Unravelling the web of dark interactions: Explainable inference of the diversity of microbial interactions

Didac Barroso-Bergada[a,*], Alireza Tamaddoni-Nezhad[b], Dany Varghese[b], Corinne Vacher[c], Nika Galic[d], Valérie Laval[e], Frédéric Suffert[e], and David A. Bohan[a]

[a]Agroécologie, AgroSup Dijon, INRAE, Université de Bourgogne Franche-Comté, Rue Sully, Dijon, France
[b]University of Surrey, Guildford, United Kindgom
[c]BIOGECO, INRAE, Univ. Bordeaux, Pessac, France
[d]Syngenta AG, Basel, Switzerland
[e]Université Paris-Saclay, INRAE, UR BIOGER, Palaiseau, France
*Corresponding author. e-mail address: barroso.didac@gmail.com

Contents

Abstract

The functional diversity of microbial communities emerges from a combination of the great number of species and the many interaction types, such as competition, mutualism, predation or parasitism, in microbial ecological networks. Understanding the relationship between microbial networks and the functions delivered by the microbial communities is a key challenge for microbial ecology, particularly as so many of these

Advances in Ecological Research
ISSN 0065-2504, https://doi.org/10.1016/bs.aecr.2023.09.005

interactions are difficult to observe and characterise. We believe that this 'Dark Web' of interactions could be unravelled using an explainable machine learning approach, called Abductive/Inductive Logic Programming (A/ILP) in the R package InfIntE, which uses mechanistic rules (interaction hypotheses) to infer directly the network structure and interaction types. Here we attempt to unravel the dark web of the plant microbiome in metabarcoding data sampled from the grapevine foliar microbiome. Using synthetic, simulated data, we first show that it is possible to satisfactorily reconstruct microbial networks using explainable machine learning. Then we confirm that the dark web of the grapevine microbiome is diverse, being composed of a range of interaction types consistent with the literature. This first attempt to use explainable machine learning to infer microbial interaction networks advances our understanding of the ecological processes that occur in microbial communities and allows us to hypothesise specific types of interaction within the grapevine microbiome. This work will have potentially valuable applications, such as the discovery of antagonistic interactions that might be used to identify potential biological control agents within the microbiome.

1. Introduction

The high taxonomic, morphological and functional diversity of microbial communities (Konopka, 2009) emerges from a combination of the great number of species and the many interaction types in the ecological networks of the microbiome. These ecological interactions result in economically and socially important ecosystem functions and disease regulation (Hacquard et al., 2017; Ishaq, 2017), and structure the communities of microbial organisms observed in nature (Faust and Raes, 2012). The abundance of any two species in a community will be determined by whether they participate in a pairwise interaction and the precise ecological interaction type, as well as being influenced by other biotic and abiotic factors (de Vries et al., 2018). Where the abundance of both species is observed to decline, this might be hypothesised to come about as a result of a competitive interaction. Predation could lead to an increase in abundance of one species at the expense of the other (Faust and Raes, 2012). Mutualistic interactions might result in both species increasing in abundance, while amensalism and commensalism would cause a differential benefit or cost to the abundance of only one of the species. It is the mixture of all these different interactions, acting between all species in the ecological network simultaneously, which determines the species richness, diversity patterns, functions and dynamics of microbial communities.

Understanding the relationship between microbial interactions and the functions delivered by the microbiome is a key challenge for microbial ecology. We expect, for example, that interactions such as competition or predation will be associated with ecosystem functions like biological control of microbial pathogens (Musetti et al., 2007; Poveda et al., 2021). The great challenge is that these interactions can be difficult to evaluate directly, due to the complexities of observing microbial species in-situ or culturing microbial species in the laboratory (Crhanova et al., 2019; Wu et al., 2019). Moreover, the types of interaction are well described in the macrobial world, but their microbial equivalent may not be obvious. Many microbial taxa also remain unknown, forming what has been termed microbial "Dark Matter" (Marcy et al., 2007). As a consequence, much of the diversity and structure of microbial ecological networks are unobserved; they are "Dark Webs". This contributes to our poor understanding of these systems and limits the advance of the science of the microbiome. In this paper, we attempt to unravel the dark web of the microbiome using direct inference of specific types of ecological interactions from DNA metabarcoding data.

Analysis of microbial species and communities in situ has been greatly facilitated by metabarcoding surveys of environmental DNA samples using generic primers (Ruppert et al., 2019; Thomsen and Willerslev, 2015). The sequencing yields information on the copies of each DNA sequence corresponding to quantitative information for the count of the different taxa. Pipelines like VSEARCH (Rognes et al., 2016) cluster the counts into operational taxonomic units (OTU) by their similarity (Pauvert et al., 2019), under the assumption that sequences with the greatest similarity represent phylogenetically similar organisms (He et al., 2015). Other pipelines like DADA2 (Callahan et al., 2016) infer high resolution versions of OTUs, called amplicon sequence variants (ASVs), which vary due to possible sequencing errors (Davis et al., 2018). DNA sequencers can only process a given number of DNA sequences in any given metabarcoding run. The sequencing data produced is therefore compositional (Gloor et al., 2017), reflecting the relative but not absolute abundance of the species within each sample, and such biases of compositionality have typically been controlled using log-transformation.

Statistical associations between counts have been used to infer possible interactions between OTUs (Faust and Raes, 2012; Röttjers and Faust, 2018; Weiss et al., 2016). SparCC (Friedman and Alm, 2012) infers networks using linear Pearson correlation between the log-transformed

components, for example. CCLasso uses Lasso to infer the correlation network (Fang et al., 2015) and SPIEC-EASI uses inverse covariance or neighbourhood selection and StARS to obtain the most stable network (Kurtz et al., 2015; Liu et al., 2010). PLN-network uses the Poisson-LogNormal model, where sequence counts follow Poisson distributions, and introduces sequencing depth as an offset. It then uses methods like StARS or EBIC for network selection (Chiquet et al., 2019). The frameworks for network inference, such as those described above, have robust theoretical statistical foundations, and the tools are typically flexible, fast and robust to noise in the metabarcoding sample data (Dohlman and Shen, 2019; Weiss et al., 2016). The ecological importance of taxa (nodes) and associations (edges) of the networks that are produced need considerable interpretation and post-analysis, however. Positive or negative statistical associations do not indicate causality and are not specific indicators of a type of ecological interaction. This interpretation problem resembles the black box problem (Castelvecchi, 2016). The tool may detect that something is happening, but it cannot provide a mechanistic understanding of the underlying process. A mechanistic interpretation is only provided post-hoc via the literature and/or subjective expert knowledge (Tamaddoni-Nezhad et al., 2021). Here, we use a form of Explanatory machine learning (EXML) (Ai et al., 2021; Gilpin et al., 2019), called Abductive/Inductive Logic Programming (A/ILP), to circumvent the black box problem by detecting and classifying ecological interaction types directly.

A/ILP infers species interactions by searching for patterns in data using ecological rules or hypotheses for each type of interaction, defined a priori by the user. Such hypotheses can reach a large complexity degree, depending on the variables entailed. However, simpler explanations of a phenomenon are normally more plausible, or at least, should be the ones first tested. In consequence, the hypotheses of interactions used in this work limit the complexity by, (1) focusing on data obtained from samples sharing the same abiotic conditions. This removes, or greatly decreases the appearance of spurious interactions caused by the species affinity/disaffinity to certain abiotic conditions; (2) Only considering as hypothesised interactions the relations that cause a direct effect on the species abundance. Once the hypotheses of interaction are defined, the link between logical ecological rules and data patterns is straightforward. For instance, a competition interaction between any two species might be hypothesised to cause a decline in both their abundances when the species co-occur. In data of many samples from a similar environment, we can calculate directions of

change in species abundances between all pairs of samples and use this pattern of change to hypothesise species interactions. A species abundance might go up, down or stay the same between any two samples, from sample X to Y, for example. Under the assumption that species 1 and 2 are undergoing competition, then we might expect that a down change in abundance for species 1, from sample X to Y, would be accompanied by a down for species 2. Finding such a pattern of change consistently across many sample pairs, might then allow us to infer that the patterns in species abundances are due to competition between species 1 and 2. Different rules could also be used to infer other interaction types from the up, down or stay pattern. This link between rule and pattern in the data thereby provides informative, explainable networks with network edges that represent hypotheses of ecological interaction types.

