CHAPTER ONE

Learning Ecological Networks from Next-Generation Sequencing Data

Corinne Vacher[*,†,1], Alireza Tamaddoni-Nezhad[‡], Stefaniya Kamenova[§,¶], Nathalie Peyrard[‖], Yann Moalic[#], Régis Sabbadin[‖], Loïc Schwaller[,††], Julien Chiquet[‡‡], M. Alex Smith[¶], Jessica Vallance[§§,¶¶], Virgil Fievet[*,†], Boris Jakuschkin[*,†], David A. Bohan[‖‖]**

[*]INRA, UMR1202, BIOGECO, Cestas, France
[†]University of Bordeaux, UMR1202, BIOGECO, Pessac, France
[‡]Department of Computing, Imperial College London, London, United Kingdom
[§]INRA, UMR1349 IGEPP, Le Rheu, France
[¶]Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada
[‖]INRA, Toulouse, France
[#]Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Institut Universitaire Européen de la Mer (IUEM), Université de Bretagne Occidentale, Plouzané, France
[**]INRA, UMR518MIA, Paris, France
[††]AgroParisTech, Paris, France
[‡‡]LaMME, UMR CNRS 8071, Université d'Evry-Val d'Esonne, Evry, France
[§§]Université de Bordeaux, ISVV, UMR1065 Santé et Agroécologie du Vignoble (SAVE), Bordeaux Sciences Agro, Villenave d'Ornon, France
[¶¶]INRA, ISVV, UMR1065 SAVE, Villenave d'Ornon, France
[‖‖]UMR1347 Agroécologie, AgroSup/UB/INRA, Pôle Ecologie des Communautés et Durabilité des Systèmes Agricoles, Dijon Cedex, France
[1]Corresponding author: e-mail address: corinne.vacher@pierroton.inra.fr

Contents

Advances in Ecological Research, Volume 54
ISSN 0065-2504
http://dx.doi.org/10.1016/bs.aecr.2015.10.004

Abstract

Species diversity, and the various interactions that occur between species, supports ecosystems functioning and benefit human societies. Monitoring the response of species interactions to human alterations of the environment is thus crucial for preserving ecosystems. Ecological networks are now the standard method for representing and simultaneously analyzing all the interactions between species. However, deciphering such networks requires considerable time and resources to observe and sample the organisms, to identify them at the species level and to characterize their interactions. Next-generation sequencing (NGS) techniques, combined with network learning and modelling, can help alleviate these constraints. They are essential for observing cryptic interactions involving microbial species, as well as short-term interactions such as those between predator and prey. Here, we present three case studies, in which species associations or interactions have been revealed with NGS. We then review several currently available statistical and machine-learning approaches that could be used for reconstructing networks of direct interactions between species, based on the NGS co-occurrence data. Future developments of these methods may allow us to discover and monitor species interactions cost-effectively, under various environmental conditions and within a replicated experimental design framework.

1. INTRODUCTION

1.1 Ecological Interactions Are Drivers of Ecosystem Functioning

For a considerable part of the history of ecology, ecologists have tried to observe and explain the relationships between biodiversity and ecosystem functioning (BEF relationship) and how these change with environmental conditions and human-derived stressors (see Bohan et al., 2013; Cardinale

et al., 2012; chapter 'Towards an integration of biodiversity-ecosystem functioning and food-web theory to evaluate relationships between multiple ecosystem services' by Hines et al.; chapter 'Linking biodiversity, ecosystem functioning and services, and ecological resilience: towards an integrative framework for improved management' by Truchy et al.; Raffaelli et al., 2014). There is now a broad consensus that species diversity supports many ecosystem functions (Bohan et al., 2013; Cardinale et al., 2012). Ecological explanations of BEF have largely invoked two effects. First, a 'sampling effect' where highly productive species are more likely to be present in a diverse species pool and consequently function increases with diversity. Second, a 'complementarity effect' where the link between diversity and function is explained as a difference in species resource requirements (Loreau and Hector, 2001). Given that some species exploit resources more efficiently than others, then across increasingly diverse pools of species, functionality will tend to increase.

Species diversity not only supports ecosystem functions, but also influences the stability of ecosystems (Haddad et al., 2011; Loreau and de Mazancourt, 2013; Tilman et al., 2006). For system-level properties, such as total biomass/productivity and biological control, resilience of ecosystem function can arise with both the sampling and complementarity effects. As interacting species react in different ways to external stress, some species will benefit while others will not (see May, 1973). Across a portfolio of species within a diverse ecosystem, the stability of the entire portfolio will be greater than the fluctuations in the function of each species, due to an effect of averaging (Tilman et al., 1998). In high-diversity systems, stability across the portfolio of species is therefore preserved in the face of variation (insurance hypothesis, see Loreau et al., 2003; Yachi and Loreau, 1999).

These effects of biodiversity on ecosystem functioning are mediated by the ecological interactions between species (Duffy et al., 2007; Thébault and Loreau, 2006). For studying diverse systems of many interacting species, ecological network approaches have become the standard method (Ings et al., 2009; Lewinsohn et al., 2006; Pocock et al., 2012). These consider species as nodes, and the interactions between species (e.g. predator–prey interactions, host–parasite interactions, plant–pollinator interactions) as a series of links. The links may be weighted (quantitative ecological network) or not. Link weight is often defined the observed frequency of the interaction or the effect of the interaction on the performance of the interacting partners (Bascompte et al., 2006; Berlow et al., 2004; Laliberté and

Tylianakis, 2010). Building upon the BEF knowledge already acquired, the aim of network approaches is to understand how the network structure leads to the 'emergence' of ecosystem functions (see chapter 'Towards an integration of biodiversity-ecosystem functioning and food-web theory to evaluate relationships between multiple ecosystem services' by Hines et al.; Thompson et al., 2012). It has long been known that the provision of a specific ecosystem function may be maximized through management of the abundance of functionally important species (Gaston, 2010), such as pollinator honeybees (Calderone, 2012; Hagen et al., 2012). But what has become clear from network approaches is that the resilience of ecosystems and the stability of their functions rely directly on species diversity and the interactions between all species within the network (Naeem et al., 2009).

One interesting development in recent years has been a growing appreciation that ecological interactions and evolutionary dynamics can interact (Hairston et al., 2005). The structure of ecological networks and evolution can feedback, one upon the other. For instance, in plant–pollinator and plant–seed disperser mutualistic networks, network structure depends on co-evolution with and between partner species (Bascompte et al., 2006; Nuismer et al., 2013). In trophic networks (food webs), link connectance and the number of trophic levels can be affected through the evolution of body–size relationships (Loeuille and Loreau, 2006) and adaptive foraging (Beckerman et al., 2006). The evolutionary history can in turn affect network modularity and nestedness (Robinson et al., 2015; Vacher et al., 2008). These changes in ecological network structure, caused by the evolution of life-history traits, affect the functioning of ecosystems (Abrams and Matsuda, 2005; Loeuille et al., 2002), as well as their long-term stability and resilience (Kondoh, 2003; Loeuille, 2010a,b; Thébault and Fontaine, 2010).

1.2 Ecological Interactions Are Altered by Anthropogenic Activity

Anthropogenically accelerated biodiversity loss and larger global-scale environmental impacts have provoked much public concern and a drive to improve our management of the planet (Cardinale et al., 2012; Raffaelli and White, 2013; Rockström et al., 2009). Starting at the Rio 'Earth Summit' in 1992, this need to protect and monitor ecosystems and their functions have more recently coalesced around the concept of ecosystem services, as the ecosystem functions that directly benefits humanity (Millennium Ecosystem Assessment, 2005; see also chapter 'Detrital

dynamics and cascading effects on supporting ecosystem services' by Mulder et al.). There is, however, little agreement on, or clear understanding of, the best practices for preventing losses in species diversity, and of how networks of interacting species may be restored (but see Pocock et al., 2012). What we do have are expectations regarding how networks may change with environmental stressors, such as agricultural intensification (Albrecht et al., 2007; Laliberté and Tylianakis, 2010; Tylianakis et al., 2007) or biological invasions (Aizen et al., 2008; Albrecht et al., 2014; Heleno et al., 2009; Lopezaraiza-Mikel et al., 2007; Vacher et al., 2010). However, predicting *a priori* exactly how a particular ecological network metric (e.g. species richness, number of links, interaction strength, connectance, modularity, nestedness) will change in response to a particular stressor remains difficult (Heleno et al., 2012). Defining ecologically relevant network metrics (Blüthgen et al., 2008; Fortuna et al., 2010; Joppa et al., 2010; Leger et al., 2015) and predicting how they affect network structural stability and the persistence of the species within it (Bascompte et al., 2003; Montoya et al., 2006; Rohr et al., 2014; Saavedra et al., 2011; Thébault and Fontaine, 2010) is also very challenging.

Simply reducing or removing anthropogenic stressors may not be enough to restore the structure of ecological networks and the functioning of ecosystems. For example, in a recent grassland experiment, plant diversity in plots that received high rates of nitrogen for 10 years had not recovered to control levels 20 years after the nitrogen inputs had stopped (Isbell et al., 2013). This suggests that 'turning back the clock' to a more benign set of management practices, even where this is feasible, might not achieve restoration goals. There is, indeed, increasing evidence that network structure and dynamics modulates the trajectory and rate of change of the response to stressors, with time lags and the presence of multiple alternative stable states, due to ecological inertia in the food web. Alternative stable states have been suggested as the reasons for the slow or non-existent biological recovery of commercial marine fisheries following reductions in fishing effort, and of freshwaters following acidification (Layer et al., 2010, 2011) and eutrophication (Scheffer et al., 2001).

1.3 Next-Generation Sequencing Can Be Used for Monitoring Ecological Interactions

While ecological networks play a central role in the development of basic and applied ecological science, most empirical networks remain at best

incomplete due to persistent methodological shortcomings that hinder our capacity to resolve species (i.e. network nodes) and their interactions (i.e. network links) (Ings et al., 2009). Resolving ecological networks require considerable (and often prohibitive) investments of time and resources to observe and sample the organisms present within an ecosystem, identify them to the species level and then characterize the possible links that exist between those species. Quantifying the importance of these links, in terms of frequency and effect on the performance of the interacting partners, is even more difficult. This means that relatively few well-resolved ecological networks have been constructed, there is little replication amongst those networks (see Pocock et al., 2012), and ultimately our understanding of network-based BEF and ecosystem change is limited.

Here, we speculate that reconstructing ecological networks directly from environmental DNA (eDNA), by combining next-generation sequencing (NGS; high-throughput sequencing) (Di Bella et al., 2013) and theoretical approaches, such as statistical modelling (Faust and Raes, 2012) and machine learning (Bohan et al., 2011), will alleviate these methodological and financial constraints. Our arguments are the following: (1) NGS platforms generate several millions of DNA sequences for a few hundred dollars (Liu et al., 2012; Quail et al., 2012); (2) they permit the characterization of DNA diversity in complex environmental samples (e.g. soil, water, plant tissues, faeces, pellet, gut content, etc.) containing hundreds of microbial species and the imprint of many macro-organisms; (3) many of these organisms can be identified at the species level by using GenBank or reference taxonomic databases such as Greengenes, BOLD, SILVA and UNITE (Abarenkov et al., 2010; DeSantis et al., 2006; Kõljalg et al., 2005; Quast et al., 2013; Ratnasingham and Hebert, 2007) and (4) methods are being developed to predict species interactions from their abundance patterns and additional information such as their functional traits or the features of the environmental samples (Bohan et al., 2011; Deng et al., 2012; Faust and Raes, 2012; Kurtz et al., 2015). With these advancements, the potential of NGS techniques for resolving complex ecological networks is enormous. Below we review the types of ecological interactions for which NGS data are the most relevant, present some examples of NGS-based ecological networks and give an insight into the theoretical approaches that may be applied to NGS data to highlight ecological interactions.

2. WHY LEARNING ECOLOGICAL NETWORKS FROM NGS DATA?

2.1 Limitations of Classical Methods for Resolving Ecological Interactions

Network ecology primarily emerged based on the observation of interactions between macro-organisms (e.g. plant–pollinator, plant–herbivore, plant–seed disperser, anemone–fish; see IWDB, https://www.nceas.ucsb.edu/interactionweb/resources.html). These observations require considerable time and resources and have limitations. Smaller species, such as microbes and parasites, have often been overlooked (Lafferty et al., 2008) despite their importance for ecosystem functioning (Ducklow, 2008; Gilbert and Neufeld, 2014; Hudson et al., 2006). As a consequence, integration of macro-organisms and micro-organisms resolved at the species level into the same networks is far from the norm (but see Vacher et al., 2008, 2010). Short-term interactions such as those between predator and prey are also difficult to observe. This lack of completeness and integration of ecological networks critically limits our understanding of BEF and should be resolved (Fontaine et al., 2011).

