar.* **64**, 135–156 (2000).
8. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5**, 1556–1573 (2007).
9. Lane, D. J. *et al.* Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl Acad. Sci. USA* **82**, 6955–6959 (1985).
10. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).
11. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557 (2005).
12. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
13. Cole, J. R. *et al.* The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37**, D141–D145 (2009).
14. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
15. Gonzalez, A. & Knight, R. Advancing analytical algorithms and pipelines for billions of microbial sequences. *Curr. Opin. Biotechnol.* **23**, 64–71 (2012).
16. De Smet, R. & Marchal, K. Advantages and limitations of current network inference methods. *Nature Rev. Microbiol.* **8**, 717–729 (2010).
17. Veiga, D. F. T., Dutta, B. & Balázsi, G. Network inference and network response identification: moving genome-scale data to the next level of biological discovery. *Mol. Biosyst.* **6**, 469–480 (2010).
18. Bonneau, R. *et al.* A predictive model for transcriptional control of physiology in a free living cell. *Cell* **131**, 1354–1365 (2007).
19. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).
20. Milns, I., Beale, C. M. & Smith, V. A. Revealing ecological networks using Bayesian network inference algorithms. *Ecology* **91**, 1892–1899 (2010).
21. Agrawal, R., Imielinski, T. & Swami, A. Mining association rules between sets of items in large databases. *ACM SIGMOD Record* **22**, 207–216 (1993).
22. Chaffron, S., Rehrauer, H., Pernthaler, J. & Von Mering, C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* **20**, 947–959 (2010).
23. Eiler, A., Heinrich, F. & Bertilsson, S. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* **6**, 330–342 (2012).
24. Steele, J. A. *et al.* Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* **5**, 1414–1425 (2011).
25. Fuhrman, J. A. Microbial community structure and its functional implications. *Nature* **459**, 193–199 (2009).
26. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive earth's biogeochemical cycles. *Science* **320**, 1034–1039 (2008).
27. Hecker, M., Lambeck, S., Toepfer, S., van Someren, E. & Guthke, R. Gene regulatory network inference: data integration in dynamic models—a review. *Biosystems* **96**, 86–103 (2009).
28. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1973).
29. Mounier, J. *et al.* Microbial interactions within a cheese microbial community. *Appl. Environ. Microbiol.* **74**, 172–181 (2008).