A/ILP has previously been used to provide an explanation of metabolic regulation (Tamaddoni-Nezhad et al., 2006) and invertebrate food webs (Bohan et al., 2011). It has never been used on microbial community sequence data, however, or to infer numerous types of interaction simultaneously. This has in part been due to the intensive computational requirements of EXML approaches and their lower robustness to ecological noise. It may also be due to the much more limited experience of logical approaches amongst ecologists.

We tested the A/ILP approach, detailed here through two steps. The first step uses synthetic data for different types of interactions, in order to verify that we can both detect and classify interactions by their correct type, with appropriate levels of significance. The A/ILP is then used at the second step to unravel the dark webs of the different interaction types from eDNA sample data. The eDNA was sampled from leaves of grapevine, *Vitis vinifera* L. during downy mildew epidemics, caused by *Plasmopara viticola* (Barroso-Bergadà et al., 2023).

2. Materials and methods

2.1 Hypothesis framework for learning microbial ecological interactions using abductive logic

Explainable approaches to inferring ecological interactions start with a clear declaration of the rules for an ecological interaction between any two, or possibly more, OTUs (Faust and Raes, 2012; Tshikantwa et al., 2018). We posit that the minimum common facts for all hypothesised interactions are

that: (i) the two OTUs undergoing an interaction should co-occur in at least one sample; and, (ii) at least one of the OTUs undergoes a change in abundance. Abundance is understood as a measure of the size of an OTU population in a sample, derived from the copies of OTU sequences in each sample. Thus, to evaluate change in abundance of all OTUs, across all samples, the copies of an OTU and the total sequence depth in any two samples, are used to construct a contingency table, with the significance of the change in OTU abundance between the samples being evaluated by a $\chi 2$-test of independence. Thus, the compositional biases inherent in treating DNA sequences as relative counts are addressed by assessing the independence of the OTU change of counts between samples from the changes in the samples sequencing depth. Significant changes are classified either as an increase (up), or as a decrease (down), in terms of the relative abundance of the first sample compared to the second. Symbolically, this can be expressed as the logic clause abundance (s1, x, y, dir) where s1 is any given OTU, (x, y) are two samples and dir describes the direction of abundance change (up or down) from sample x to sample y. Abundance changes are computed in this way across all OTUs in all samples. I In addition, the presence (yes) or absence (no) of an OTU in sample x can be expressed as the clause: presence (s1, x, yes/no).

The abductive logic process uses these clauses to find possible explanations (effects) for the observed changes in abundance and presence using hypotheses for ecological interactions that reflect the existing state of ecological knowledge. We hypothesise that an interaction will have occurred where the presence of one OTU s1 produced a consistent effect on the abundance of OTU s2 in the samples. The relationships for an interaction effect can be described as:

$$\begin{aligned} &\text{abundance}(s2,\ x,\ y,\ up) \\ \text{effect\_up}(s1, s2)\ \textit{if}: \quad &\text{presence}(s1,\ x,\ no) \\ &\text{presence}(s1,\ y,\ yes) \end{aligned}$$

$$\begin{aligned} &\text{abundance}(s2,\ x,\ y,\ down) \\ \text{effect\_down}(s1, s2)\ \text{if}: \quad &\text{presence}(s1,\ x,\ no) \\ &\text{presence}(s1,\ y,\ yes) \end{aligned} \tag{1}$$

Here, the upper part of the relation describes how OTU s2 has a greater abundance in sample y than in sample x, due to the presence of OTU s1 in sample y and its absence in sample x. Should this pattern be consistent across different sample pair combinations, then the abduction process

would infer an up effect of s1 on s2. A pattern is considered consistent if the number of observations supporting such pattern is considerably larger than the number of observations contradicting it.

The abduction process computes a compression value as a numerical measure representing the amount of observations that support each abduced effect (Muggleton and Bryant, 2000), and therefore indicates how likely an interaction is to have occurred. To be likely, an interaction should give a greater effect in one direction than the other, all other things being equal. We therefore compute an overall statistic for the likelihood of interaction, I, as the difference between the compressions for the up and down effects (Barroso-Bergada et al., 2022).

2.1.1 Detection of significant interactions

For each pair of OTUs, the value of I is treated as the weight of a directed edge in a network. Setting a threshold, λ for I, selects inferred edges. $\lambda = 0$ would select all possible edges, while a $\lambda = \max(I)$ would deliver an empty network with no edges selected. $\max(I)$ is dependent on the number of observations, and it is not possible to establish a common λ for the datasets. We therefore select significant edges empirically using a subsampling methodology called StARS (Liu et al., 2010). StARS subsamples 80% of the samples, multiple times, and performs the abduction of network edges. The most stable network of interactions is then identified using the frequency that the edges appear at different values of λ. We use 50 resamplings and 50 λ values increasing from 0 to $\max(I)$. The standard number of subsamples and length of lambda path was selected with a restrictive stability threshold of 0.01 to minimise the number of false positive interactions (Müller et al., 2016).

2.1.2 Classification of interaction types

The StARS procedure selects those edges of a network that have a consistent direction of effect and may therefore be treated as significant. The direction of these detected interactions, up, down or no effect may be used to classify interaction types directly (Derocles et al., 2018; Faust and Raes, 2012). Thus, effect_up(s2, s1) and s2 effect_up(s1, s2) might be classified as mutualism since both OTUs benefit. effect_down(s2, s1) and effect_down(s1, s2), by contrast, could be assigned to a competition interaction. Across all possible inferred pairwise combinations of up, down and no change, it becomes possible to classify directly ecological interactions of mutualism, predation, competition, commensalism and amensalism (Table 1, Faust and Raes, 2012), that can often not be observed or measured in microbial experiments.

Table 1 Interaction types as described in Derocles et al., 2018.

Type of interaction	Effect on taxa A abundance	Effect on taxa B abundance	Description
Amensalism	0	↓	Taxa A causes a decrease (down) on the abundance of taxa B without suffering any effect on the abundance
Commensalism	0	↑	Taxa B increases (up) its abundance thanks to the effect of Taxa A
Competition	↓	↓	Both taxa abundance decreases (down) by the effect of the other. This can be caused by direct competition (directly harming the other taxa) or exploitation competition (they need the same resource an in consequence there is less available)
Mutualism	↑	↑	The abundance of both taxa increases (up) by the effect of the other.
Neutralism	0	0	Both taxa co-occur but there is no effect on their abundance, and therefore, no interaction.
Parasitism or Predation	↑	↓	Taxa A develops (up) at the expense (down) of taxa B.

2.1.3 Modelling ecological exclusion

Ecological interactions that cause the abundance of an OTU to decrease can lead, *in extremis*, to its exclusion. Some OTUs may also depend upon the presence of a second OTU in order to exist within a sample. The hypotheses of interaction do not explicitly account for this possibility that could affect the numbers of zeros in the data and the detection of interactions. We expand the hypothesis of interaction framework to entail exclusion and mutual dependence cases. This can be described by logical clauses with the form abundance(s1, x, y, app/dis), where app symbolises a qualitative change from 0 to a numbers of counts of OTU s1 between the samples x and y, and dis symbolises the change from a positive number of counts to 0. Significance of change is evaluated using a $\chi 2$-test of independence. The theory is expanded as:

$$\begin{aligned} &\text{abundance}(s2,\ x,\ y,\ up/app) \\ \text{effect\_up}(s1, s2)\ \text{if:}\quad &\text{presence}(s1,\ x,\ no) \\ &\text{presence}(s1,\ y,\ yes) \end{aligned}$$

$$\begin{aligned} &\text{abundance}(s2,\ x,\ y,\ down/dis) \\ \text{effect\_down}(s1, s2)\ \text{if:}\quad &\text{presence}(s1,\ x,\ no) \\ &\text{presence}(s1,\ y,\ yes) \end{aligned} \tag{2}$$

It is important to note that for this 'with exclusion' formulation of the theory the effect caused by s1 on s2 or vice versa, is consistent irrespective of whether the result is an *up/down* or *app/dis*. This explicitly includes the possibility that an interaction could cause both a reduction in the abundance of an OTU, and, ultimately, its exclusion. This formulation also allows the I statistic to be computed as previously described.