Micro-organisms have not been fully integrated in network ecology largely due to the enormous difficulty associated with identifying the interacting microbes. Microbial interactions have long been studied using laboratory culture-dependent methods. However, testing the hundreds of combinations of culture conditions necessary to find the conditions required to grow successfully a single microbial species is often prohibitive (Lok, 2015). As a consequence, an estimated 85–99% of bacteria and archaea cannot yet be grown in the lab, drastically limiting our knowledge of microbial life. Culturing is not only a (inadvertently) selective environment, but also a labour intensive and tedious task that cannot be extended to the whole microbial community in a given environment. Co-culture experiments are traditionally used to detect antagonistic or mutualistic interactions between microbial taxa but the extrapolation of results to natural conditions is risky (Sher et al., 2011). High-throughput cultivation platforms, such as microfluidic chips and iChips, are currently revolutionizing this field of research. They are based on complex cultivation media, closer to those of the natural environment and may support multiple microbial species (Lok, 2015). Culturing and co-culturing microbes

nevertheless remain a complicated task (Haruta et al., 2009). Many scientists have, therefore, chosen to bypass it entirely and move to sequencing directly microbial DNA.

Many scientists studying trophic networks have also chosen to use DNA sequencing for characterizing predator–prey or host–parasitoid interactions. The visual and microscopic examination of guts, faeces, or pellets (Hengeveld, 1980; Pisanu et al., 2011) and the rearing of parasitoids from hosts (Eveleigh et al., 2007; Müller et al., 1999) are indeed time consuming, labour intensive and highly dependent on the observer's experience. They tend to overlook cryptic species and are very difficult to compare between systems or researchers (reviewed by Symondson, 2002). Direct methods have expanded recently to include the analysis of ratios of naturally occurring stable isotopes (mostly carbon and nitrogen), permitting the integration of trophic information over an extended timescale (several days to several months) (Boecklen et al., 2011). Nevertheless, the technique can only be applied to a small number of prey species (Moore and Semmens, 2008) and often does not permit obtaining accurate trophic data at individual level (Vanderklift and Ponsard, 2003). Other biomarkers, such as fatty acids (reviewed by Traugott et al., 2013) have similar constraints.

2.2 Advantages of NGS for Identifying Species and Their Interactions

The emergence of NGS techniques and associated bioinformatic pipelines has boosted the characterization of species and their interactions over the last 10 years. In particular, it has boosted the characterization of microbial communities (Di Bella et al., 2013; Hibbett et al., 2009) and facilitated the quantification of trophic links (Smith et al., 2011; Staudacher et al., 2015). NGS indeed overcomes the limitations of culture-dependent methods and provides a more thorough description of microbial communities. We are now, in principle, able to detect any microbial organism in nature, even those that cannot be isolated or grown in the lab. The first step is to collect environmental samples (e.g. samples of soil or water, tissues of plant, animal, etc.) and extract total DNA from these samples. The taxonomic description of microbial communities with NGS techniques then relies on the amplification and sequencing of DNA barcodes (Chakraborty et al., 2014). Similarly, the sequencing of faeces, pellet, or gut content gives an insight into the diet of an organism with little or no need for *a priori* information about the

target prey species (e.g. Boyer et al., 2013; Brown et al., 2012; Quéméré et al., 2013; Shehzad et al., 2012). Moreover, the use of tags (i.e. unique identifiers) to recover data from each sample after sequencing (Clarke et al., 2014; Taberlet et al., 2012) facilitates the mass screening of several hundreds of individual samples (e.g. Kartzinel et al., 2015; Mollot et al., 2014).

A DNA barcode, in its simplest definition, is one or small number of short genetic sequences taken from a standardized portion of the genome that are used to identify species (Schlaeppi and Bulgarelli, 2015). Different barcode regions are used for different taxonomic groups. The most popular DNA barcode for identifying fungal species is the internal transcribed spacer (ITS) region of the nuclear ribosomal repeat unit (Schoch et al., 2012), except for Glomeromycota (Öpik et al., 2014; Stockinger et al., 2010). The 16S small ribosomal subunit gene (16S rRNA) is the 'gold standard' for characterizing bacterial community composition (Sun et al., 2013), although other barcode regions have been proposed (Chakraborty et al., 2014). The 16S rRNA gene contains conserved regions as well as nine hypervariable regions (V1–V9). One or more of these hypervariable regions are usually sequenced (Gloor et al., 2010; Sun et al., 2013). Universal barcode genes are also available for algae and protozoa, but, to our knowledge, not for viruses (Chakraborty et al., 2014). The standardized barcode region for eukaryotic animals is part of the cytochrome *c* oxidase 1 (CO1) mitochondrial gene and a growing library of identified specimens exists (see BOLD, http://www.boldsystems.org/; Ratnasingham and Hebert, 2007). In the case of trophic ecology, the best barcode depends upon the model organism, its likely preys, and the question being addressed (Pompanon et al., 2012). Generally, as dietary samples are complex mixtures of highly degraded DNA, the best barcode should target very short DNA fragments with the same amplification efficiency across very distantly related taxa. These constraints sometimes preclude species-level taxonomic assignment of preys as the probability to encounter variable sites decreases with the sequence length. Therefore, classical DNA-based methods such diagnostic PCR followed by Sanger sequencing remains very useful for the quantitative assessment of predator–prey (Clare et al., 2009; Rougerie et al., 2011) and host–parasitoid interactions (Condon et al., 2014; Kaartinen et al., 2010; Smith et al., 2008). These molecular-based techniques have been used successfully to reveal previously unseen trophic interactions (Derocles et al., 2015; Wirta et al., 2014).

3. EXAMPLES OF NGS-BASED ECOLOGICAL NETWORKS AND THEIR APPLICATIONS

3.1 Deciphering Pathobiomes Using NGS-Based Microbial Networks for Improving Biological Control

Population growth and changes in dietary habits are likely to double human demand for food in the coming decade. However, in addition to these issues, there are inherent inefficiencies in modern agriculture where plant pathogens can claim 10–16% of the global food harvest. Reducing this percentage has become, therefore, a priority for achieving food security (Chakraborty and Newton, 2011). Biodiverse ecosystems can support natural antagonists of pathogens and they have been argued to limit or prevent disease development (Berendsen et al., 2012; Penton et al., 2014). The notion that the pathogens live and interact with other organisms in their environment has led to the development of the pathobiome concept (i.e. the pathogen species integrated within its biotic environment; Vayssier-Taussat et al., 2014). Elucidating the components of the pathobiome is one of the prerequisites for understanding the persistence, transmission and evolution of pathogen species, and for improving methods of biological control with naturally occurring antagonists.

Microbial networks give insight into the relationships between a pathogen species and the other micro-organisms interacting with the host species. Their study may enable the identification of potential pathogen antagonists and may reveal network topological properties driving the stability of the residential microbial community, including its invasibility by pathogens (Desprez-Loustau et al., 2015; Kemen, 2014). Microbial networks typically represent the patterns of co-existence between microbial taxa, irrespective of the underlying mechanism. Nodes represent microbial taxa and links indicate spatial or temporal associations between taxa. Such networks are referred to as microbial co-occurrence networks (Aires et al., 2015; Barberán et al., 2012; Kara et al., 2013; Navarrete et al., 2015), microbial association networks (Chow et al., 2014; Faust and Raes, 2012; Fuhrman, 2009), microbial correlation networks (Duran-Pinedo et al., 2011; Friedman and Alm, 2012), or networks of co-existing microbes (Chaffron et al., 2010). *Direct ecological interactions* may account for the observed co-existence patterns. As with macroscopic organisms, microbes interact with each other directly, through pairwise ecological interactions such as competition, predation, parasitism, mutualism, commensalism, or

amenalism (Faust and Raes, 2012). These interactions can involve micro-organisms belonging to the same kingdom or different kingdoms (Frey-Klett et al., 2011). Competition, which is considered to be the dominant type of microbe–microbe interaction (Foster and Bell, 2012; Hibbing et al., 2010), is expected to produce co-exclusions between microbial taxa. Amensalism is also expected to produce co-exclusions while mutualism and commensalism are expected to yield co-associations between microbial taxa. The outcome of parasitism or predation, in terms of spatial or temporal association, is less straightforward to predict, with the consumer relying on, but potentially depleting, its prey. Associations between microbial taxa may also be explained by two other mechanisms: because they both depend on the presence of a third microbial taxa (*indirect ecological interaction*) or because they are adapted to similar environmental conditions (*shared environmental preference*) (Ovaskainen et al., 2010).

Here, we reconstructed the microbial association network (Fig. 1) of the leaves of an oak tree (*Quercus robur* L.) susceptible to the fungal pathogen *Erysiphe alphitoides*, the causal agent of oak powdery mildew (Mougou et al., 2008). Forty leaves were sampled and DNA was extracted from four 0.5 cm^2 discs per leaf. Fungal taxa were characterized by 454 pyrosequencing of the ITS1 region (as described in Cordier et al., 2012). Bacterial taxa were characterized by sequencing the hypervariable V6 region of the 16S rRNA gene on an Illumina platform (as described in Gloor et al., 2010). The taxonomic composition of the samples was obtained by filtering and clustering the sequences using QIIME (Caporaso et al., 2010) and Usearch (Edgar, 2013). The software SparCC (Friedman and Alm, 2012) was then used to compute the correlations between the abundances of microbial taxa across samples. Only significant positive correlations (co-associations) and significant negative associations (co-exclusions) were retained as links in the microbial network. The network was visualized using Gephi (Bastian and Heymann, 2009). Its analysis revealed that *E. alphitoides* has a particular connectivity behaviour, as it is predominantly connected to the network through strong negative links. The presence of the pathogen is thus associated to the absence of other micro-organisms. To go further into the mechanistic interpretation of this network, the microbial association network should be turned into a microbial interaction network having only *direct* ecological interactions as links. This may be done by firstly using environmental conditions to screen out associations that indicate a *shared environmental preference* and, secondly, using network inference methods to predict *direct* links between microbial taxa (Faust and Raes, 2012; Kurtz et al., 2015;

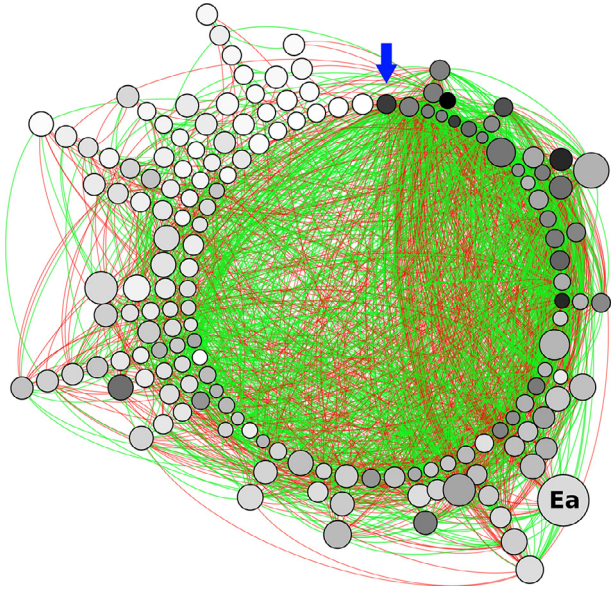

Figure 1 Microbial association network of the leaves of an oak tree (*Q. robur* L.) susceptible to the foliar fungal pathogen *E. alphitoides* (Ea). Each node represents a microbial taxon (either bacterial or fungal) and each link represents a significant correlation between their abundances. Red (grey in the print version) and green (grey in the print version) links indicate co-exclusions and co-associations, respectively. The arrow indicates the node with the highest degree (i.e. the highest total number of links). Degree decreases clockwise, with nodes stacked on the same line having the same degree. The size of the nodes is inversely proportional to the sum of the correlation coefficients: larger nodes have more number and/or stronger negative associations. Darker nodes have higher betweenness centrality (calculated on the absolute values of associations), suggesting that they are topological keystone taxa. *E. alphitoides* is predominantly connected to the network through strong negative links (co-exclusions) but is not a good candidate for topological keystone species.

Schwaller et al., 2015). Such network would allow us to identify the microbial species that directly impede or facilitate pathogen development.