© 2012 Macmillan Publishers Limited. All rights reserved

30. Hoffmann, K. H. *et al.* Power law rank-abundance models for marine phage communities. *FEMS Microbiol. Lett.* **273**, 224–228 (2007).
31. Thingstad, T. F. Elements of a theory for the mechanisms controlling abundance, diversity and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45**, 1320–1328 (2000).
32. Rodriguez-Brito, B. *et al.* Viral and microbial community dynamics in four aquatic environments. *ISME J.* **4**, 739–751 (2010).
33. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).
34. Zhou, J. *et al.* Functional molecular ecological networks. *mBio* **1**, e00169–e00110 (2010).
35. Gilbert, J. A. *et al.* Defining seasonal marine microbial community dynamics. *ISME J.* **6**, 298–308 (2012).
36. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
37. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
38. Freilich, S. *et al.* The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res.* **38**, 3857–3868 (2010).
39. Ruan, Q. *et al.* Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22**, 2532–2538 (2006).
40. Fuhrman, J. A. & Steele, J. A. Community structure of marine bacterioplankton: patterns, networks, and relationships to function aquat. *Microb. Ecol.* **53**, 69–81 (2008).
41. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
42. Faust, K. & Sathirapongsasuti, J. F. *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* (in the press).
43. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
44. Kolenbrander, P. E. *et al.* Communication among oral bacteria. *Microbiol. Mol. Biol. Rev.* **66**, 486–505 (2002).
45. Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proc. Natl Acad. Sci. USA* **108**, 4680–4687 (2011).
46. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* **7**, e39315 (2012).
47. Morgan, J. L., Darling, A. E. & Eisen, J. A. Metagenomic sequencing of an *in vitro*-simulated microbial community. *PLoS ONE* **5**, e10209 (2010).
48. Raes, J. & Bork, P. Molecular eco-systems biology: towards an understanding of community function. *Nature Rev. Microbiol.* **6**, 693–699 (2008).
49. Aitchison, J. A concise guide to compositional data analysis. *Laboratório de Estatística e Geoinformação* [online], http://www.leg.ufpr.br/lib/exe/fetch.php/pessoais:abtmartins:a_concise_guide_to_compositional_data_analysis.pdf (2003).
50. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA* **103**, 12115–12120 (2006).
51. Reshef, D. N. *et al.* Detecting novel associations in large data sets. *Science* **334**, 1518–1524 (2011).
52. Kuczynski, J. *et al.* Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature Methods* **7**, 813–819 (2010).
53. Gotelli, N. J. Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621 (2000).
54. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
55. Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7** (Suppl. 1), S7 (2006).
56. Opgen-Rhein, R. & Strimmer, K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst. Biol.* **1**, 37 (2007).
57. Schaefer, J. & Strimmer, K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **21**, 754–764 (2005).
58. Schaefer, J. & Strimmer, K. A. Shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* **4**, 32 (2005).
59. Freilich, S. *et al.* Competitive and cooperative metabolic interactions in bacterial communities. *Nature Commun.* **2**, 589 (2011).
60. Brohée, S. *et al.* NeAT: a toolbox for the analysis of biological networks, clusters, classes and pathways. *Nucleic Acids Res.* **36**, W444–W451 (2008).
61. Hu, Z. *et al.* VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Res.* **37**, W115–W121 (2009).
62. Smoot, M., Ono, K., Ruscheinski, J., Wang, P. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
63. Lima-Mendez, G. & van Helden, J. The powerful law of the power law and other myths in network biology. *Mol. Biosyst.* **5**, 1482–1493 (2009).
64. Paine, R. T. A note on trophic complexity and community stability. *Am. Nat.* **103**, 91–93 (1969).
65. May, R. M. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. *Science* **186**, 645–647 (1974).
66. Becks, L., Hilker, F. M., Malchow, H., Jürgens, K. & Arndt, H. Experimental demonstration of chaos in a microbial food web. *Nature* **435**, 1226–1229 (2005).
67. Harcombe, W. Novel cooperation experimentally evolved between species. *Evolution* **64**, 2166–2172 (2010).
68. Kaeberlein, T., Lewis, K. & Epstein, S. S. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296**, 1127–1129 (2002).
69. Nichols, D. *et al.* Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. *Appl. Environ. Microbiol.* **76**, 2445–2450 (2010).
70. Valm, A. M. *et al.* Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc. Natl Acad. Sci. USA* **108**, 4152–4157 (2011).
71. Park, J., Kerner, A., Burns, M. A. & Lin, X. N. Microdroplet-enabled highly parallel co-cultivation of microbial communities. *PLoS ONE* **6**, e17019 (2011).
72. Kim, H. J., Du, W. & Ismagilov, R. F. Complex function by design using spatially pre-structured synthetic microbial communities: degradation of pentachlorophenol in the presence of Hg(II). *Integr. Biol.* **3**, 126–133 (2011).
73. Holling, C. S. Resilience and stability of ecological systems. *Annu. Rev. Ecol. Evol. Syst.* **4**, 1–23 (1973).
74. Ives, A. R. & Carpenter, S. R. Stability and diversity of ecosystems. *Science* **317**, 58–62 (2007).
75. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
76. Lewontin, R. C. The meaning of stability. *Brookhaven Symp. Biol.* **22**, 13–23 (1969).