2.1.4 Implementation of the abductive process for inferring ecological interactions

A full description of the logical process of abduction and of A/ILP is not given here, but can be found in (Muggleton and Bryant, 2000; Tamaddoni-Nezhad et al., 2006). We detail the specific implementation of abductive network inference in a new R package, InfIntE (INFerence of INTeractions using Explainable machine learning, Fig. 1). InfIntE parses the OTU count data into logical clauses using the R base package tools (Core Team, 2018). The logical clauses and interaction hypotheses are run in PyGol, which implements A/ILP based on Meta Inverse Entailment (MIE). The hypothesis searching mechanisms of MIE combining both

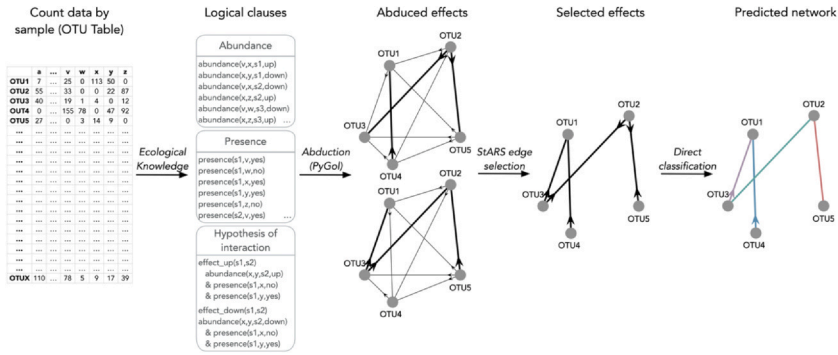

Fig. 1 Schematic diagram representing the InfIntE pipeline. The pipeline performs: conversion of abundance data contained in the OTU table to logical clauses, based upon our ecological knowledge; abduction of interaction effects from the logical clauses using PyGol; selection of important edges using StARS; and the direct classification of interaction types depicted by the different edge colours. Arrows show the direction of effects or interactions. Edges without an arrow represent interactions without direction.

top-down and bottom-up approaches as the hypothesis is more generalised than bottom clause and more specific than meta theory (Varghese et al., 2022).

InfIntE uses the R package reticulate (Ushey et al., 2022) to provide the logical clauses to PyGol and then retrieve the abduced compression values and I statistic. Pygol searches for the best abducible effects on OTU abundance that explain the observation using maximum compression. Model selection by StARS is then conducted in the R pulsar package (Müller et al., 2016). A custom pulsar function uses the meta clause to abduce subsets of the OTU table and retrieves the interaction networks along the λ path.

2.2 Experiment 1: Generating synthetic, ecological-like data for verification

Numerous approaches have been proposed to simulate the effect of interactions in metabarcoding data, based on the assumption that interactions cause a change in the sequence counts of OTUs involved. We use the broadly-accepted linear ecological-like models proposed by Weiss et al. (2016) (Tackmann et al., 2019; Weiss et al., 2016). These models simulate changes in counts in an OTU table, caused by specific interaction types. The Weiss et al. (2016) models produce the OTU tables for the counts of p non-interacting OTUs over n samples using a log-normal distribution

(Shoemaker et al., 2017). These counts are then forced to increase or decrease as a function of the counts of the interacting OTUs, modulated by a strength of interaction, s. The generated OTU tables consist of abundances of p OTUs simulating either amensalism, commensalism, competition or mutualism pairwise interactions as proposed in Faust and Raes (2012). Each simulated interaction type has a different effect on the abundance of the OTUs involved in the interaction. For example, in mutualism, the abundance of both OTUs will increase in the samples where they co-occur. To introduce compositionality to the data, the relative counts of an OTU in a sample are used as a probability to sample a multinomial distribution at a common sample size.

The number of samples is an important variable in network inference (Berry and Widder, 2014). We generate OTU tables with different number of samples, n, to assess the effects of sampling effort. For each $n = 20$, 30, 40.90, we create three OTU tables with strengths of interaction, $s = 2$, 3 and 5, and $p = 80$ OTUs to obtain a total of 72 OTU tables incorporating the four types of interactions. The number of $p = 80$ OTUs was chosen to reflect the number of abundant OTUs typically observed in real metabarcoding microbial datasets from agriculture, such as the one used in this work.

2.2.1 Inference and detection of interactions from simulated data

InfIntE was used to infer interaction networks for each of the simulated OTU tables, for interaction hypotheses both with and without exclusion. The area under the receiver operating characteristic curve (AUC) (Fan et al., 2006) was then evaluated. The AUC was treated as a measure of how well the tool detected interactions that we knew to be present (real), or absent (false) in the simulated dataset. Interaction inference was also done for the statistical inference tools, SparCC (Friedman and Alm, 2012) and SPIEC-EASI glasso (Kurtz et al., 2015), as a comparison for interaction detection between logical and statistical networks inference tools. The SparCC inference was done in FastSpar v1.0.0 (Watts et al., 2019) and SPIEC-EASI glasso was run in the R package SpiecEasi v1.1.2, both with their respective default settings. The I statistic was computed with and without exclusion hypotheses in InfIntE. SparCC correlations were obtained directly and SPIEC-EASI correlations were obtained from the inverse covariance matrix at $\lambda = 0$. Given that these logical and statistical tools produce either classified interaction or correlational networks respectively, the largest value of I or correlation, obtained for each pair of OTUs, was used to compare the performance of the tools.

2.2.2 Evaluating the accuracy of interaction detection and classification in simulated datasets

For interactions with and without exclusion, the accuracy of InfIntE and the StARS procedure to detect simulated interactions, was computed using the function:

$$\text{Accuracy} = (\text{TP} + \text{TN})/\text{N} \tag{3}$$

where true positives, TP, are the true real simulated interactions detected by the StARS selection, TN are the true non-interacting pairs of OTUs within the simulated dataset, and N is the total number of possible interactions that might exist in a fully saturated network. Accuracy was selected as an overall measure to summarise the total capacity of the tools to accurately detect interactions. The evaluation of SparCC was performed using the default bootstrapping procedure, with 999 permutations. The SPIEC-EASI pipeline uses the StARS procedure to select important interactions as a function of edge stability. The default parameters of StARS selection in SPIEC-EASI are 20 subsamples and a stability threshold of 0.05.

2.3 Experiment 2: Inferring networks from real data

We used InfIntE to reconstruct microbiome interaction networks occurring in leaves of European cultivated grapevine, *Vitis vinifera*. We reconstructed 9 networks, each corresponding to a different vineyard in France, using qPCR and metabarcoding data described and analysed in Barroso-Bergadà et al. (2023). This dataset includes the relative abundance of 650 fungal ASVs in 534 leaf samples (60 per vineyard). All leaf samples collected within a vineyard were collected in similar conditions (on the same day and on the same rows of vines). The abundance of the oomycete pathogen species *Plasmopara viticola* was evaluated by qPCR in each leaf sample. For each vineyard, patterns of change in the abundance of fungal ASVs between samples were calculated as described above. In addition, patterns of change in the abundance of *P. viticola* between samples were computed using the qPCR data. The up/down was considered significant if the logarithmic absolute amount of *P. viticola* DNA differed by an absolute value of 0.05 between samples.

Fungi–fungi interaction networks, including the oomycete *P. viticola*, were then inferred using the InfIntE package (https://github.com/didacb/InfIntE). Network inference was performed for the top 80 most abundant fungal ASVs, using all 60 samples from each vineyard.

2.3.1 Evaluating the significance of interactions using predictions of change

Most microbial community interactions are unknown or poorly understood. There is no complete and understood microbial network that might be used to evaluate the accuracy of the interactions detected and classified by InfIntE, of which we are aware. It is necessary, therefore, to evaluate the significance of inferred interactions using the data itself. Changes in OTU abundance, caused by an ecological interaction, can be used to compute predictive accuracy. InfIntE does this using a k-fold cross-validation that predicts abundance changes using the I statistic. Abundance change observations are randomly divided into 5 equal size, 20%, folds of the dataset. Interaction inference is then performed using InfIntE on 4 random selections of these folds, leaving one fold for validation. The I statistic for the effects of OTUs on, for example, s1 in sample y, are retained and the up/down is predicted from the I sum value of effect up and effect down. Predictive accuracy is then tested by computing the number of correctly predicted abundance changes across the validation fold.

2.4 Statistical analysis

All statistical analyses were conducted in the R v4.1.3 (Core Team, 2022). Plots were made using ggplot2 v3.3.5 (Wilkinson, 2011) and cowplot v1.1.1 (Wilke, 2019). The AUC of each inferred network was measured using the pROC package v1.18.0 (Robin et al., 2011).