3.2 Studying the Hologenome Theory of Evolution Using NGS-Based Microbial Networks

Humans are drastically and rapidly altering the environment, including climate, and many species may not be able of adapting quickly enough to these new conditions (Carroll et al., 2014). The mismatch between the phenotype of agriculturally important plants and new climate is an overarching

challenge in forestry, agriculture and conservation biology. A widespread debate concerns whether to use local versus external sources of genetic material for replanting to best anticipate responses to climate change (Carroll et al., 2014). The hologenome theory of evolution is relevant to this debate. It posits that the holobiont (the plant or animal with all of its associated micro-organisms) is a unit of selection in evolution. According to this theory, host and 'associates' genomes act as a consortium that copes with environmental change by promoting adaptation and evolving as a whole. During periods of rapid environmental change, a diverse microbial symbiont community can aid the holobiont in surviving, multiplying and 'buying the time' necessary for the host genome to evolve (Zilber-Rosenberg and Rosenberg, 2008). There is indeed empirical evidence that host speciation is under the influence of its interactions with micro-organisms (Janson et al., 2008) and that microbial communities can be one determinant shaping the adaptation and evolution of higher organisms (Dittami et al., 2015; Rosenberg et al., 2007) .

NGS techniques are useful tools to characterize the tight associations of eukaryotes with microbes. In a recent study, they were used to characterize the internal microbial communities of the algae *Caulerpa* at the scale of the Mediterranean Sea (Aires et al., 2015). These endophytic microbial communities are hypothesized to play important roles in development, defence and metabolic activities of their host algae. Three *Caulerpa* species were sampled and the bacterial lineages within each sampling unit were identified through high-throughput sequencing of the hypervariable V4 region of the 16S rRNA gene. Then, networks of co-occurring bacterial lineages were analyzed by combining percolation theory (Moalic et al., 2012; Stauffer and Aharony, 1994) and community detection algorithms (Fortunato, 2010; Lancichinetti and Fortunato, 2009; Leger et al., 2014). Here, the Bray–Curtis index was used to assess the similarity in the distribution of bacterial lineages across samples: two lineages are linked if their distribution similarity is higher than 0.62, which is the percolation threshold of the network. Second, modules were delimited by using the leading eigenvector algorithm (Fig. 2). The results revealed that a very large fraction of the bacterial community is species-specific, even in areas where distinct *Caulerpa* species occur in sympatry. These species-specific bacterial lineages account for the modular structure of the co-occurrence network. Such specificity of endophytic bacterial communities is coherent with the hologenome theory of evolution. As several *Caulerpa* taxa have extended their range through invasion in different parts of the world, including the Mediterranean Sea, future

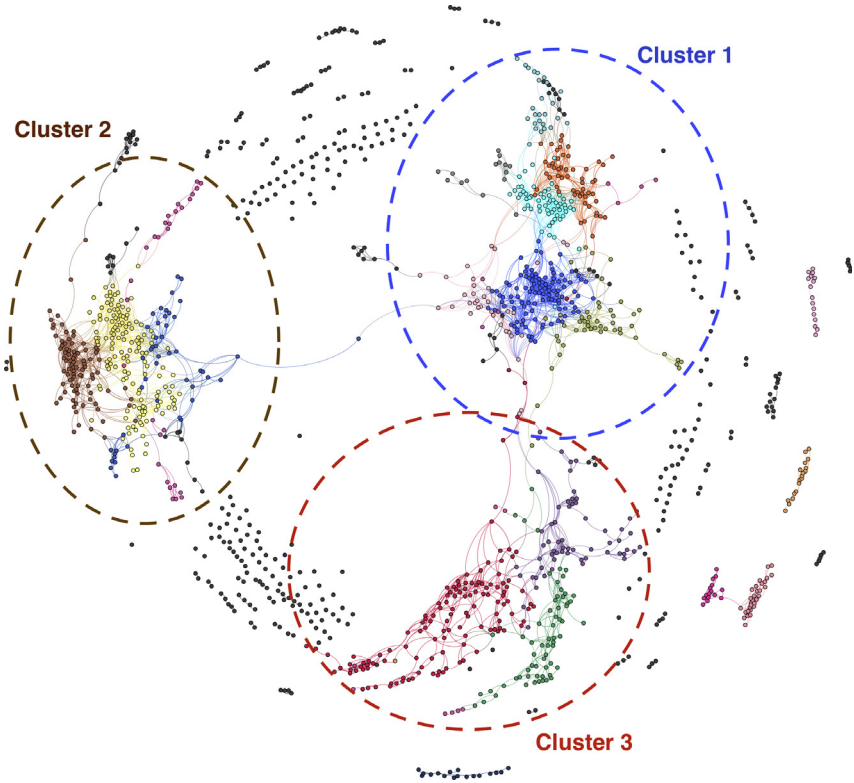

Figure 2 Associations between the endophytic bacterial lineages of the algae *Caulerpa* at the scale of the Mediterranean Sea. Each node is a bacterial lineage. Two lineages are linked if they tend to co-occur in the same samples. The Bray–Curtis index was used to assess the similarity in their distribution across samples: two lineages are linked if their distribution similarity is higher than 0.62, which is the percolation threshold of the network. Nodes are coloured according to their belonging to a module. Modules were delimited by using the leading eigenvector algorithm. The modules form three larger clusters that are, respectively, dominated by bacterial lineages specific of *C. prolifera*, *C. racemosa var. turbinata-uvifera*, and *C. racemosa var. cylindracea*.

studies should investigate the role of these microbial communities on the invasiveness of their host.

3.3 Testing the Niche Partitioning Theory with NGS-Based Trophic Networks

The niche partitioning theory is central to our understanding of biodiversity. The term niche partitioning refers to the process by which natural selection drives competing species into different patterns of resource use or different

niches (Hector and Hooper, 2002; MacArhur, 1958). This differentiation of ecological niches reduces competition and promotes co-existence between species (Chesson, 2000; Levine and HilleRisLambers, 2009). NGS techniques promise to bring significant changes in our understanding of niche partitioning because they provide information about the entire diet range of a species, while also highlighting new and unexpected trophic links. For instance, Ibanez et al. (2013) carried out a trait-based estimation of the trophic niche width of four species of grasshoppers through choice experiments and NGS study of their diet. They showed that observed trophic niche breadth in generalist herbivorous insects depends on both species-specific food preferences and habitat diversity, and that it is not an intrinsic property of the species as usually considered in theoretical studies. Kartzinel et al. (2015) also used NGS techniques, in combination with stable isotope analyses, to investigate trophic niche partitioning among sympatric large mammalian herbivores in Kenya. They observed unsuspected fine-scale resource partitioning even between species within the same trophic guild. Consequently, through this study, NGS methods illuminated mechanisms behind the large diversity of co-existing herbivorous mammal species observed in African savannas.

Here, we used NGS techniques to resolve the trophic interactions in a community of carabid beetles inhabiting European arable landscapes. These beetles significantly contribute to the biological control of pests (Kromp, 1999) but their contribution is unpredictable given their broad diet spectrum including alternative preys such as other natural pest enemies. We used NGS approach in order to investigate changes in carabid diet among 14 common arable species by analyzing the prey DNA contained in their guts. The carabid beetles were sampled in six arable fields in two different agroecosystems in Brittany, France. To cover the whole carabid diet spectrum, four barcode genes were combined, including mitochondrial 16S and COI for characterizing animal prey (Bienert et al., 2012; De Barba et al., 2014) and the chloroplast trnL for detecting consumed plant species (Taberlet et al., 2007). We obtained bipartite ecological networks showing an extensive degree of niche overlapping between carabid species (Fig. 3), confirming their generalist feeding behaviour. The use of NGS also allowed us for the first time to quantify carabid community contribution to ecosystem services (the consumption of several animal and plant pest species) and dis-services (the consumption of other service-providing organisms) (Fig. 4). These findings suggest that important ecological functions as pest regulation and intraguild predation may not be associated with the identity of particular species but are

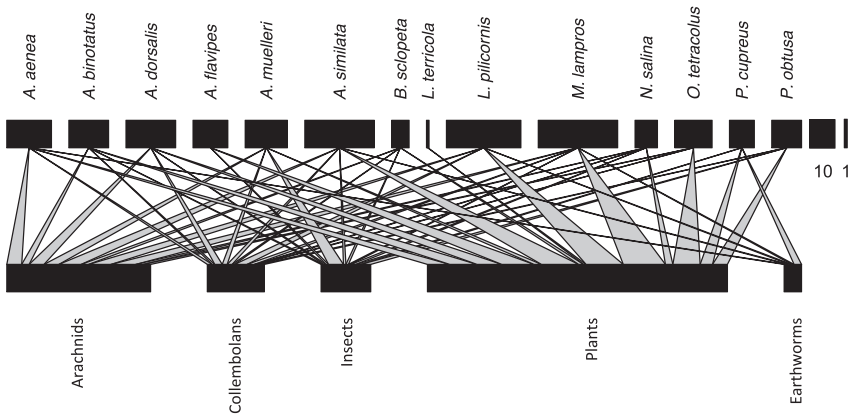

Figure 3 Quantitative predator–prey network recovered by analysing gut contents of 1414 carabid species using an NGS approach. Lower bars represent prey category (arachnids, collembolans, insects, plants, and earthworms) abundance and upper bars represent the abundance of carabid species positive for at least one prey category. Scales are indicated at right. Linkage width indicates frequency of each trophic interaction. Carabid beetles were sampled in six different fields.

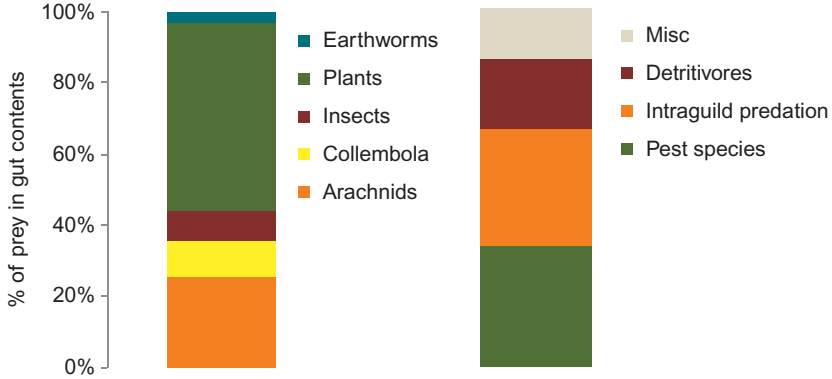

Figure 4 The relative proportions of five major prey categories (arachnids, collembolans, insects, plants, and earthworms) recovered in gut contents of 503 individuals from 14 carabid species using NGS meta-barcoding approach. Using their taxonomic identity, all animal and plant preys were assigned into three functional categories (pest species, detritivore species, and other natural predators). All preys that did not fit any functional category (e.g. tree or non-pest plant DNA) were grouped into a forth category (miscellaneous). This approach allowed us to quantify the relative contribution of the carabid community to ecosystem services (i.e. pest control) and dis-services (i.e. control of other service-providing organisms) within agroecosystems.

more likely an emergent property of the community probably modulated by extrinsic environmental factors. Further investigations are required for see how important agronomical features such as the cropping system or management intensity impact the structure of the trophic web.

3.4 Challenges to Be Addressed to Get Predictive Insights from NGS-Based Networks

In the three-aforementioned examples, the NGS-based ecological networks were static (Figs. 1–3). To predict the impact of environmental stressors on ecosystem services, future research should move from static networks to dynamic ones. In particular, it should explore whether time-varying networks have 'early warning' properties that can predict abrupt ecological transitions or 'tipping points' (Dakos and Bascompte, 2014; Faust et al., 2015). Metagenomic time series data can provide useful information on the dynamics of microbial communities (Fuhrman et al., 2015). Future research should also go beyond community composition by integrating data about functional gene expression, transcripts, proteins and metabolites. These data will yield crucial information, in addition to current co-occurrence patterns, about 'who?' does 'what?' and 'when?' within the community (Faust et al., 2015).

Future research should also aim at validating the ecological interactions learned from NGS data. In the first example (Fig. 1), co-association and co-exclusion patterns indicate which pairs of microbial taxa are the most likely to interact together. But experimental validation of these interactions is required. Reference networks, with well-known ecological interactions, are necessary to assess the relevance of the methods of network reconstruction. For instance, a machine-learning methodology to reconstruct networks from co-occurrence data and species functional traits has been demonstrated (Bohan et al., 2011). This produced hypothetical food webs in an agricultural system that bore all the hallmarks of real food webs and were validated against both the literature and direct molecular biological data for specific trophic interactions (Davey et al., 2013). Other methods have also been used successfully to reconstruct known ecological networks based on species abundance data (Aderhold et al., 2012; Faisal et al., 2010; Milns et al., 2010). The current challenge is to test all these methods on occurrence or abundance data derived from NGS datasets, by using a well-resolved empirical network as a reference.