77. Petraitis, P. S. & Dudgeon, S. R. Detection of alternative stable states in marine communities. *J. Exp. Mar. Biol. Ecol.* **300**, 343–371 (2004).
78. Connell, J. H. & Sousa, W. P. On the evidence needed to judge ecological stability or persistence. *Am. Nat.* **121**, 789–824 (1983).
79. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
80. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108**, 4578–4585 (2011).
81. Fierer, N., Nemergut, D., Knight, R. & Craine, J. M. Changes through time: integrating microorganisms into the study of succession. *Res. Microbiol.* **161**, 635–642 (2010).
82. Trosvik, P., Stenseth, N. C. & Rudi, K. Convergent temporal dynamics of the human infant gut microbiota. *ISME J.* **4**, 151–158 (2010).
83. Rosindell, J., Hubbell, S. P. & Etienne, R. S. The unified neutral theory of biodiversity and biogeography at age ten. *Trends Ecol. Evol.* **26**, 340–348 (2011).
84. West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. Social evolution theory for microorganisms. *Nature Rev. Microbiol.* **4**, 597–607 (2006).
85. Zomorrodi, A. R. & Maranas, C. D. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput. Biol.* **8**, e1002363 (2012).
86. Karlsson, F. H., Nookaew, I., Petranovic, D. & Nielsen, J. Prospects for systems biology and modeling of the gut microbiome. *Trends Biotechnol.* **29**, 251–258 (2011).
87. Hastings, A. Transients: the key to long-term ecological understanding? *Trends Ecol. Evol.* **19**, 39–45 (2004).
88. DeJongh, M. *et al.* Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinformatics* **8**, I39 (2007).
89. Hanly, T. J. & Henson, M. A. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol. Bioeng.* **108**, 376–385 (2011).
90. Curtis, T. P., Head, I. M. & Graham, D. W. Theoretical ecology for engineering biology. *Environ. Sci. Technol.* **37**, 64A–70A (2003).
91. Dunham, M. J. Synthetic ecology: a model system for cooperation. *Proc. Natl Acad. Sci. USA* **104**, 1741–1742 (2007).
92. Balagaddé, F. K. *et al.* A synthetic *Escherichia coli* predator–prey ecosystem. *Mol. Syst. Biol.* **4**, 187 (2008).
93. Ruder, W. C., Lu, T. & Collins, J. J. Synthetic biology moving into the clinic. *Science* **333**, 1248–1252 (2011).
94. Werner, J. J. *et al.* Bacterial community structures are unique and resilient in full-scale bioenergy systems. *Proc. Natl Acad. Sci. USA* 22 Feb 2011 (doi:10.1073/pnas.1015676108).
95. Marsh, P. D. Are dental diseases examples of ecological catastrophes? *Microbiology* **149**, 279–294 (2003).
96. Maloy, K. J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474**, 298–306 (2011).
97. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
98. Turnbaugh, P. J., Bäckhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
99. Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* **333**, 101–104 (2011).
100. Abbeele, P. V.d. *et al.* Microbial community development in a dynamic gut model is reproducible, colon region specific, and selective for bacteroidetes and clostridium cluster IX. *Appl. Environ. Microbiol.* **76**, 5237–5246 (2010).
101. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14 (2009).
102. Diamond, J. M. in *Ecology and Evolution of Communities* (eds Cody, M. & Diamond, J. M.) 342–444 (Harvard Univ. Press, 1975).
103. Connor, E. F. & Simberloff, D. The assembly of species communities: chance or competition? *Ecology* **60**, 1132–1140 (1979).
104. Woodcock, S. *et al.* Neutral assembly of bacterial communities. *FEMS Microbiol. Ecol.* **62**, 171–180 (2007).
105. Wootton, J. T. Field parameterization and experimental test of the neutral theory of biodiversity. *Nature* **433**, 309–312 (2005).
106. Horner-Devine, M. C. *et al.* A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* **88**, 1345–1353 (2007).
107. Sloan, W. T. *et al.* Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* **8**, 732–740 (2006).
108. Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
109. Borgelt, C. & Kruse, R. in *COMPSTAT 2002 — Proceedings in Computational Statistics: 15th Symposium held in Berlin, Germany, 2002* (eds Härdle, W. & Rönz, B.) 395–400 (Physica-Verlag, 2002).
110. Lallich, S., Teytaud, O. & Prudhomme, E. *Association Rule Interestingness: Measure and Statistical Validation* (eds Guillet, F. & Hamilton J. H.) (Springer, 2007).
111. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–61 (1960).
112. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

© 2012 Macmillan Publishers Limited. All rights reserved

113. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
114. van Dongen, S. Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Analysis Appl.* **30**, 121–141 (2008).
115. Clauset, A., Newman, M. E. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **70**, 066111 (2004).
116. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).

Acknowledgements
K.F. and J.R. are supported by the Research Foundation Flanders (FWO), the Flemish agency for Innovation by Science and Technology (IWT) and the Brussels Institute for Research and Innovation. We would like to acknowledge G. Lima-Mendez, S. Chaffron and all other members of the Raes laboratory, as well as D. Gonze for helpful comments and discussions. We would also like to apologize to all authors whose work could not be included owing to space restraints. In adition, we thank our reviewers, whose criticisms and suggestions helped to improve this Review.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Jeroen Raes's homepage: http://systemsbiology.vub.ac.be
In-house tool for microbial association network construction: http://systemsbiology.vub.ac.be/conet

© 2012 Macmillan Publishers Limited. All rights reserved