3. Results

3.1 Experiment 1: Generating synthetic, ecological-like networks

3.1.1 Modelling exclusion increases I statistic predictive power

The InfIntE AUC for the I statistic was higher when interactions were inferred with than without exclusion (Fig. 2A). This difference was greater for datasets with higher strengths of interaction and for smaller sample sizes. Sample size had an important effect on the AUC, independent of the strength of interaction and the hypothesis of interaction used, plateauing at about 60 samples. The InfIntE AUC varied by up to 20% between hypotheses without and with exclusion and from 20 to 90 samples. The AUC obtained for the InfIntE I statistic with exclusion

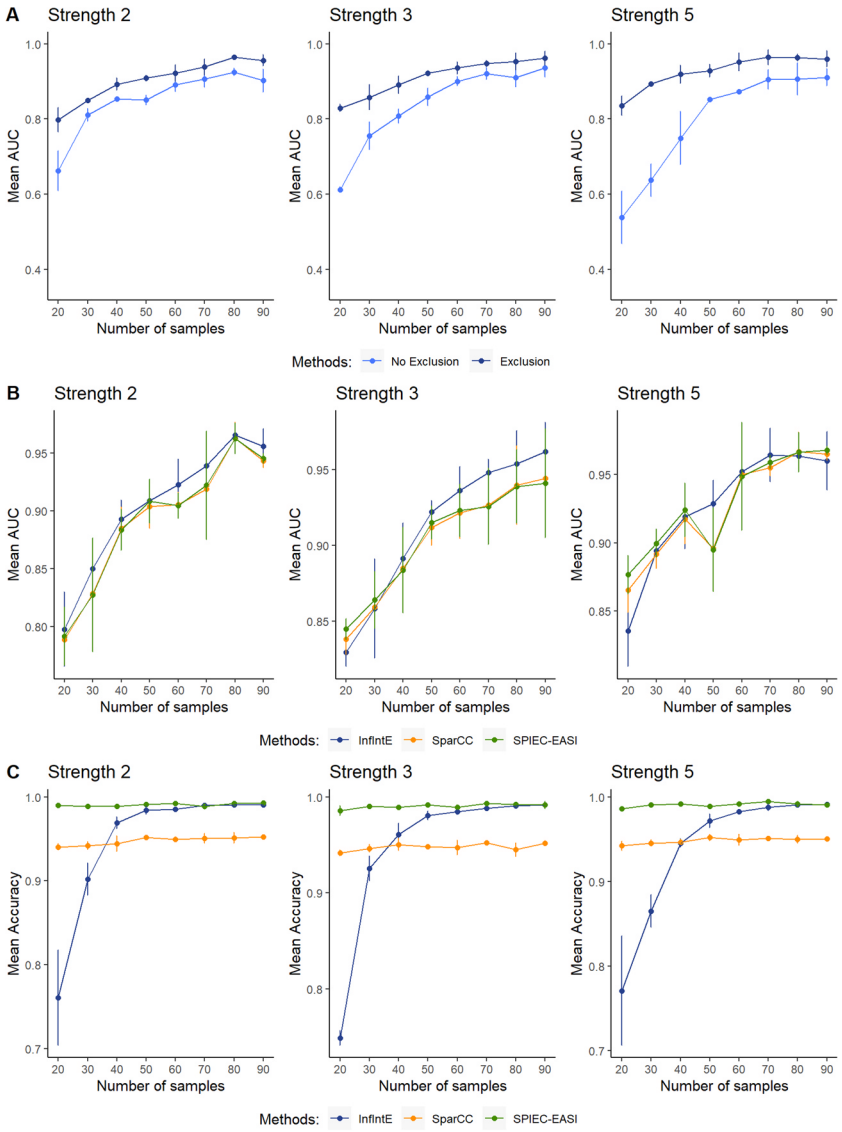

Fig. 2 Relationship between number of samples and interaction inference performance for different strengths of interaction. Datasets were computer-generated simulating four different interaction types: amensalism, commensalism, competition and mutualism. (A) Area under the roc curve values (AUC) (Fan et al., 2006) obtained by I statistic with and without exclusion. Larger AUC values represent better specificity and sensitivity in interaction detection. I statistic is used by InflntE as a numeric measure of interaction. (B) Area under the roc curve values (AUC) obtained by InflntE's I statistic and SparCC and SPIEC-EASI correlation like measures. InflntE used the hypothesis of interactions including exclusion. SparCC and SPIEC-EASI were executed with default settings. (C) Accuracy of interaction detection computed as described in Section 2.2.2. InflntE used the hypothesis of interactions including exclusion. SparCC and SPIEC-EASI were executed with default settings.

was similar to the AUC obtained by the correlation values computed in SparCC or in SPIEC-EASI (Fig. 2B). The AUC for SparCC and SPIEC-EASI, varied with the number of samples. The strength of simulated ecological interaction had a negligible effect. No significant difference was found in AUC values between the three network inference tools.

3.1.2 Accuracy of InfIntE detection with sample size

The accuracy of InfIntE with exclusion varied significantly with the sample size (Fig. 2C), increasing to a plateau at approximately 60 samples. InfIntE accuracy did not change with the strength of simulated interaction. SparCC and SPIEC-EASI showed levels of accuracy that did not depend on the number of samples, but SPIEC-EASI accuracy was higher than SparCC. InfIntE accuracy was lower than SPIEC-EASI and SparCC at small sample sizes, but was comparable to SPIEC-EASI at higher sampling effort.

3.1.3 InfIntE classification of simulated interactions

The InfIntE classification of interactions at 60 samples, was dependent on the interaction type simulated (Fig. 3). More than 90% of the synthetic commensalism interactions were detected, and of all of these were classified correctly. Most mutualism interactions were detected by InfIntE, and the majority were classified correctly. Those wrongly classified were classed as commensalism. Slightly less than 20% of competition interactions were detected by InfIntE, but nearly all were classified correctly. Synthetic amensalism interactions were not detected by InfIntE. InfIntE produced low numbers of false positives. Where these occurred, all detected links were classified as commensalism.

SparCC detected as associations almost all the synthetic mutualism, commensalism and competition interactions. About 25% of the simulated amensalism interactions were detected. This detection came at the cost of an elevated rate of false positive detections. SPIEC-EASI showed association detection performance similar to InfIntE, detecting associations for the majority of synthetic commensalism and mutualism interactions. It had similar difficulty in detecting synthetic associations from competition and poor performance in detecting amensalism. SPIEC-EASI produced very low rates of false positives.

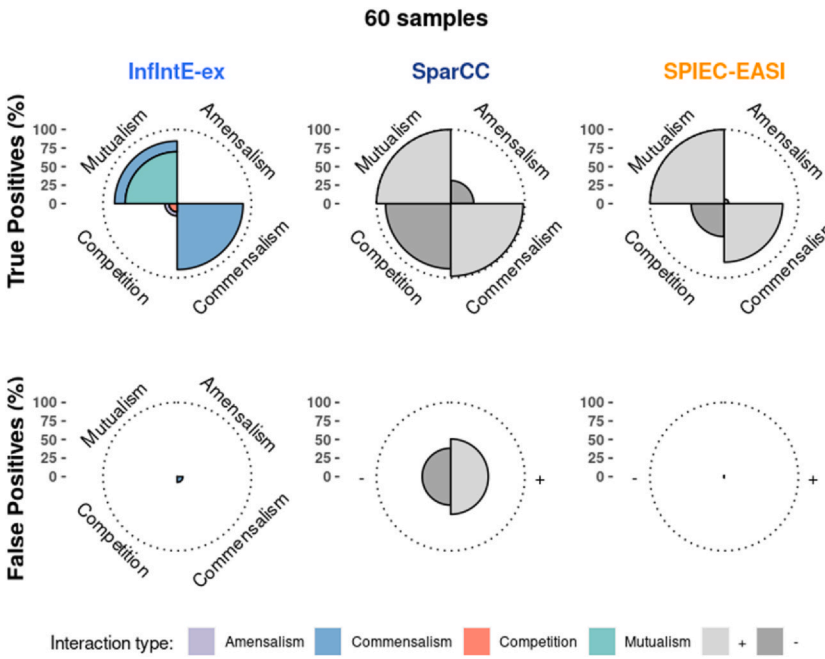

Fig. 3 Nightingale rose charts comparing the percentage of correct interaction classification by types. OTU tables were synthetically generated simulating groups of 60 replicated samples mixing four different types of interactions: amensalism, commensalism, competition and mutualism. Inference of interactions was performed using InfIntE, SparCC and SPIEC-EASI. First row charts show the percentage of each interaction type correctly detected by each tool. Second row charts show the percentage of false positives proposed by each tool over the total possible false positives. Each petal of the rose charts is coloured in function of the classification of the detected interaction type given for each inference tool. InfIntE automatically classifies interactions in amensalism, commensalism, competition and mutualism while SparCC and SPIEC-EASI return positive (+) or negative (−) associations. InfIntE correctly detects and classifies most mutualism and commensalism as well as around 20% of competition with few false positives. SparCC detects most interactions at the expense of a large amount of false positives. SPIEC-EASI has a similar performance to InfIntE, but classifying the interactions only in positive and negative correlations.