Finally, future research must continue to improve the qualitative and quantitative reliability of NGS data, while keeping the costs down. Future

research must continue to evaluate and reduce the biases in species composition due to primers choice and amplification by PCR (Berry et al., 2011; Gonzalez et al., 2012; Lee et al., 2012; Patin et al., 2013; Pinto and Raskin, 2012), to characterize the noise profiles of the various sequencing platforms and to develop bioinformatic pipelines for taking them into account (Gaspar and Thomas, 2013; Gilles et al., 2011; Reeder and Knight, 2011). Future research must also develop new statistical methods for estimating species richness and diversity from NGS data (Haegeman et al., 2013) and improve the taxonomic assignment of the species. For instance, the combination of several barcodes targeting restricted groups of organisms could improve the taxonomic resolution of preys in dietary samples (e.g. Deagle et al., 2009). However, combining of several barcode markers could increase considerably the costs of NGS techniques. Multiplexing several barcodes within the same PCR would help improving the cost effectiveness (De Barba et al., 2014), but these still require time-consuming methodological development. Additionally, the direct comparison of the relative proportion of prey groups, by comparison of their barcode sequence counts, is highly questionable due to the step of variation in amplification among barcodes by PCR (Taberlet et al., 2012). The rapid development of direct shotgun sequencing techniques, which requires no PCR amplification step, promises to resolve these problems. In two recent studies, the shotgun sequencing of long fragments of nuclear, mitochondrial and chloroplast DNA fragments from an herbivorous mammal (Srivathsan et al., 2014) and a carnivorous coleopteran (Paula et al., 2015) provided reliable and quantitative species-level identification of their consumed preys. The shotgun sequencing simultaneously revealed detailed information about the host's genetics, as well as information about the diversity of host's parasites and bacterial symbionts. These findings open several exciting perspectives for merging multiple types of ecological interactions in the same network (Fontaine et al., 2011; Kéfi et al., 2012), or for assessing the impact of host genetic differentiation or speciation into ecological network dynamics (e.g. Robinson et al., 2015).

4. THEORETICAL METHODS FOR DECIPHERING ECOLOGICAL NETWORKS FROM NGS DATA

4.1 The Input Data

Meta-barcoding, also referred to as amplicon-based community profiling, generates a list of operational taxonomic units (OTUs) and their distribution in the environmental samples. Each OTU is a group of related sequences

(with a similarity percentage higher than 97%, in most cases) (Schlaeppi and Bulgarelli, 2015). Taxonomic assignment of OTUs is usually performed by comparing the most abundant sequence of each OTU to the sequences deposited in GenBank or in curated databases such as Greengenes, BOLD, SILVA or UNITE (Abarenkov et al., 2010; DeSantis et al., 2006; Kõljalg et al., 2005; Quast et al., 2013; Ratnasingham and Hebert, 2007). The ecological relevance of the number of sequences per OTU and per sample is a matter of debate as the quantitative link between sequence reads and species relative abundance is not always straightforward due to methodological biases. The storage of environmental samples, before DNA extraction, may cause biases in the community composition revealed by NGS technique (U'Ren et al., 2014). In the case of gut samples, differences in digestion rates between types of prey may also cause bias (Deagle et al., 2009). Finally, whatever the type of environmental sample, variations in the amplification efficiency of the barcode region among species may distort the community composition (Berry et al., 2011; Gonzalez et al., 2012; Lee et al., 2012; Patin et al., 2013; Pinto and Raskin, 2012). Studies of plant pathogen species have, however, shown that pathogen abundances measured by visual symptoms correlate with their relative abundance in NGS datasets (Sapkota et al., 2015). The number of sequences of a given OTU in a given sample therefore contains some information on the OTU abundance in that sample, despite potential biases introduced by PCR (Cotton et al., 2014). This is important, because associations or interactions between OTUs are often derived from sequence counts (Faust and Raes, 2012; Friedman and Alm, 2012; Kurtz et al., 2015).

The datasets required for reconstructing ecological networks from NGS data are the OTU table, giving the estimated abundance of each OTU in each sample, with or without some additional information on the OTUs (e.g. functional traits of the corresponding species) and on the environmental samples (e.g. host species, abiotic conditions at the time of sampling) (Table 1). Association networks can be derived from the OTU table, as shown in the two first examples (Figs. 1 and 2). In these examples, the nodes of the association networks were fungal and/or bacterial OTUs. The links were the significant correlations between their abundance across samples (Fig. 1) or the similarity in their distribution across samples (Fig. 2). In the first example, links were positive or negative depending on the correlation sign. Such correlations should not be calculated based on raw sequence counts because variations in the total number of sequences per sample (a technical bias) may yield spurious results. Correlations between normalized

Table 1 Description of the Data Entering in the Construction of NGS-Based Ecological Networks, with Their Names in the Two Main Fields of Research (Statistical Inference of Networks and Logic-Based Machine-Learning Algorithms)

Data Description	Type of Data	Name Given in Statistical Modelling	Name Given in Logic-Based Machine Learning
Measured occurrence or abundance of species/OTUs in the studied sites/samples	Input	Observed variables	Observable predicates
Additional information on the species/OTUs (e.g. functional traits) or on the sites/samples (e.g. abiotic environment)	Input	Covariates	Background knowledge
Rules of interaction (e.g. species/OTU *X* can eat *Y* if *X* is bigger than *Y*)	Input or output	Constraints on the space of possible networks to explore	Logical rules which can be part of background knowledge or can be learned from input data
Hypothetical relationships between species/OTUs that can be visualized as a network	Output	Inferred edges	Abduced links

counts may also yield spurious results and several methods have been proposed to circumvent this issue (Deng et al., 2012; Faust and Raes, 2012; Friedman and Alm, 2012; Kurtz et al., 2015).

Such networks representing *spatial or temporal associations* between microbial OTUs are not entirely satisfactory because they do not match the mechanistic framework typically used in network ecology. As mentioned above, network ecology developed mainly on the observation of *direct ecological interactions* between several sets of macro-organisms (e.g. predator–prey, plant–pollinator, plant–herbivore, plant–seed disperser, anemone–fish interactions) (Ings et al., 2009). The predator–prey network recovered by using NGS techniques to describe gut contents (Fig. 3) is similar in terms of interaction type, but not the microbial association networks of the two first examples (Figs. 1 and 2). To go beyond these association networks, the first challenge is to distinguish associations due to *direct ecological interactions* and those due to *indirect ecological interactions* (Kurtz et al., 2015; Schwaller et al., 2015). The second challenge is to take into account environmental

variables into network models (Faust and Raes, 2012) in order to remove associations due to *shared environmental preference* (Ovaskainen et al., 2010). The third challenge is to integrate ecological knowledge on OTUs, in the form of 'rules of interaction'. This integration may improve the reliability of the network, as shown previously for macro-organisms (Bohan et al., 2011), and limits the space of possible networks that has to be explored.

Below we review the methods that can be used for addressing these challenges. These methods can be classified into two categories: statistical inference of networks and logic-based machine-learning algorithms. Both categories use similar input data and aim at constructing plausible and testable networks, which explain the observed data the best. However, they constitute different fields of research and therefore use a different vocabulary. The main terms are summarized in Table 1.

4.2 Inferring Ecological Interactions Using Statistical Models

Statistical models of graphs (Lauritzen, 1996; Whittaker, 1990) are a natural approach to depict the intricate interactions and relationships between species (or OTUs). They have proved to be a valuable framework for modelling *direct* relationships between species (or OTUs), based on conditional dependencies in occurrence and abundance data (Table 1). Here, we describe two popular approaches that might have value for reconstructing networks from NGS data.

4.2.1 Bayesian Networks and Dynamic Bayesian Networks

Bayesian Networks (BNs, Jensen and Nielsen, 2007) and their temporally explicit extension Dynamic Bayesian Networks (DBNs) framework (Dean and Kanazawa, 1989) have been used to reconstruct interactions from occurrence or abundance data. In ecology, BNs have been used to model species interactions and species–habitat relationships (Aderhold et al., 2012; Milns et al., 2010), while DBNs have been used to infer microbial networks from time series (Faust et al., 2015). The structure of BNs and DBNs can be derived from occurrence or abundance data, and possibly covariates representing environmental conditions, by using score-and-search techniques (Daly et al., 2011; Friedman et al., 1998). A score function (e.g. the Bayesian Information Criterion) is used to measure the fit between the observed data and an inferred network. Rules of iteration then move from one possible network structure to another in the space of possible network structures to evaluate fit. The aim is to find a network that maximizes the fit. Exact optimization of the score is out of reach because

of the high number of possible network structures. Various heuristics have been proposed to improve the search and its efficiency, including the 'greedy search' where only local improvements of the network structure are explored.

4.2.2 Gaussian Graphical Models with Sparse Regularization

Coupling Gaussian graphical models (GGM) with sparse regularization has become a popular method of inference in the last decade, because it allows us to deal with large networks involving thousands of nodes by setting many of the links to be zero. In the standard Gaussian setting, when the data are normally distributed, the approach assumes that the joint distribution of all OTUs follows a multivariate Gaussian distribution. The associated covariance matrix entirely describes the dependency structure between the OTUs. The precision matrix, which is formed as the inverse of the covariance matrix, is then a direct proxy for the expected network structure as the matrix entries are proportional to the partial covariances between the OTU abundances. Therefore, learning the graphical structure of conditional dependencies can be reduced to a problem of variable selection to optimize fit. This task can be efficiently performed by means of sparse regularization such as the Lasso (Tibshirani, 2011). Various lasso algorithms have been developed, including the 'neighbourhood selection' (Meinshausen and Bühlmann, 2006) and the 'Graphical Lasso' (Friedman et al., 2008). Some implementations can deal with millions of nodes.

To overcome the Gaussian assumption and comply with sequence count data like those obtained with NGS techniques, research is being done to broaden the applicability of GGM. Two lines of research have emerged on this topic. The first is to transform the original data into a 'Gaussian' setting, via simple transformations in order to use the well-understood GGM with sparse regularization framework (Liu et al., 2009). The second, which we do not discuss here, relies on the use of statistical models tailored to count data (Yang et al., 2013). The technique of data transformation has been successfully applied to ecological network inference (Kurtz et al., 2015), using a general workflow, as follows: (i) the OTU count data are preprocessed and normalized to meet the Gaussian assumption; (ii) a standard sparse GGM inference method (either 'neighbourhood selection' or 'graphical-Lasso') is used to select networks, describing direct links between OTUs; and, (iii) step (ii) is iterated several times on many random subsamples of the original data. The final network only retains the most stable edges, which appear robust because they are selected in most of the subsamples. This three-step

strategy—normalization plus inference plus stabilization—has also been successfully used on other types of genomic data (Marbach et al., 2012).

An advantage of the GGM framework is that it is well suited to theoretical analysis, which provides insights into the 'data' situations where the methods may either be useful or not. Most strikingly, an analysis of network selection consistency by Ravikumar et al. (2011) suggested that it should be possible to infer network structure when a network has many nodes and the sample size remains moderate, under the provision that there are no nodes in the network with very high connectance. Such statistical results are practically important because they give some confidence that networks may be inferred from NGS-based data where the number of samples is typically much smaller than the number of OTUs.

4.3 Learning Ecological Interactions Using Logic-Based Machine-Learning Algorithms

Like statistical models of graphs, the purpose of logic-based machine-learning algorithms is to construct plausible and testable networks, which best explain the observed data. The possibility of easily integrating the existing background knowledge into the learning process, e.g. in the form of 'rules of interaction', is a key advantage of logic-based machine-learning algorithms over statistical models (Table 1). Logic-based machine-learning algorithms have already been used successfully to automatically generate trophic networks directly from species occurrence data combined with background knowledge, including information about the species body size and functional groups (Bohan et al., 2011; Tamaddoni-Nezhad et al., 2013). They have also been applied successfully to other problems in ecology, in particular to model population dynamics (e.g. modelling phytoplankton growth for the Danish Lake Glumsø; Todorovski et al., 1998), and their relevance has also been demonstrated in many challenging domains in computational biology including predictive toxicology (e.g. King et al., 1996), pharmacophore design (e.g. Srinivasan et al., 2006) and protein structure prediction (e.g. Cootes et al., 2003).

Below, we describe two types of logic-based machine-learning algorithms that may be used to learn ecological networks from NGS data. Both types are capable of learning networks from species (or OTU) occurrence or abundance data. The first type, called inductive logic programming (ILP; Muggleton, 1991), is capable of using background knowledge, such as the existing knowledge about the species and their environment to hypothesize (learn) interactions (Table 1). In the case where background

knowledge may be incomplete or a subject of the learning itself, a new approach called meta-interpretive learning (MIL; Muggleton et al., 2014) may be used. To our knowledge, neither of these approaches have yet been used to learn ecological networks from NGS data, but the first approach has been used for classical ecological association data (co-occurrences in spatially replicated field sampling) to learn predator–prey networks (Bohan et al., 2011; Tamaddoni-Nezhad et al., 2013). Ecological interactions recovered from the high-throughput sequencing of predator gut contents (e.g. Fig. 3) or predicted by functional traits (e.g. body size or gape width), may be included as background knowledge, or may be used to validate the machine-learnt network. The second approach is likely to be more appropriate to the learning of microbial networks (e.g. Figs. 1 and 2) where background knowledge on microbial OTUs is often scarce.

4.3.1 Inductive Logic Programming

ILP systems (Muggleton, 1991) use a given set of positive and negative examples $E = \{E^+ \cup E^-\}$, i.e., E is the union of positive and negative examples, background knowledge B to construct a hypothesis, H, that explains E^+ relative to B such that the extended theory is self-consistent, i.e. $B \cup H$ logically implies E^+ and $B \cup H \cup E^-$ is logically consistent. The components E, B and H are each represented as logic programmes. In the case of machine learning of trophic networks, ILP systems can be used to learn ground hypotheses H in the form of trophic relations between species (or OTUs). Background knowledge includes logical rules, such as $R \subseteq B$ (i.e. R is a subset of B), to describe the species occurrence or abundance (observable predicates) in terms of the trophic interactions (abducible predicates) (Table 1).

This approach was used to learn a trophic network from an extensive Vortis suction sampling of invertebrates in 257 arable fields across the UK (Bohan et al., 2011). These fields were part of the farm scale evaluations (FSE) of genetically modified, herbicide-tolerant (GMHT) crops. The change in invertebrates abundance data with the GMHT treatment was regarded as the primary observational data for the learning: observable predicates were represented by *abundance(X, S, up)* (or *abundance(X, S, down)*) expressing the fact that the GMHT treatment increased or decreased the abundance of species X at site S. The aim was to learn abducible predicates *eats(X, Y)*, capturing the hypothesis that species X eats species Y. Additional information on the species were used to constrain the search for abducible predicate *eats(X, Y)*, by assuming that X should be a predator and bigger

than *Y*. Predicates *predator(X)* and *bigger_than(X, Y)* were provided as part of the background knowledge. This information was integrated within a logical rule describing the observable predicates (*abundance*) in terms of the abducible predicates (*eats*):

`abundance(X, S, Dir) if predator(X), bigger_than(X, Y), eats(X, Y), abundance(Y, S, Dir)` where *Dir* can be either up or down. This rule expresses the inference that following a perturbation in the ecosystem (i.e. the introduction of GMHT crops), the increased (or decreased) abundance of species *X* at site *S* can be explained by *X* eating species *Y*, which is lower in the food chain, and by changes in the abundance of species *Y*. Given this model and the observable data, the Abductive ILP system Progol 5.0 was used to generate a set of ground hypotheses (i.e. hypothetical trophic links) which was visualized as a trophic network (Bohan et al., 2011). The initial study was then extended by learning the trophic network from a larger dataset (Tamaddoni-Nezhad et al., 2013). In both cases, a probabilistic ILP approach, called Hypothesis Frequency Estimation (HFE, Tamaddoni-nezhad et al., 2012), was used for estimating probabilities of hypothetical trophic links. These probabilities were represented as the thickness of trophic links in Bohan et al. (2011) and Tamaddoni-Nezhad et al. (2013). Ecologists who examined the first machine-learned food web (Bohan et al., 2011) found that many of the learnt trophic links were corroborated by the literature. In particular, links ascribed with high probability by machine learning were shown to correspond well with those having multiple references in the literature. Novel, high probability links were also suggested, and some of these have recently been tested and confirmed by subsequent empirical studies. For example, in the hypothesized food webs, some species of spiders always appeared as prey for other predators; a result that was unexpected because spiders are obligate predators. This hypothesis was tested using molecular analysis of predator guts and it was found that in this system spiders do appear to play an important role as prey (Davey et al., 2013). This finding was reconfirmed in Section 3.3 and Fig. 3. Thus, even though some of the hypothesized links were unexpected, these were in fact confirmed later and this provided an extremely stringent test for the machine-learning approach.

4.3.2 Meta-Interpretive Learning

In the machine-learning settings described in the previous section, the search for trophic links was constrained by additional information on the species

(e.g. body size, trophic behaviour) which were provided as part of background knowledge. The logical rule stated that X may eat Y if X is a predator bigger than Y. However, for most communities and ecosystems, including microbial communities, this kind of background knowledge may not be available or it may be incomplete. MIL (Muggleton et al., 2014) is a new machine-learning approach capable of predicate invention and recursive rule learning. This new approach can be used for learning both the interactions between species (or OTUs) and the 'rules of interaction' directly from species occurrence or abundance data. In this case, background knowledge does not include any specific knowledge on the species but includes higher-order meta-rules, $M \subseteq B$, which are activated during the proving of examples in order to generate hypotheses, H. A recent study showed that MIL can be used to re-construct a simplified food web and learn interaction rules directly from data (Tamaddoni-Nezhad et al., 2015). We believe that this new learning setting will be useful for learning ecological networks from NGS data whenever the interaction rules are not known before hand.

5. CONCLUSION

Most interactions between species are difficult to observe, and as a consequence the ecological networks that we typically reconstruct and analyze to understand ecosystem function are incomplete. Through the examples given in this review, we have argued that NGS techniques permit the characterization of biodiversity in complex environmental samples (e.g. soil, water, plant tissues, faeces, pellet, gut content, etc.) containing hundreds of microbial OTUs and multiple macro-organism species. In the lists of co-occurring species and OTUs that we can discern with NGS techniques, there are the 'ghosts of interactions past' from which we could learn robust and more complete ecological networks, at any specified taxonomic resolution and for organisms from all the Kingdoms of life. In this review, we have then showed that various statistical and machine-learning approaches are available for performing such network reconstruction, but the majority have yet to be applied to NGS data. We believe that the combination of these theoretical approaches with cost-effective NGS techniques will allow us to study species interactions under all environmental conditions at high replication. However, three outstanding challenges remain to be overcome to achieve these aims. These are to:

- improve the qualitative and quantitative reliability of NGS data, while keeping the costs down;
- use well-characterized networks of ecological interactions, in order to test the validity of the various methods of network learning and reconstruction;
- develop theoretical approaches to enable the learning of temporally dynamic ecological networks.

Should these three challenges be met, then we foresee a step change in our ability to measure, understand and monitor the world's ecosystems and the functions and services they provide.

ACKNOWLEDGEMENTS

We thank the *métaprogramme EcoServ* (INRA) for funding the meeting that gave rise to this chapter. The work presented in Fig. 1 was funded by the METAPHORE project (AIP Bioressource) and a grant from the French Ministry of Research and Education (MENRT no. 2011/AF/57). We thank Cecile Robin for helpful discussions. The work presented in Fig. 2 was supported by the Portuguese Foundation for Science and Technology 609 (FCT) and FEDER with the project IBISA (PTDC/MAR/64749/2006). We thank Tania Aires, Ester Serrao and Sophie Arnaud-Haond for providing the dataset. The study presented in Figs. 3 and 4 was funded by the two French ANR projects, Landscaphid (ANR-09-STRA-05) and Peerless (ANR-12-AGRO-0006). The second author acknowledges the support of an EPSRC 'Pathways to Impact Award' during the writing of this paper.

D.B. is supported by two ANR projects, PEERLESS (ANR-12-AGRO-0006) and AGROBIOSE (ANR-13-AGRO-0001).

GLOSSARY

Barcode A short genetic sequence taken from a standardized portion of the genome that is used to identify species.

Biodiversity The variety of life, including variation among genes, species, and functional traits. It is often measured as: richness is a measure of the number of unique life forms; evenness is a measure of the equitability among life forms; and heterogeneity is the dissimilarity among life forms (Cardinale et al., 2012).

Conditional dependency The relationship between two variables conditioned on all other variables. In a Gaussian setting, conditional dependency is measured by partial correlation.

Diagnostic PCR A PCR assay which is used to test samples for the presence of DNA from a specific species or a group of organisms.

Ecosystem functions (or functioning) Ecological processes that control the fluxes of energy, nutrients, and organic matter through an environment. Primary production,

for instance, is an ecosystem function. It is the process by which plants use sunlight to convert inorganic matter into new biological tissue (Cardinale et al., 2012).

Environmental DNA (eDNA) Genetic material obtained directly from environmental samples (soil, sediment, water, etc.).

Graph A mathematical object where entities (represented by vertices) are connected with links (called edges). In graphical models, graphs are used to depict the structure of conditional dependencies between variables.

Holobiont The functional entity composed by a macro-species and its associated symbiotic microbes.

Logic-based machine learning A form of machine learning which uses logic-based inference and representation. For example, in inductive logic programming (ILP), the training examples, background knowledge, and the learned hypotheses are all represented as logic programmes.

Meta-barcoding (amplicon-based community profiling) A method of biodiversity assessment combining DNA-based identification and high-throughput sequencing. It uses universal PCR primers to mass-amplify genetic markers from mass collections of organisms or from environmental samples.

Next-generation sequencing (high-throughput sequencing) Technologies that parallelize the sequencing process, producing thousands, or millions of sequences concurrently.

Niche partitioning theory refers to evolutionary and ecological processes leading to differential resource exploitation between species in response to interspecific competition.

Operational taxonomic units (OTUs) are usually defined as clusters of similar barcode sequences (16S rDNA, ITS etc.), frequently intended to represent some degree of taxonomic relatedness.

OTU table Matrix giving the number of sequences per OTU and per sample.

Pathobiome The pathogenic agent plus the members of its biotic environment.

Sanger sequencing (chain-termination method) Method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during *in vitro* DNA replication. Developed by Frederick Sanger and colleagues in 1977, it was the most widely used sequencing method for approximately 25 years. It can be used for fairly short strands of 100–1000 bp.

Shotgun sequencing Method used for sequencing long DNA strands that was one of the precursor technologies for full genome sequencing. DNA is fragmented into numerous small fragments that are sequenced using the chain termination method to obtain reads. Multiple overlapping reads for the target DNA are obtained and computer programmes then use the overlapping ends of different reads to assemble them into a continuous sequence.

Sparse regularization Technique from constrained optimization used to force some entries to zero in a vector of parameters. When applied to network inference, it allows to deal with thousands of nodes, by coercing many of the edges to zero.

Statistical inference The process of drawing conclusions on a population based on data.

Tag A unique DNA sequence ligated to fragments within a sequencing library for downstream sorting and identification. Tags are typically a component of adapters or PCR primers and are between 8 and 12 bp. Libraries with unique tags can be pooled together and sequenced in the same sequencing run (=multiplexing). Reads are later identified and sorted via bioinformatic pipelines.

REFERENCES

Abarenkov, K., Nilsson, R.H., Larsson, K.H., Alexander, I.J., Eberhardt, U., Erland, S., Hoiland, K., Kjoller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A.F.S., Tedersoo, L., Ursing, B.M., Vralstad, T., Liimatainen, K., Peintner, U., Koljalg, U., 2010. The UNITE database for molecular identification of fungi—recent updates and future perspectives. New Phytol. 186, 281–285.

Abrams, P.A., Matsuda, H., 2005. The effect of adaptive change in the prey on the dynamics of an exploited predator population. Can. J. Fish. Aquat. Sci. 62, 758–766.

Aderhold, A., Husmeier, D., Lennon, J.J., Beale, C.M., Smith, V.A., 2012. Hierarchical Bayesian models in ecology: reconstructing species interaction networks from non-homogeneous species abundance data. Ecol. Inform. 11, 55–64.

Aires, T., Moalic, Y., Serrao, E.A., Arnaud-Haond, S., 2015. Hologenome theory supported by co-occurrence networks of species-specific bacterial communities in siphonous algae (Caulerpa). FEMS Microbiol. Ecol. 91 (7): fiv067.

Aizen, M.A., Morales, C.L., Morales, J.M., 2008. Invasive mutualists erode native pollination webs. PLoS Biol. 6, e31.

Albrecht, M., Duelli, P., Schmid, B., Müller, C.B., 2007. Interaction diversity within quantified insect food webs in restored and adjacent intensively managed meadows. J. Anim. Ecol. 76, 1015–1025.

Albrecht, M., Padrón, B., Bartomeus, I., Traveset, A., 2014. Consequences of plant invasions on compartmentalization and species' roles in plant–pollinator networks. Proc. R. Soc. B 281, 20140773.