3.2 Experiment 2: Inferring complex networks, the Dark Web, from real data

3.2.1 Structure of interaction types in the foliar networks of grapevine

The interaction networks inferred for each of the nine vineyards (Fig. 4), using InfIntE, comprise all the different types of ecological interaction (Derocles et al., 2018, Table 1). The networks did not display unconnected

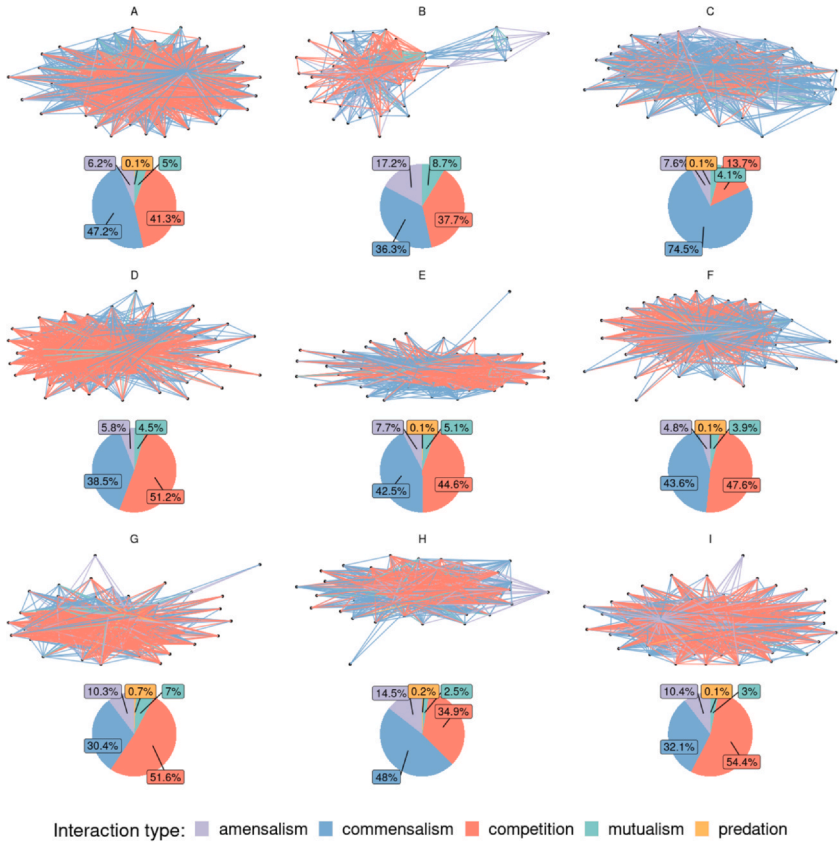

Fig. 4 Interaction networks predicted for each of the 9 different vineyards in the dataset, inferred using InfIntE. Each vineyard dataset was composed by 60 samples. The edge colours follow the interaction colour typology in A. The pie chart associated with each network indicates the relative percentage of each interaction type in the network.

subnetworks, and the total number of hypothesised interactions varied from 551 to 1410, depending on the vineyard. The predominant interactions were commensalism and competition, each representing at least 30% of the total number of interactions in most networks. Amensalism interactions never made up more than 17% of interactions and mutualism varied from 2.5% to 8.7%. Interactions classified as predation never represented more than 0.7% of the inferred interactions. Predictive accuracy was estimated to be approximately 75% of the observations in the test fold (Fig. 5), when at least 50% of the sample dataset was used. No difference in accuracy was found between the vineyards for a given standard percentage of the sample.

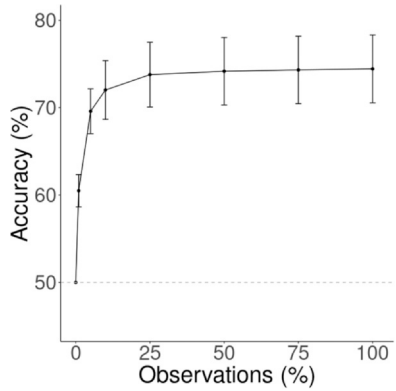

Fig. 5 Accuracy of abundance change prediction in grapevine metabarcoding data as a function of the number of observations used for the inference. Each dataset consisted in 60 samples. Each point presents the mean accuracy of prediction for different fold combinations of the 9 vineyards.

3.2.2 Identifying potential biological control agents using InfIntE inference

A total of 20 fungal ASVs, that could be assigned to 12 different fungal species and 2 genera, were identified, using the InfIntE package, as potential antagonists of *P. viticola* in at least one vineyard (Table 2). The list of potential antagonists included ASVs assigned to *Aureobasidium pullulans*, the *Alternaria* genus and the *Fusarium genus*, in coherence with expectations from the grapevine literature (Ghule et al., 2018; Harm et al., 2011; Musetti et al., 2006). It also included ASVs assigned to the genera Cladosporium (Amanelah Baharvandi and Zafari, 2015; Becker et al., 2020; Köhl et al., 2019), Phlebia (White and Boddy, 1992), Sporobolomyces (Filonow et al., 1996; Janisiewicz and Bors, 1995; Li et al., 2017) and Vishniacozyma (Becker et al., 2020; Gramisci et al., 2018; Lutz et al., 2013), in coherence with the biocontrol literature for other foliar pathosystems. ASVs assigned to *Mycosphaerella tassiana* and two species of the Filobasidium genus were also classified as antagonistic to *P. viticola* by inference, but had not previously been described in the literature.

4. Discussion

Our work shows that it is possible to begin to explore the Dark Web of microbial interactions, through the direct classification of interactions using explainable machine learning. The combination of A/ILP and simple

Table 2 Potential *Plasmopara viticola* antagonists found by InflntE. The table shows the fungal species found to have a potential interaction able to reduce the abundance of *P. viticola*. A bibliographical search in Google Scholar, Pubmed and Science Direct was conducted to identify whether potential antagonists have previously been described as biocontrol agents of *P. viticola* or other pathogens in the literature. The keywords used for the search were the name of the potential antagonist, "*Plasmopara viticola*", "biocontrol" and "antagonist". Those taxa identified with an asterisk were not automatically assigned to a taxonomic grouping in UNITE and required manual curation and assignment using BLAST. When there was more than one OTU assigned to the same species having the same interaction it is noted with *xn*, where *n* is the number of OTUs.