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., Taberlet, P., 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol. Ecol. Resour. 14, 306–323.

Barberán, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J. 6, 343–351.

Bascompte, J., Jordano, P., Melian, C.J., Olesen, J.M., 2003. The nested assembly of plant–animal mutualistic networks. Proc. Natl. Acad. Sci. U. S. A. 100, 9383–9387.

Bascompte, J., Jordano, P., Olesen, J.M., 2006. Asymmetric coevolutionary networks facilitate biodiversity maintenance. Science 312, 431–433.

Bastian, M., Heymann, S., 2009. Gephi: an open source software for exploring and manipulating networks. In: International AAAI Conference on Weblogs and Social Media.

Beckerman, A.P., Petchey, O.L., Warren, P.H., 2006. Foraging biology predicts food web complexity. Proc. Natl. Acad. Sci. U. S. A. 103, 13745–13749.

Berry, D., Ben Mahfoudh, K., Wagner, M., Loy, A., 2011. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. Appl. Environ. Microbiol. 77, 7846–7849.

Berlow, E.L., Neutel, A.M., Cohen, J.E., De Ruiter, P.C., Ebenman, B., Emmerson, M., Fox, J.W., Jansen, V.A., Jones, J.I., Kokkoris, G.D., Logofet, D.O., Mckane, A.J., Montoya, J.M., Petchey, O., 2004. Interaction strengths in food webs: issues and opportunities. J. Anim. Ecol. 73, 585–598.

Cotton, T.E.A., Dumbrell, A.J., Helgason, T., 2014. What goes in must come out: testing for biases in molecular analysis of arbuscular mycorrhizal fungal communities. PLoS One 9, e109234.

Derocles, S.A.P., Evans, D.M., Nichols, P.C., Evans, S.A., Lunt, D.H., 2015. Determining plant—leaf miner—parasitoid interactions: a DNA barcoding approach. PLoS One 10 (2), e0117872.

Di Bella, J.M., Bao, Y., Gloor, G.B., Burton, J.P., Reid, G., 2013. High throughput sequencing methods and analysis for microbiome research. J. Microbiol. Methods 95, 401–414.

Berendsen, R.L., Pieterse, C.M.J., Bakker, P.A.H.M., 2012. The rhizosphere microbiome and plant health. Trends Plant Sci. 17, 478–486.
Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J.-J., Taberlet, P., 2012. Tracking earthworm communities from soil DNA. Mol. Ecol. 21, 2017–2030.
Blüthgen, N., Fründ, J., Vázquez, D.P., Menzel, F., Bluthgen, N., Frund, J., Vazquez, D.P., 2008. What do interaction network metrics tell us about specialization and biological traits? Ecology 89, 3387–3399.
Boecklen, W.J., Yarnes, C.T., Cook, B.A., James, A.C., 2011. On the use of stable isotopes in trophic ecology. Annu. Rev. Ecol. Evol. Syst. 42, 411–440.
Bohan, D.A., Caron-Lormier, G., Muggleton, S., Raybould, A., Tamaddoni-Nezhad, A., 2011. Automated discovery of food webs from ecological data using logic-based machine learning. PLoS One 6, e29028.
Bohan, D.A., Raybould, A., Mulder, C., Woodward, G., Tamaddoni-nezhad, A., Blüthgen, N., Pocock, M.J.O., Muggleton, S., Evans, D.M., Astegiano, J., Massol, F., Loeuille, N., Petit, S., Macfadyen, S., 2013. Networking agroecology: integrating the diversity of agroecosystem interactions. Adv. Ecol. Res. 49, 1–67.
Boyer, S., Wratten, S.D., Holyoake, A., Abdelkrim, J., Cruickshank, R.H., 2013. Using next-generation sequencing to analyse the diet of a highly endangered land snail (*Powelliphanta augusta*) feeding on endemic earthworms. PLoS One 8, e75962.
Brown, D.S., Jarman, S.N., Symondson, W.O.C., 2012. Pyrosequencing of prey DNA in reptile faeces: analysis of earthworm consumption by slow worms. Mol. Ecol. Resour. 12, 259–266.
Calderone, N.W., 2012. Insect pollinated crops, insect pollinators and US agriculture: trend analysis of aggregate data for the period 1992–2009. PLoS One 7, e37235.
Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., Mcdonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336.
Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact on humanity. Nature 486, 59–67.
Carroll, S.P., Jørgensen, P.S., Michael, T., Bergstrom, C.T., Denison, R.F., Gluckman, P., Smith, T.B., Strauss, S.Y., Tabashnik, B.E., 2014. Applying evolutionary biology to address global challenges. Science 346, 1–16.
Chaffron, S., Rehrauer, H., Pernthaler, J., Mering, C., 2010. A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res. 2010, 947–959.
Chakraborty, C., Doss, C.G.P., Patra, B.C., Bandyopadhyay, S., 2014. DNA barcoding to map the microbial communities: current advances and future directions. Appl. Microbiol. Biotechnol. 98, 3425–3436.
Chakraborty, S., Newton, A.C., 2011. Climate change, plant diseases and food security: an overview. Plant Pathol. 60, 2–14.
Chesson, P., 2000. Mechanisms of maintenance of species diversity. Annu. Rev. Ecol. Syst. 31, 343–366.
Chow, C.E.T., Kim, D.Y., Sachdeva, R., Caron, D.A., Fuhrman, J.A., 2014. Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. ISME J. 8, 816–829.

Clare, E.L., Fraser, E.E., Braid, H.E., Fenton, M.B., Hebert, P.D.N., 2009. Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. Mol. Ecol. 18, 2532–2542.

Clarke, L.J., Czechowski, P., Soubrier, J., Stevens, M.I., Cooper, A., 2014. Modular tagging of amplicons using a single PCR for high-throughput sequencing. Mol. Ecol. Resour. 14, 117–121.

Condon, M.A., Scheffer, S.J., Lewis, M.L., Wharton, R., Adams, D.C., Forbes, A.A., 2014. Lethal interactions between parasites and prey increase niche diversity in a tropical community. Science 343, 1240–1244.

Cootes, A.P., Muggleton, S.H., Sternberg, M.J.E., 2003. The automatic discovery of structural principles describing protein fold space. J. Mol. Biol. 330, 839–850.

Cordier, T., Robin, C., Capdevielle, X., Fabreguettes, O., Desprez-Loustau, M.-L., Vacher, C., 2012. The composition of phyllosphere fungal assemblages of European beech (*Fagus sylvatica*) varies significantly along an elevation gradient. New Phytol. 196, 510–519.

Daly, R., Shen, Q., Aitken, S., 2011. Learning Bayesian networks: approaches and issues. Knowl. Eng. Rev. 26, 99–157.

Dakos, V., Bascompte, J., 2014. Critical slowing down as early warning for the onset of collapse in mutualistic communities. Proc. Natl. Acad. Sci. 11, 17546–17551.

Davey, J.S., Vaughan, I.P., Andrew King, R., Bell, J.R., Bohan, D.A., Bruford, M.W., Holland, J.M., Symondson, W.O.C., 2013. Intraguild predation in winter wheat: prey choice by a common epigeal carabid consuming spiders. J. Appl. Ecol. 50, 271–279.

Deagle, B.E., Kirkwood, R., Jarman, S.N., 2009. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Mol. Ecol. 18, 2022–2038.

Dean, T., Kanazawa, K., 1989. A model for reasoning about persistence and causation. Comput. Intell. 5, 142–150.

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

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072.

Desprez-Loustau, M.-L., Aguayo, J., Dutech, C., Hayden, K.J., Husson, C., Jakushkin, B., Marçais, B., Piou, D., Robin, C., Vacher, C., 2015. An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. Ann. For. Sci. http://dx.doi.org/10.1007/s13595-015-0487-4.

Dittami, S.M., Duboscq-Bidot, L., Perennou, M., Gobet, A., Corre, E., Boyen, C., Tonon, T., 2015. Host–microbe interactions as a driver of acclimation to salinity gradients in brown algal cultures. ISME J. http://dx.doi.org/10.1038/ismej.2015.104.

Ducklow, H., 2008. Microbial services: challenges for microbial ecologists in a changing world. Aquat. Microb. Ecol. 53, 13–19.

Duffy, J.E., Carinale, B.J., France, K.E., McIntyre, P.B., Thebault, E., Loreau, M., 2007. The functional role of biodiversity in ecosystems: incorporating trophic complexity. Ecol. Lett. 10, 522–538.

Duran-Pinedo, A.E., Paster, B., Teles, R., Frias-Lopez, J., 2011. Correlation network analysis applied to complex biofilm communities. PLoS One 6, e28438.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998.

Eveleigh, E.S., McCann, K.S., McCarthy, P.C., Pollock, S.J., Lucarotti, C.J., Morin, B., McDougall, G.A., Strongman, D.B., Huber, J.T., Umbanhowar, J., Faria, L.D.B., 2007. Fluctuations in density of an outbreak species drive diversity cascades in food webs. Proc. Natl. Acad. Sci. 104, 16976–16981.

Faisal, A., Dondelinger, F., Husmeier, D., Beale, C.M., 2010. Inferring species interaction networks from species abundance data: a comparative evaluation of various statistical and machine learning methods. Ecol. Inform. 5, 451–464.
Faust, K., Lahti, L., Gonze, D., de Vos, W.M., Raes, J., 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Curr. Opin. Microbiol. 25, 56–66.
Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. Nat. Rev. Microbiol. 10, 538–550.
Fontaine, C., Guimaraes, P.R., Kefi, S., Loeuille, N., Memmott, J., van der Putten, W.H., van Veen, F.J.F., Thebault, E., 2011. The ecological and evolutionary implications of merging different types of networks. Ecol. Lett. 14, 1170–1181.
Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D., Krasnov, B.R., Poulin, R., Bascompte, J., 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? J. Anim. Ecol. 79, 811–817.
Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486, 75–174.
Foster, K.R., Bell, T., 2012. Competition, not cooperation, dominates interactions among culturable microbial species. Curr. Biol. 22, 1845–1850.
Frey-Klett, P., Burlinson, P., Deveau, A., Barret, M., Tarkka, M., Sarniguet, A., 2011. Bacterial–fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists. Microbiol. Mol. Biol. Rev. 75, 583–609.
Friedman, J., Alm, E.J., 2012. Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 8, e1002687.
Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441.
Friedman, N., Murphy, K., Russell, S., 1998. Learning the structure of dynamic probabilistic networks. In: Proceedings of the Fourteenth Conference on Uncertainity in Artificial Intelligence, pp. 139–147.
Fuhrman, J.A., 2009. Microbial community structure and its functional implications. Nature 459, 193–199.
Fuhrman, J.A., Cram, J.A., Needham, D.M., 2015. Marine microbial community dynamics and their ecological interpretation. Nat. Rev. Microbiol. 13, 133–146.
Gaston, K.J., 2010. Ecology. Valuing common species. Science 327, 154–155.
Gaspar, J.M., Thomas, W.K., 2013. Assessing the consequences of denoising marker-based metagenomic data. PLoS One 8, e60458.
Gilbert, J.A., Neufeld, J.D., 2014. Life in a world without microbes. PLoS Biol. 12, 1–3.
Gilles, A.A., Meglecz, E., Pech, N., Ferreira, S.S., Malausa, T., Martin, J.-F., 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 12, 245.
Gloor, G.B., Hummelen, R., Macklaim, J.M., Dickson, R.J., Fernandes, A.D., MacPhee, R., Reid, G., 2010. Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. PLoS One 5, e15406.
Gonzalez, J.M., Portillo, M.C., Belda-Ferre, P., Mira, A., 2012. Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. PLoS One 7, e29973.
Haddad, N.M., Crutsinger, G.M., Gross, K., Haarstad, J., Tilman, D., 2011. Plant diversity and the stability of foodwebs. Ecol. Lett. 14, 42–46.
Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., Weitz, J.S., 2013. Robust estimation of microbial diversity in theory and in practice. ISME J. 7, 1092–1101.
Hagen, M., Kissling, W.D., Rasmussen, C., De Aguiar, M.A.M., Brown, L.E., Carstensen, D.W., Alves-Dos-Santos, I., Dupont, Y.L., Edwards, F.K., Genini, J., Guimarães, P.R., Jenkins, G.B., Jordano, P., Kaiser-Bunbury, C.N., Ledger, M.E., Maia, K.P., Marquitti, F.M.D., Mclaughlin, Ó., Morellato, L.P.C., O'Gorman, E.J.,