Name	Vineyard	Interaction	Bibliography against plasmopara	Bibliography biocontrol
Cladosporium delicatulum	I	competition		Köhl et al. (2019); Amanelah Baharvandi and Zafari (2015); Becker et al. (2020)
Mycosphaerella tassiana	I	competition		
Alternaria sp.*	A	amensalism	Musetti et al. (2006)	
*Alternaria alternata**	I	competition	Musetti et al. (2006, 2007)	
Alternaria brassicae	B	competition	Duhan et al. (2021)	
*Aureobasidium pullulans**	I	competition	Harm et al. (2011)	
Filobasidium chernovii	Ix2	competition		
*Filobasidium magnum**	D	competition		
Fusarium sp.*	A, B, E	competition	Ghule et al. (2018); Bakshi et al. (2001)	

(*continued*)

Table 2 Potential *Plasmopara viticola* antagonists found by InflntE. The table shows the fungal species found to have a potential interaction able to reduce the abundance of *P. viticola*. A bibliographical search in Google Scholar, Pubmed and Science Direct was conducted to identify whether potential antagonists have previously been described as biocontrol agents of *P. viticola* or other pathogens in the literature. The keywords used for the search were the name of the potential antagonist, "*Plasmopara viticola*", "biocontrol" and "antagonist". Those taxa identified with an asterisk were not automatically assigned to a taxonomic grouping in UNITE and required manual curation and assignment using BLAST. When there was more than one OTU assigned to the same species having the same interaction it is noted with *xn*, where *n* is the number of OTUs. (*cont'd*)

Name	Vineyard	Interaction	Bibliography against plasmopara	Bibliography biocontrol
Phlebia rufa	E	amensalism		White and Boddy (1992)
Sporobolomyces roseus	Ix3	competition		Janisiewicz and Bors (1995); Filonow et al. (1996)
*Sporobolomyces pararoseus**	A,G	competition		Li et al. (2017) (in grapes)
Vishniacozyma victoriae	B,C	amensalism, competition		Gramisci et al. (2018); Lutz et al. (2013)
Vishniacozyma carnescens	D	amensalism		Becker et al. (2020)

hypotheses for ecological interactions, and coded within the InfIntE R package, works satisfactorily. It shows accuracy of detection and classification of simulated interactions, given enough samples. We use it to detect and classify interaction types from real DNA sequence data. Validation reflects well the interactions that have previously been proposed in literature (Table 1). Our explainable machine learning is a step towards making the link between generic ecological hypotheses that scientists define and the inference of specific types of interaction within the microbiome that can be validated through experimentation. We find that the dark web of the microbiome is diverse, being composed of a range of interaction types consistent with the hypotheses of interaction we propose. These include antagonistic interactions that might be used to identify biocontrol agents of the pathogen P. viticola within the foliar microbiome of grapevines (Table 2).

The work also shows potential for improvement, by refining the description of existing hypotheses and the definition of new hypotheses for types of interaction not considered here. Explainable A/ILP emphasises the description of ecological interactions. The simple interaction hypotheses we have defined describe the process by which two species interact, contingent on the way that their respective abundances change (Derocles et al., 2018; Faust and Raes, 2012). We then extended the definition of these hypotheses to include the ecological process of exclusion, whereby the action of one species can cause the exclusion of the other, showing a marked increase in the power to discriminate true interactions. This indicates that our simple generic interaction hypotheses are only one representation of ecological interaction and could be further improved to infer interaction networks which better reflect the ecological reality of microbial interactions. The work of Bohan et al. (2011) showed that an A/ILP predation hypothesis, which included species traits, produced explainable invertebrate food webs. The networks that are inferred by A/ILP are therefore dependent upon and reflect the quality of our ecological knowledge and theory for the hypothesis of interaction.

Similarly, the hypotheses of interactions can be potentially expanded to include the sample abiotic conditions. Our work uses randomly taken samples obtained from the same vineyard at the same time to infer interactions. This hinders the possibility of introducing apparent interactions caused by abiotic factors (Derocles et al., 2018) in the inferred networks. In other cases, where the abiotic effects are more likely to cause a disruption on the inference, the introduction of known abiotic factors to the

hypotheses of interaction could allow to improve the inference. This would require, however, to correctly identify and measure such factors, adding new biases to the process.

There is no standard dataset in microbial ecology for evaluating network inference tools (Röttjers and Faust, 2018). We used simulation models to generate synthetic datasets of ecological-like interactions to evaluate and test the combined performance of InfIntE and our simple ecological rules. While these datasets are simplified approximations of the different typologies of interaction (Faust and Raes, 2012), our analysis suggests that the I statistic, computed in InfIntE, was able to discriminate interactions of different types accurately where sufficient samples were available. The sensitivity of inference to sample size is a matter of considerable debate in the literature (Hirano and Takemoto, 2019). For simulated datasets, we found that InfIntE network inference stabilised at around 60 samples. The strength of the simulated interactions seemed not to have an important effect on the performance of the A/ILP inference. At the point where interactions were strong enough to be detected, further increases of strength did not improve the overall detection accuracy.

Where the number of samples was large enough, InfIntE and our interaction hypotheses could detect simulated interactions with an accuracy comparable to SparCC and SPIEC-EASI that use correlation-like statistics. Both SparCC and SPIEC-EASI can detect interactions from 20 samples, suggesting that they may infer association networks at reduced sampling effort, but with the need for subjective expert interpretation. The InfIntE approach detects the simulated interactions with a probability similar to those that SparCC and SPIEC-EASI infer a possible association of type unknown. Moreover, the InfIntE approach does this with many fewer false positives than SparCC and a similar level of false positives to SPIEC-EASI. Direct classification by InfIntE is dependent upon the type of interaction simulated, however. Competition and amensal interactions, which can cause exclusion of OTUs, present difficulties of detection and classification. This result is consistent with the findings of Weiss et al. (2016), who found that the detection of simulated detrimental interactions for all interaction inference tools is difficult due to the loss of information that comes with exclusion. InfIntE detects detrimental interactions less well than SPIEC-EASI, and while incorporating exclusion in the hypothesis of interaction improves accuracy of detection, InfIntE has lower performance. We note, however, that those competition interactions that are detected by InfIntE are classified correctly with a high probability and a near zero rate of false

positives. This suggests that where competition interactions are detected and classified in a dataset, classification should be accurate even though some competition interactions may not be detected. In general, given this early stage in the development of A/ILP approaches for learning microbial interactions, the InfIntE approach detects interactions to an accuracy similar to SparCC and SPIEC-EASI, and does so with the benefit of direct, automatic classification.

Using InfIntE on real metabarcoding data for the fungal microbiome of grapevine leaves produced networks that varied between the nine vineyards of the dataset. Only one interaction was common to all nine vineyards. This result is consistent with the few consensus associations that have been found among vineyards in previous works using correlation-based tools like SparCC or SPIEC-EASI (Barroso-Bergadà et al., 2021). Each network showed a diversity of interaction types, including interactions that difficult to observe using classical ecological approaches. The most frequently found interaction types were classified as commensalism and competition interactions, likely based on energy transfer chains (Tshikantwa et al., 2018) or the exploitation of resources, such as space (Lloyd and Allen, 2015). Predation interactions were rarely classified, and this may be due to problems of detection with this type of interaction being masked by other interaction types (Derocles et al., 2018). The inclusion of qPCR data for the grapevine pathogen, *P. viticola*, allowed the prediction of subnetworks centred on the pathogen, with a view to understanding its ecology and the potential for management using microbial biocontrol (Koledenkova et al., 2022). A total of 14 fungal taxa were hypothesised to be antagonistic to *P. viticola*, with the predominant interactions inferred being of competition and more rarely amensalism. The list of potential antagonists included five taxa identified previously in the literature as *P. viticola* antagonists (Bakshi et al., 2001; Duhan et al., 2021; Ghule et al., 2018; Harm et al., 2011; Musetti et al., 2007, 2006). Six of the remaining taxa have also been proposed as biocontrol agents of other pathogens in different foliar systems. InfIntE proposed *Mycosphaerella tassiana* and two species from the Filobasidium genus as potential antagonists that are currently unknown. These literature validations give us some confidence that the InfIntE approach correctly classifies interactions, allowing us to unravel an approximation of the dark web of microbial interactions and the pathobiome.

Our InfIntE approach introduces the use of explainable machine learning to microbial interaction inference. Explainable machine learning

has had a limited use in ecology due, at least in part, to its relatively intense computational requirements and lower robustness to noise in comparison to statistical learning. The duration of an explainable approach increases exponentially with the size of the dataset, limiting interactivity and scale of the learning that can be done. Moreover, explainable machine learning approaches have been largely restricted to researchers in the field of logic with its distinct theoretical framework and suite of methodologies. InfIntE uses a new implementation of A/ILP called PyGol, which is available as a Python library (Varghese et al., 2022). Ecologists are more familiar with Python, facilitating the use of Pygol. PyGol is also much faster than previous A/ILP implementations and appears more robust to noise in datasets. This has reduced the run time for explainable machine learning, from several days to a few hours, for 60 samples and 80 OTUs. Most importantly, this promotes a much more interactive experimental approach to the inference, because the time costs of mistakes in coding are much lower. A/ILP allows other sources of information, such as from databases like Funguild or Faprotax, and functional ecological descriptions in place of taxonomy, to be included in the logical hypotheses. This would lead to greater descriptive precision, potentially improving further the inference process and our understanding of these Dark Webs.

Acknowledgment

We thank all members of the BCMicrobiome project for their support for this work providing the metabarcoding data, published in Barroso-Bergadà et al. (2023). The support of the NGB project consortium (ANR-17-CE32-0011) was instrumental in the development of this work. We also would like to thank Sophie Weiss for sharing the methods to create ecological-like computer generated datasets.

Ethics declarations

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the Agence Nationale de la Recherche, Grant Award Number: ANR-17-CE32-0011, and SYNGENTA CROP PROTECTION AG. Corinne Vacher and David A. Bohan acknowledge the support of the Learn-Biocontrol project, funded by the INRAE MEM metaprogramme, and the BCMicrobiome project funded by the Consortium Biocontrôle. Corinne Vacher acknowledges the support of the ANR (VITAE project, 20-PCPA-0010). Alireza Tamaddoni-Nezhad was supported by the EPSRC

Network Plus grant on Human-Like Computing (HLC). Dany Varghese was supported by Vice Chancellor's PhD Scholarship Award at the University of Surrey.

Data availability

The analysis script and the scripts to generate synthetic ASV tables have been deposited in Recherche Data Gouv (https://doi.org/10.57745/UDTORW). The bioinformatic process to obtain the fungal ASV table is described in Barroso-Bergadà et al. (2023). InfIntE R package is in a GitHub repository (https://github.com/didacb/InfIntE) and can be directly downloaded and installed in a R environment.

References

Ai, L., Muggleton, S.H., Hocquette, C., Gromowski, M., Schmid, U., 2021. Beneficial and harmful explanatory machine learning. Mach. Learn. 110, 695–721. https://doi.org/10.1007/s10994-020-05941-0.

Amanelah Baharvandi, H., Zafari, D., 2015. Identification of Cladosporium delicatulum as a mycoparasite of Taphrina pruni. Arch. Phytopathol. Plant Prot. 48, 688–697. https://doi.org/10.1080/03235408.2015.1099886.

Bakshi, S., Sztejnberg, A., Yarden, O., 2001. Isolation and characterization of a cold-tolerant strain of Fusarium proliferatum, a biocontrol agent of grape downy mildew. Phytopathology 91, 1062–1068. https://doi.org/10.1094/PHYTO.2001.91.11.1062.

Barroso-Bergadà, D., Pauvert, C., Vallance, J., Delière, L., Bohan, D.A., Buée, M., et al., 2021. Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: strengths and pitfalls. Mol. Ecol. Resour. 21, 762–780. https://doi.org/10.1111/1755-0998.13302.

Barroso-Bergada, D., Tamaddoni-Nezhad, A., Muggleton, S.H., Vacher, C., Galic, N., Bohan, D.A., 2022. Machine learning of microbial interactions using abductive ILP and hypothesis frequency/compression estimation. Springer. https://doi.org/10.1007/978-3-030-97454-1_3.

Barroso-Bergadà, D., Delmotte, F., Faivre d'Arcier, J., Massot, M., Chancerel, E., Demeaux, I., et al., 2023. Leaf microbiome data for European cultivated grapevine (Vitis vinifera) during Downy Mildew (Plasmopara viticola) epidemics in three wine-producing regions in France. PhytoFrontiersTM 3. https://doi.org/10.1094/phytofr-11-22-0138-a.

Becker, R., Ulrich, K., Behrendt, U., Kube, M., Ulrich, A., 2020. Analyzing ash leaf-colonizing fungal communities for their biological control of Hymenoscyphus fraxineus. Front. Microbiol. 11, 590944. https://doi.org/10.3389/fmicb.2020.590944.

Berry, D., Widder, S., 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front. Microbiol. 5, 219. https://doi.org/10.3389/fmicb.2014.00219.

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. https://doi.org/10.1371/journal.pone.0029028.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869.

Castelvecchi, D., 2016. Can we open the black box of AI? Nature 538, 20–23. https://doi.org/10.1038/538020a.

Chiquet, J., Mariadassou, M., Robin, S., 2019. Variational inference of sparse network from count data, in: 36th Int. Conf. Mach. Learning, ICML 2019, pp. 1988–1997.

Crhanova, M., Karasova, D., Juricova, H., Matiasovicova, J., Jahodarova, E., Kubasova, T., et al., 2019. Systematic culturomics shows that half of chicken caecal microbiota members can be grown in vitro except for two lineages of Clostridiales and a single lineage of Bacteroidetes. Microorganisms 7, 496. https://doi.org/10.3390/microorganisms7110496.

Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., Callahan, B.J., 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6, 1–14. https://doi.org/10.1186/s40168-018-0605-2.

de Vries, F.T., Griffiths, R.I., Bailey, M., Craig, H., Girlanda, M., Gweon, H.S., et al., 2018. Soil bacterial networks are less stable under drought than fungal networks. Nat. Commun. 9, 1–12. https://doi.org/10.1038/s41467-018-05516-7.

Derocles, S.A.P., Bohan, D.A., Dumbrell, A.J., Kitson, J.J.N., Massol, F., Pauvert, C., et al., 2018. Biomonitoring for the 21st century: integrating next-generation sequencing into ecological network analysis. Advances in Ecological ResearchAcademic Press Inc., pp. 1–62. ⟨https://doi.org/10.1016/bs.aecr.2017.12.001⟩.

Dohlman, A.B., Shen, X., 2019. Mapping the microbial interactome: statistical and experimental approaches for microbiome network inference. Exp. Biol. Med. 244, 445–458. https://doi.org/10.1177/1535370219836771.

Duhan, D., Gajbhiye, S., Jaswal, R., Singh, R.P., Sharma, T.R., Rajarammohan, S., 2021. Functional characterization of the Nep1-like protein effectors of the necrotrophic pathogen – Alternaria brassicae. Front. Microbiol. 12. https://doi.org/10.3389/fmicb.2021.738617.

Fan, J., Upadhye, S., Worster, A., 2006. Understanding receiver operating characteristic (ROC) curves. Can. J. Emerg. Med. 8, 19–20. https://doi.org/10.1017/S1481803500013336.

Fang, H., Huang, C., Zhao, H., Deng, M., 2015. CCLasso: correlation inference for compositional data through Lasso. Bioinformatics 31, 3172–3180. https://doi.org/10.1093/bioinformatics/btv349.

Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. Nat. Rev. Microbiol. 10, 538–550. https://doi.org/10.1038/nrmicro2832.

Filonow, A.B., Vishniac, H.S., Anderson, J.A., Janisiewicz, W.J., 1996. Biological control of Botrytis cinerea in apple by yeasts from various habitats and their putative mechanisms of antagonism. Biol. Control 7, 212–220. https://doi.org/10.1006/bcon.1996.0086.

Friedman, J., Alm, E.J., 2012. Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 8, 1–11. https://doi.org/10.1371/journal.pcbi.1002687.

Ghule, M.R., Sawant, I.S., Sawant, S.D., Sharma, R., Shouche, Y.S., 2018. Identification of Fusarium species as putative mycoparasites of Plasmopara viticola causing downy mildew in grapevines. Australas. Plant Dis. Notes 13, 16. https://doi.org/10.1007/s13314-018-0297-2.

Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L., 2019. Explaining explanations: an overview of interpretability of machine learning, in: Proc. - 2018 IEEE 5th Int. Conf. Data Sci. Adv. Analytics 2018 (DSAA). https://doi.org/10.1109/DSAA.2018.00018.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome datasets are compositional: and this is not optional. Front. Microbiol. 8, 1–6. https://doi.org/10.3389/fmicb.2017.02224.

Gramisci, B.R., Lutz, M.C., Lopes, C.A., Sangorrín, M.P., 2018. Enhancing the efficacy of yeast biocontrol agents against postharvest pathogens through nutrient profiling and the use of other additives. Biol. Control 121, 151–158. https://doi.org/10.1016/j.biocontrol.2018.03.001.

Hacquard, S., Spaepen, S., Garrido-Oter, R., Schulze-Lefert, P., 2017. Interplay between innate immunity and the plant microbiota. Annu. Rev. Phytopathol. 55, 565–589. https://doi.org/10.1146/annurev-phyto-080516-035623.