Trøjelsgaard, K., Tylianakis, J.M., Vidal, M.M., Woodward, G., Olesen, J.M., 2012. Biodiversity, species interactions and ecological networks in a fragmented world. Adv. Ecol. Res. 46, 89–120.
Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T., Fox, J.A., 2005. Rapid evolution and the convergence of ecological and evolutionary time. Ecol. Lett. 8, 1114–1127.
Haruta, S., Kato, S., Yamamoto, K., Igarashi, Y., 2009. Intertwined interspecies relationships: approaches to untangle the microbial network. Environ. Microbiol. 11, 2963–2969.
Hector, A., Hooper, R., 2002. Ecological experiment. Science 295, 639–640.
Heleno, R., Devoto, M., Pocock, M., 2012. Connectance of species interaction networks and conservation value: is it any good to be well connected? Ecol. Indic. 14, 7–10.
Heleno, R.H., Ceia, R.S., Ramos, J.A., Memmott, J., 2009. Effects of alien plants on insect abundance and biomass: a food-web approach. Conserv. Biol. 23, 410–419.
Hengeveld, R., 1980. Polyphagy, oligophagy and food specialization in ground beetles (Coleoptera, Carabidae). Neth. J. Zool. 30, 564–584.
Hibbett, D.S., Ohman, A., Kirk, P.M., 2009. Fungal ecology catches fire. New Phytol. 184, 279–282.
Hibbing, M.E., Fuqua, C., Parsek, M.R., Peterson, S.B., 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nat. Rev. Microbiol. 8, 15–25.
Hudson, P.J., Dobson, A.P., Lafferty, K.D., 2006. Is a healthy ecosystem one that is rich in parasites? Trends Ecol. Evol. 21, 381–385.
Ibanez, S., Manneville, O., Miquel, C., Taberlet, P., Valentini, A., Aubert, S., Coissac, E., Colace, M.P., Duparc, Q., Lavorel, S., Moretti, M., 2013. Plant functional traits reveal the relative contribution of habitat and food preferences to the diet of grasshoppers. Oecologia 173, 1459–1470.
Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E., Lewis, H.M., Olesen, J.M., Veen, V., Frank, F.J., Warren, P.H., Woodward, G., 2009. Ecological networks—beyond food webs. J. Anim. Ecol. 78, 253–269.
Isbell, F., Tilman, D., Polasky, S., Binder, S., Hawthorne, P., 2013. Low biodiversity state persists two decades after cessation of nutrient enrichment. Ecol. Lett. 16, 454–460.
Janson, E.M., Stireman, J.O., Singer, M.S., Abbot, P., 2008. Phytophagous insect–microbe mutualisms and adaptive evolutionary diversification. Evolution 62, 997–1012.
Jensen, F.V., Nielsen, T.D., 2007. Bayesian Networks and Decision Graphs. Springer, New York.
Joppa, L.N., Montoya, J.M., Sole, R., Sanderson, J., Pimm, S.L., 2010. On nestedness in ecological networks. Evol. Ecol. Res. 12, 35–46.
Kaartinen, R., Stone, G.N., Hearn, J., Lohse, K., Roslin, T., 2010. Revealing secret liaisons: DNA barcoding changes our understanding of food webs. Ecol. Entomol. 35, 623–638.
Kara, E.L., Hanson, P.C., Hu, Y.H., Winslow, L., McMahon, K.D., 2013. A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA. ISME J. 7, 680–684.
Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., Wang, W., Pringle, R.M., 2015. DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proc. Natl. Acad. Sci. 26, 819–824.
Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., Boit, A., Joppa, L.N., Lafferty, K.D., Williams, R.J., Martinez, N.D., Menge, B.A., Blanchette, C.A., Iles, A.C., Brose, U., 2012. More than a meal … integrating non-feeding interactions into food webs. Ecol. Lett. 15, 291–300.

Kemen, E., 2014. Microbe–microbe interactions determine oomycete and fungal host colonization. Curr. Opin. Plant Biol. 20, 75–81.
King, R.D., Muggleton, S.H., Srinivasan, A., Sternberg, M.J., 1996. Structure–activity relationships derived by machine learning: the use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. Proc. Natl. Acad. Sci. 93, 438–442.
Kõljalg, U., Larsson, K.-H., Abarenkov, K., Nilsson, R.H., Alexander, I.J., Eberhardt, U., Erland, S., Høiland, K., Kjøller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A.F.S., Tedersoo, L., Vrålstad, T., Ursing, B.M., 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol. 166, 1063–1068.
Kondoh, M., 2003. Foraging adaptation and the relationship between food-web complexity and stability. Science 299, 1388–1391.
Kromp, B., 1999. Carabid beetles in sustainable agriculture: a review on pest control efficacy, cultivation impacts and enhancement. Agric. Ecosyst. Environ. 74, 187–228.
Kurtz, Z.D., Mueller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A., 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput. Biol. 11 (5), e1004226.
Lafferty, K.D., Allesina, S., Arim, M., Briggs, C.J., De Leo, G., Dobson, A.P., Dunne, J.A., Johnson, P.T.J., Kuris, A.M., Marcogliese, D.J., Martinez, N.D., Memmott, J., Marquet, P.A., McLaughlin, J.P., Mordecai, E.A., Pascual, M., Poulin, R., Thieltges, D.W., 2008. Parasites in food webs: the ultimate missing links. Ecol. Lett. 11, 533–546.
Laliberté, E., Tylianakis, J.M., 2010. Deforestation homogenizes tropical parasitoid-host networks. Ecology 91, 1740–1747.
Lancichinetti, A., Fortunato, S., 2009. Community detection algorithms: a comparative analysis. Phys. Rev. E 80, 056117.
Lauritzen, S.L., 1996. Graphical Models. Oxford Science Publications, New York, NY.
Layer, K., Hildrew, A.G., Jenkins, G.B., Riede, J., Rossiter, S.J., Townsend, C.R., Woodward, G., 2011. Long-term dynamics of a well-characterised food web: four decades of acidification and recovery in the Broadstone Stream model system. Adv. Ecol. Res. 44, 69–117.
Layer, K., Riede, J.O., Hildrew, A.G., Woodward, G., 2010. Food web structure and stability in 20 streams across a wide pH gradient. Adv. Ecol. Res. 42, 265–299.
Lee, C.K., Herbold, C.W., Polson, S.W., Wommack, K.E., Williamson, S.J., McDonald, I.R., Cary, S.C., 2012. Groundtruthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing. PLoS One 7, e44224.
Leger, J.-B., Vacher, C., Daudin, J., 2015. Clustering methods differ in their ability to detect patterns in species interaction networks. Methods Ecol. Evol. 6, 474–481.
Leger, J.B., Vacher, C., Daudin, J.J., 2014. Detection of structurally homogeneous subsets in graphs. Stat. Comput. 24, 675–692.
Levine, J.M., HilleRisLambers, J., 2009. The importance of niches for the maintenance of species diversity. Nature 461, 254–257.
Lewinsohn, T.M., Prado, P.I., Jordano, P., Bascompte, J., Olesen, J.M., 2006. Structure in plant–animal interaction assemblages. Oikos 113, 174–184.
Liu, H., Laferty, J., Wasserman, L., 2009. The nonparanormal: semiparametric estimation of high dimensional undirected graphs. J. Mach. Learn. Res. 10, 2295–2328.
Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012, 1–11.
Loeuille, N., 2010a. Consequences of adaptive foraging in diverse communities. Funct. Ecol. 24, 18–27.

Loeuille, N., 2010b. Influence of evolution on the stability of ecological communities. Ecol. Lett. 13, 1536–1545.
Loeuille, N., Loreau, M., 2006. Evolution of body size in food webs: does the energetic equivalence rule hold? Ecol. Lett. 9, 171–178.
Loeuille, N., Loreau, M., Ferrière, R., 2002. Consequences of plant–herbivore coevolution on the dynamics and functioning of ecosystems. J. Theor. Biol. 217, 369–381.
Lok, C., 2015. Mining the microbial dark matter. Nature 522, 270–273.
Lopezaraiza-Mikel, M.E., Hayes, R.B., Whalley, M.R., Memmott, J., 2007. The impact of an alien plant on a native plant–pollinator network: an experimental approach. Ecol. Lett. 10, 539–550.
Loreau, M., Hector, A., 2001. Partitioning selection and complementarity in biodiversity experiments. Nature 412, 72–76.
Loreau, M., de Mazancourt, C., 2013. Biodiversity and ecosystem stability: a synthesis of underlying mechanisms. Ecol. Lett. 16, 106–115.
Loreau, M., Mouquet, N., Gonzalez, A., 2003. Biodiversity as spatial insurance in heterogeneous landscapes. Proc. Natl. Acad. Sci. 100, 12765–12770.
MacArhur, R.H., 1958. Population ecology of some warblers of Northeastern coniferous forests. Ecology 39, 599–618.
Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Kellis, M., Collins, J.J., Stolovitzky, G., 2012. Wisdom of crowds for robust gene network inference. Nat. Methods 9, 796–804.
May, R.M., 1973. Stability and complexity in model ecosystems. Monogr. Popul. Biol. 6, 1–235.
Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the Lasso. Ann. Stat. 34, 1436–1462.
Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-Being: Synthesis. Island Press, Washington, DC.
Milns, I., Beale, C.M., Smith, V.A., 2010. Revealing ecological networks using Bayesian network inference algorithms. Ecology 91, 1892–1899.
Moalic, Y., Desbruyères, D., Duarte, C.M., Rozenfeld, A.F., Bachraty, C., Arnaud-Haond, S., 2012. Biogeography revisited with network theory: retracing the history of hydrothermal vent communities. Syst. Biol. 61, 127–137.
Mollot, G., Duyck, P.-F., Lefeuvre, P., Lescourret, F., Martin, J.-F., Piry, S., Canard, E., Tixier, P., 2014. Cover cropping alters the diet of arthropods in a banana plantation: a metabarcoding approach. PLoS One 9, e93740.
Montoya, J.M., Pimm, S.L., Sole, R.V., 2006. Ecological networks and their fragility. Nature 442, 259–264.
Moore, J.W., Semmens, B.X., 2008. Incorporating uncertainty and prior information into stable isotope mixing models. Ecol. Lett. 11, 470–480.
Mougou, A., Dutech, C., Desprez-Loustau, M.-L., 2008. New insights into the identity and origin of the causal agent of oak powdery mildew in Europe. For. Pathol. 38, 275–287.
Muggleton, S., 1991. Inductive logic programming. New Generat. Comput. 8, 295–318.
Muggleton, S.H., Lin, D., Pahlavi, N., Tamaddoni-Nezhad, A., 2014. Meta-interpretive learning: application to grammatical inference. Mach. Learn. 94, 25–49.
Müller, C.B., Adriaanse, I.C.T., Belshaw, R., Godfray, H.C.J., 1999. The structure of an aphid–parasitoid community. J. Anim. Ecol. 68, 346–370.
Naeem, S., Bunker, D.E., Hector, A., Loreau, M., Perrings, C., 2009. Biodiversity, Ecosystem Functioning, and Human Wellbeing—An Ecological and Economic Perspective. Oxford University Press, Oxford, UK.
Navarrete, A.A., Tsai, S.M., Mendes, L.W., Faust, K., de Hollander, M., Cassman, N.A., Raes, J., van Veen, J.A., Kuramae, E.E., 2015. Soil microbiome responses to the short-term effects of Amazonian deforestation. Mol. Ecol. 24 (10), 2433–2448.