Harm, A., Kassemeyer, H.H., Seibicke, T., Regner, F., 2011. Evaluation of chemical and natural resistance inducers against downy mildew (Plasmopara viticola) in grapevine. Am. J. Enol. Vitic. 62, 184–192. https://doi.org/10.5344/ajev.2011.09054.

He, Y., Caporaso, J.G., Jiang, X.-T., Sheng, H.-F., Huse, S.M., Rideout, J.R., et al., 2015. Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. Microbiome 3, 20. https://doi.org/10.1186/s40168-015-0081-x.

Hirano, H., Takemoto, K., 2019. Difficulty in inferring microbial community structure based on co-occurrence network approaches. BMC Bioinform. 20, 329. https://doi.org/10.1186/s12859-019-2915-1.

Janisiewicz, W.J., Bors, B., 1995. Development of a microbial community of bacterial and yeast antagonists to control wound-invading postharvest pathogens of fruits. Appl. Environ. Microbiol. 61, 3261–3267. https://doi.org/10.1128/aem.61.9.3261-3267.1995.

Köhl, J., de Geijn, H.G., van Haas, L.G., de Henken, B., Hauschild, R., Hilscher, U., et al., 2019. Stepwise screening of candidate antagonists for biological control of Blumeria graminis f. sp. tritici. Biol. Control 136, 104008. https://doi.org/10.1016/j.biocontrol.2019.104008.

Koledenkova, K., Esmaeel, Q., Jacquard, C., Nowak, J., Clément, C., Ait Barka, E., 2022. Plasmopara viticola the causal agent of Downy mildew of grapevine: from its taxonomy to disease management. Front. Microbiol. 13, 889472. https://doi.org/10.3389/fmicb.2022.889472.

Konopka, A., 2009. What is microbial community ecology. ISME J. 3, 1223–1230. https://doi.org/10.1038/ismej.2009.88.

Kurtz, Z.D., Müller, 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, 1–25. https://doi.org/10.1371/journal.pcbi.1004226.

Ishaq, S.L., 2017. Plant-microbial interactions in agriculture and the use of farming systems to improve diversity and productivity. AIMS Microbiol. 3, 335–353. https://doi.org/10.3934/microbiol.2017.2.335.

Li, C., Zhang, N., Li, B., Xu, Q., Song, J., Wei, N., et al., 2017. Increased torulene accumulation in red yeast Sporidiobolus pararoseus NGR as stress response to high salt conditions. Food Chem. 237, 1041–1047. https://doi.org/10.1016/j.foodchem.2017.06.033.

Liu, H., Roeder, K., Wasserman, L., 2010. Stability approach to regularization selection (StARS) for high dimensional graphical models, in: Adv. Neural Inf. Process. Syst. 23: 24th Annu. Conf. Neural Inf. Process. Syst. 2010, NIPS 2010.

Lloyd, D.P., Allen, R.J., 2015. Competition for space during bacterial colonization of a surface. J. R. Soc. Interface 12, 0608. https://doi.org/10.1098/rsif.2015.0608.

Lutz, M.C., Lopes, C.A., Rodriguez, M.E., Sosa, M.C., Sangorrín, M.P., 2013. Efficacy and putative mode of action of native and commercial antagonistic yeasts against postharvest pathogens of pear. Int. J. Food Microbiol. 164, 166–172. https://doi.org/10.1016/j.ijfoodmicro.2013.04.005.

Marcy, Y., Ouverney, C., Bik, E.M., Lösekann, T., Ivanova, N., Martin, H.G., et al., 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc. Natl Acad. Sci. U. S. A. 104, 11889–11894. https://doi.org/10.1073/pnas.0704662104.

Muggleton, S.H., Bryant, C.H., 2000. Theory completion using inverse entailment. Lecture Notes in Computer Science. Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformaticshttps://doi.org/10.1007/3-540-44960-4_8.

Müller, C.L., Bonneau, R., Kurtz, Z., 2016. Generalized Stability Approach for Regularized Graphical Models.

Musetti, R., Vecchione, A., Stringher, L., Borselli, S., Zulini, L., Marzani, C., et al., 2006. Inhibition of sporulation and ultrastructural alterations of grapevine downy mildew by the endophytic fungus Alternaria alternata. Phytopathology 96, 689–698. https://doi.org/10.1094/PHYTO-96-0689.

Musetti, R., Polizzotto, R., Vecchione, A., Borselli, S., Zulini, L., D'Ambrosio, M., et al., 2007. Antifungal activity of diketopiperazines extracted from Alternaria alternata against Plasmopara viticola: an ultrastructural study. Micron 38, 643–650. https://doi.org/10.1016/j.micron.2006.09.001.

Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., et al., 2019. Bioinformatics matters: the accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. Fungal Ecol. 41, 23–33. https://doi.org/10.1016/j.funeco.2019.03.005.

Poveda, J., Roeschlin, R.A., Marano, M.R., Favaro, M.A., 2021. Microorganisms as biocontrol agents against bacterial citrus diseases. Biol. Control. 158, 104602. https://doi.org/10.1016/j.biocontrol.2021.104602.

R. Core Team, 2018. R: a language and environment for statistical computing.

Core Team, R., 2022. R Core Team 2021 R: A Language and Environment for Statistical ComputingR Foundation for Statistical Computing 2. ⟨https://www.R-project.org/⟩.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., et al., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinforma. 12. https://doi.org/10.1186/1471-2105-12-77.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 2016, e2584. https://doi.org/10.7717/peerj.2584.

Röttjers, L., Faust, K., 2018. From hairballs to hypotheses–biological insights from microbial networks. FEMS Microbiol. Rev. 42, 761–780. https://doi.org/10.1093/femsre/fuy030.

Ruppert, K.M., Kline, R.J., Rahman, M.S., 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. Glob. Ecol. Conserv. 17, e00547. https://doi.org/10.1016/j.gecco.2019.e00547.

Shoemaker, W.R., Locey, K.J., Lennon, J.T., 2017. A macroecological theory of microbial biodiversity. Nat. Ecol. Evol. 1, 107. https://doi.org/10.1038/s41559-017-0107.

Tackmann, J., Matias Rodrigues, J.F., von Mering, C., 2019. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. Cell Syst. 9, 286–296.e8. https://doi.org/10.1016/j.cels.2019.08.002.

Tamaddoni-Nezhad, A., Bohan, D.A., Milani, G.A., Raybould, A., Muggleton, S., 2021. Human-Machine Scientific Discovery.In Human-Like Machine Intelligence. Oxford University Press, pp. 279–315.

Tamaddoni-Nezhad, A., Chaleil, R., Kakas, A., Muggleton, S., 2006. Application of abductive ILP to learning metabolic network inhibition from temporal data. Machine Learning 64, 209–230. https://doi.org/10.1007/s10994-006-8988-x.

Thomsen, P.F., Willerslev, E., 2015. Environmental DNA - an emerging tool in conservation for monitoring past and present biodiversity. Biol. Conserv. 183, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019.

Tshikantwa, T.S., Ullah, M.W., He, F., Yang, G., 2018. Current trends and potential applications of microbial interactions for human welfare. Front. Microbiol. 9, 1156. https://doi.org/10.3389/fmicb.2018.01156.

Ushey, K., Allaire, J., Tang, Y., 2022. reticulate: interface to 'Python'. ⟨https://CRAN.R-project.org/package=reticulate⟩.

Varghese, D., Barroso-Bergadà, D., Bohan, D.A., Tamaddoni-Nezhad, A., 2022. Efficient abductive learning of microbial interactions using Meta Inverse Entailment. Springer.

Watts, S.C., Ritchie, S.C., Inouye, M., Holt, K.E., 2019. FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics 35, 1064–1066. https://doi.org/10.1093/bioinformatics/bty734.

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J. 10, 1669–1681. https://doi.org/10.1038/ismej.2015.235.

White, N.A., Boddy, L., 1992. Extracellular enzyme localization during interspecific fungal interactions. FEMS Microbiol. Lett. 98. https://doi.org/10.1111/j.1574-6968.1992.tb05493.x.

Wilke, C.O., 2019. cowplot: streamlined plot theme plot annotations "ggplot2."

Wilkinson, L., 2011. ggplot2: elegant graphics for data analysis by WICKHAM, H. Biometrics 67, 678–679. https://doi.org/10.1111/j.1541-0420.2011.01616.x.

Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X., Xiang, M., 2019. Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology 127–140. https://doi.org/10.1080/21501203.2019.1614106.