Nuismer, S.L., Jordano, P., Bascompte, J., 2013. Coevolution and the architecture of mutualistic networks. Evolution 67, 338–354.
Öpik, M., Davison, J., Moora, M., Zobel, M., 2014. DNA-based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences. Botany 92, 135–147.
Ovaskainen, O., Hottola, J., Siitonen, J., 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91, 2514–2521.
Patin, N.V., Kunin, V., Lidström, U., Ashby, M.N., 2013. Effects of OTU Clustering and PCR Artifacts on Microbial Diversity Estimates. Microb. Ecol. 65, 709–719.
Paula, D.P., Linard, B., Andow, D.A., Sujii, E.R., Pires, C.S.S., Vogler, A.P., 2015. Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics. Mol. Ecol. Resour. 15, 880–892.
Penton, C.R., Gupta, V.V.S.R., Tiedje, J.M., Neate, S.M., Ophel-Keller, K., Gillings, M., Harvey, P., Pham, A., Roget, D.K., 2014. Fungal community structure in disease suppressive soils assessed by 28S LSU gene sequencing. PLoS One 9, e93893.
Pinto, A.J., Raskin, L., 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One 7, e43093.
Pisanu, B., Caut, S., Gutjahr, S., Vernon, P., Chapuis, J.L., 2011. Introduced black rats *Rattus rattus* on Ile de la Possession (Iles Crozet, subantarctic): diet and trophic position in food webs. Polar Biol. 34, 169–180.
Pocock, M.J.O., Evans, D.M., Memmott, J., 2012. The robustness and restoration of a network of ecological networks. Science 335, 973–977.
Pompanon, F., Deagle, B.E., Symondson, W.O.C., Brown, D.S., Jarman, S.N., Taberlet, P., 2012. Who is eating what: diet assessment using next generation sequencing. Mol. Ecol. 21, 1931–1950.
Quail, M., Smith, M.E., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion torrent, Pacific biosciences and Illumina MiSeq sequencers. BMC Genomics 13, 341.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, 590–596.
Quéméré, E., Hibert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., Champeau, J., Rabarivola, C., Nusbaumer, L., Chatelain, C., Gautier, L., Ranirison, P., Crouau-Roy, B., Taberlet, P., Chikhi, L., 2013. A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. PLoS One 8, e58971.
Raffaelli, D., White, P.C.L., 2013. Ecosystems and their services in a changing world. An ecological perspective. Adv. Ecol. Res. 48, 1–70.
Raffaelli, D.G., Bullock, J.M., Cinderby, S., Durance, I., Emmett, B., Harris, J., Hicks, K., Oliver, T.H., Paterson, D., White, P.C.L., 2014. Adv. Ecol. Res. 51, 41–77.
Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system. Mol. Ecol. Notes 7, 355–364.
Ravikumar, P., Wainwright, M.J., Raskutti, G., Yu, B., 2011. High-dimensional covariance estimation by minimizing $\ell 1$-penalized log-determinant divergence. Electron. J. Stat. 5, 935–980.
Reeder, J., Knight, R., 2011. Rapid denoising of pyrosequencing amplicon data: exploiting the rank-abundance distribution. Nat. Methods 7, 668–669.
Robinson, K.M., Hauzy, C., Loeuille, N., Albrectsen, B.R., 2015. Relative impacts of environmental variation and evolutionary history on the nestedness and modularity of tree–herbivore networks. Ecol. Evol. 5 (14), 2898–2915.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., Foley, J.A., 2009. A safe operating space for humanity. Nature 461, 472–475.

Rohr, R.P., Saavedra, S., Bascompte, J., 2014. On the structural stability of mutualistic systems. Science 345, 1253497.

Rosenberg, E., Koren, O., Reshef, L., Efrony, R., Zilber-Rosenberg, I., 2007. The role of microorganisms in coral health, disease and evolution. Nat. Rev. Micro. 5 (5), 355–362.

Rougerie, R., Smith, M.A., Fernandez-Triana, J., Lopez-Vaamonde, C., Ratnasingham, S., Hebert, P.D.N., 2011. Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval host. Mol. Ecol. 20, 179–186.

Saavedra, S., Stouffer, D.B., Uzzi, B., Bascompte, J., 2011. Strong contributors to network persistence are the most vulnerable to extinction. Nature 478, 233–235.

Sapkota, R., Knorr, K., Jørgensen, L.N., O'Hanlon, K.A., Nicolaisen, M., 2015. Host genotype is an important determinant of the cereal phyllosphere mycobiome. New Phytol. 207, 1134–1144.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. Nature 413, 591–596.

Schlaeppi, K., Bulgarelli, D., 2015. The plant microbiome at work. MPMI 28, 212–217.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W., Miller, A.N., Wingfield, M.J., Aime, M.C., An, K.-D., Bai, F.-Y., Barreto, R.W., Begerow, D., Bergeron, M.-J., Blackwell, M., Boekhout, T., Bogale, M., Boonyuen, N., Burgaz, A.R., Buyck, B., Cai, L., Cai, Q., Cardinali, G., Chaverri, P., Coppins, B.J., Crespo, A., Cubas, P., Cummings, C., Damm, U., de Beer, Z.W., de Hoog, G.S., Del-Prado, R., Dentinger, B., Dieguez-Uribeondo, J., Divakar, P.K., Douglas, B., Duenas, M., Duong, T.A., Eberhardt, U., Edwards, J.E., Elshahed, M.S., Fliegerova, K., Furtado, M., Garcia, M.A., Ge, Z.-W., Griffith, G.W., Griffiths, K., Groenewald, J.Z., Groenewald, M., Grube, M., Gryzenhout, M., Guo, L.-D., Hagen, F., Hambleton, S., Hamelin, R.C., Hansen, K., Harrold, P., Heller, G., Herrera, C., Hirayama, K., Hirooka, Y., Ho, H.-M., Hoffmann, K., Hofstetter, V., Hognabba, F., Hollingsworth, P.M., Hong, S.-B., Hosaka, K., Houbraken, J., Hughes, K., Huhtinen, S., Hyde, K.D., James, T., Johnson, E.M., Johnson, J.E., Johnston, P.R., Jones, E.B.G., Kelly, L.J., Kirk, P.M., Knapp, D.G., Koljalg, U., Kovacs, G.M., Kurtzman, C.P., Landvik, S., Leavitt, S.D., Liggenstoffer, A.S., Liimatainen, K., Lombard, L., Luangsa-ard, J.J., Lumbsch, H.T., Maganti, H., Maharachchikumbura, S.S.N., Martin, M.P., May, T.W., McTaggart, A.R., Methven, A.S., Meyer, W., Moncalvo, J.-M., Mongkolsamrit, S., Nagy, L.G., Nilsson, R.H., Niskanen, T., Nyilasi, I., Okada, G., Okane, I., Olariaga, I., Otte, J., Papp, T., Park, D., Petkovits, T., Pino-Bodas, R., Quaedvlieg, W., Raja, H.A., Redecker, D., Rintoul, T.L., Ruibal, C., Sarmiento-Ramirez, J.M., Schmitt, I., Schussler, A., Shearer, C., Sotome, K., Stefani, F.O.P., Stenroos, S., Stielow, B., Stockinger, H., Suetrong, S., Suh, S.-O., Sung, G.-H., Suzuki, M., Tanaka, K., Tedersoo, L., Telleria, M.T., Tretter, E., Untereiner, W.A., Urbina, H., Vagvolgyi, C., Vialle, A., Vu, T.D., Walther, G., Wang, Q.-M., Wang, Y., Weir, B.S., Weiss, M., White, M.M., Xu, J., Yahr, R., Yang, Z.L., Yurkov, A., Zamora, J.-C., Zhang, N., Zhuang, W.-Y., Schindel, D., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. 109, 6241–6246.

Schwaller, L., Robin, S., Stumpf, M., 2015. Bayesian Inference of Graphical Model Structures Using Trees. arXiv:1504.02723.

Shehzad, W., Riaz, T., Nawaz, M.A., Miquel, C., Poillot, C., Shah, S.A., Pompanon, F., Coissac, E., Taberlet, P., 2012. Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. Mol. Ecol. 21, 1951–1965.

Sher, D., Thompson, J.W., Kashtan, N., Croal, L., Chisholm, S.W., 2011. Response of Prochlorococcus ecotypes to co-culture with diverse marine bacteria. ISME J. 5, 1125–1132.

Smith, M.A., Eveleigh, E.S., McCann, K.S., Merilo, M.T., McCarthy, P.C., Van Rooyen, K.I., 2011. Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. PLoS One 6, e14424.

Smith, M.A., Rodriguez, J.J., Whitfield, J.B., Deans, A.R., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proc. Natl. Acad. Sci. 105, 12359–12364.

Srinivasan, A., Page, D., Camacho, R., King, R., 2006. Quantitative pharmacophore models with inductive logic programming. Mach. Learn. 64, 65–90.

Srivathsan, A., Sha, J.C.M., Vogler, A.P., Meier, R., 2014. Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). Mol. Ecol. Resour. 15, 250–261.

Staudacher, K., Jonsson, M., Traugott, M., 2015. Diagnostic PCR assays to unravel food web interactions in cereal crops with focus on biological control of aphids. J. Pest Sci. http://dx.doi.org/10.1007/s10340-015-0685-8.

Stauffer, D., Aharony, A., 1994. Introduction to Percolation Theory. CRC Press, Taylor & Francis Group, London.

Stockinger, H., Krüger, M., Schüssler, A., 2010. DNA barcoding of arbuscular mycorrhizal fungi. New Phytol. 187, 461–474.

Sun, D.-L., Jiang, X., Wu, Q.L., Zhou, N.-Y., 2013. Intragenomic heterogeneity in 16S rRNA genes causes overestimation of prokaryotic diversity. Appl. Environ. Microbiol. 79, 5787.

Symondson, W.O.C., 2002. Molecular identification of prey in predator diets. Mol. Ecol. 11, 627–641.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21, 2045–2050.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., Willerslev, E., 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 35, e14.

Tamaddoni-Nezhad, A., Bohan, D., Raybould, A., Muggleton, S.H., 2012. Machine learning a probabilistic network of ecological interactions. In: Proceedings of the 21st International Conference on Inductive Logic Programming, LNAI 7207, Springer, Berlin, pp. 332–346.

Tamaddoni-Nezhad, A., Bohan, D.A., Raybould, A., Muggleton, S., 2015. Towards machine learning of predictive models from ecological data. In: Proceedings of the International Conference on Inductive Logic Programming, Springer, Berlin.

Tamaddoni-Nezhad, A., Milani, G.A., Raybould, A., Muggleton, S., Bohan, D.A., 2013. Construction and validation of food webs using logic-based machine learning and text mining. Adv. Ecol. Res. 49, 225–289.

Thébault, E., Fontaine, C., 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329, 853–856.

Thébault, E., Loreau, M., 2006. The relationship between biodiversity and ecosystem functioning in food webs. Ecol. Res. 21, 17–25.

Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., Martinez, N.D., Rantala, H., Romanuk, T.N., Stouffer, D.B., Tylianakis, J.M.,

2012. Food webs: reconciling the structure and function of biodiversity. Trends Ecol. Evol. 27, 689–697.
Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. J. R. Stat. Soc. Ser. B 73, 273–282.
Tilman, D., Lehman, C.L., Bristow, C.E., 1998. Diversity–stability relationships: statistical inevitability or ecological consequence? Am. Nat. 151, 277–282.
Tilman, D., Reich, P.B., Knops, J.M.H., 2006. Biodiversity and ecosystem stability in a decade-long grassland experiment. Nature 441, 629–632.
Todorovski, L., Džeroski, S., Kompare, B., 1998. Modelling and prediction of phytoplankton growth with equation discovery. Ecol. Modell. 113, 71–81.
Traugott, M., Kamenova, S., Ruess, L., 2013. Empirically characterising trophic networks: what emerging DNA-based methods, stable isotope and fatty acid analyses can offer. Adv. Ecol. Res. 49, 177–224.
Tylianakis, J.M., Tscharntke, T., Lewis, O.T., 2007. Habitat modification alters the structure of tropical host–parasitoid food webs. Nature 445, 202–205.
U'Ren, J.M., Riddle, J.M., Monacell, J.T., Carbone, I., Miadlikowska, J., Arnold, A.E., 2014. Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. Mol. Ecol. Resour. 14, 1032–1048.
Vacher, C., Daudin, J.-J., Piou, D., Desprez-Loustau, M.-L., 2010. Ecological integration of alien species into a tree-parasitic fungus network. Biol. Invasions 12, 3249–3259.
Vacher, C., Piou, D., Desprez-Loustau, M.-L., 2008. Architecture of an antagonistic tree/fungus network: the asymmetric influence of past evolutionary history. PLoS One 3, e1740.
Vanderklift, M.A., Ponsard, S., 2003. Sources of variation in consumer-diet? 15N enrichment: a meta-analysis. Oecologia 136, 169–182.
Vayssier-Taussat, M., Albina, E., Citti, C., Cosson, J.-F., Jacques, M.-A., Lebrun, M.-H., Le Loir, Y., Ogliastro, M., Petit, M.-A., Roumagnac, P., Candresse, T., 2014. Shifting the paradigm from pathogens to pathobiome: new concepts in the light of meta-omics. Front. Cell. Infect. Microbiol. 4, 29.
Whittaker, J., 1990. Graphical Models in Applied Multivariate Statistics. Wiley Publishing, New York.
Wirta, H., Hebert, P.D.N., Kaartinen, R., Prosser, S., Várkonyi, G., Roslin, T., 2014. Complementary molecular information changes our perception of food web structure. Proc. Natl. Acad. Sci. 111, 1885–1890.
Yachi, S., Loreau, M., 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proc. Natl. Acad. Sci. 96, 1463–1468.
Yang, E., Ravikumar, P., Allen, G.I., Liu, Z., 2013. On Poisson graphical models. Adv. Neural Inform. Process. Syst. 26, 1718–1726.
Zilber-Rosenberg, I., Rosenberg, E., 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol. Rev. 32, 723–735.