



HAL
open science

Uncovering microbial food webs using machine learning

Janna Barel, Owen Petchey, Abir Ghaffouli, Vincent E.J. Jassey

► **To cite this version:**

Janna Barel, Owen Petchey, Abir Ghaffouli, Vincent E.J. Jassey. Uncovering microbial food webs using machine learning. *Soil Biology and Biochemistry*, 2023, 186, pp.109174. 10.1016/j.soilbio.2023.109174 . hal-04234479

HAL Id: hal-04234479

<https://hal.science/hal-04234479>

Submitted on 10 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Uncovering microbial food webs using machine learning**

2

3 Janna M. Barel^{1,2} (OrCID: 0000-0001-8520-8289), Owen L. Petchey³ (OrCID: 0000-0002-7724-1633),

4 Abir Ghaffouli¹, Vincent E.J. Jassey^{1*} (OrCID: 0000-0002-1450-2437)

5

6 ¹ Laboratoire Ecology Fonctionnelle et Environnement, Université Paul Sabatier III, CNRS, F-31062

7 Toulouse, France

8 ² Aquatic Ecology & Environmental Biology, Radboud Institute for Biological and Environmental

9 Sciences, Faculty of Science, Radboud University Nijmegen, AJ 6525 Nijmegen, The Netherlands

10 ³ Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich,

11 Switzerland

12

13 * corresponding author: vincent.jassey@univ-tlse3.fr

14

15 **Running title** (45 characters): Untangling microbial food webs

16

17 **Authorship:** JMB and VEJJ contributed equally to the project. VEJJ had the original idea. VEJJ and

18 JMB conceived the study in close consultation with OLP. JMB assembled the database of microbial

19 feeding links with the help of VEJJ. VEJJ provided the case study data. VEJJ and JMB developed the

20 modelling framework with the help of AG and OLP. VEJJ performed the statistical analyses with the

21 help of JMB. VEJJ wrote the manuscript in close consultation with JMB and OLP. All authors

22 contributed to the final form of the manuscript.

23 **Abstract**

24 Microbial trophic interactions are an important aspect of microbiomes in any ecosystem. They can
25 reveal how microbial diversity modulates ecosystem functioning. However, uncovering microbial
26 feeding interactions is a challenge because direct observation of predation is difficult with classical
27 approaches such as behaviour and gut contents analyses. To overcome this issue, recent developments
28 in trait-matching and machine-learning approaches are promising for successfully inferring microbial
29 feeding links. Here, we tested the ability of six machine-learning algorithms for predicting microbial
30 feeding links, based on species traits and taxonomy. By incorporating organism speed, size and
31 abundance into the model predictions, we further estimated the probability of feeding links occurring.
32 We found that the boost-regression trees predicted feeding links between microbes best. Sensitivity
33 analyses showed that feeding predictions were robust against false feeding links and faulty predictors
34 in the training set, and capable of predicting feeding links for empirical datasets containing up to 50%
35 of new taxa. We cross-validated our predictions using an empirical dataset from a *Sphagnum*-
36 dominated peatland with direct feeding observations for two dominant testate amoeba predators. The
37 feeding habits of the two testate amoeba species were comparable between microscopic observations
38 and model predictions. Machine learning thus offers a means to develop robust models for studying
39 microbial food webs in ecosystems. They offer a route to combine traditional observations with DNA-
40 based sampling strategies to upscale soil biodiversity research along ecological gradients.

41

42 **Keywords:** microbiomes, feeding habits, predation, food web modelling, micro-eukaryotes, feeding
43 link strength, peatlands

44 **1. Introduction**

45 Microorganisms are the biogeochemical engineers of life on Earth (Falkowski et al., 2008). Their activity
46 cycles major elements, such as H, C, N, O, S, and P, supporting life in higher trophic levels. In particular,
47 microorganisms govern terrestrial carbon cycling through their role in greenhouse gas emissions (Singh
48 et al., 2010), carbon cycling (Schimel and Schaeffer, 2012; Liang et al., 2017), plant growth (Guo et al.,
49 2022), and primary productivity (Schimel and Schaeffer, 2012; Jassey et al., 2022). In light of current
50 and future climate predictions, understanding and predicting the response of microorganisms— from
51 bacteria to protists— and the ecosystem services they provide to environmental change has never been
52 more important (Singh et al., 2010; Jansson and Hofmockel, 2020; Naylor et al., 2020; Guerra et al.,
53 2022).

54 High-throughput sequencing and global analyses have been fundamental in revealing the
55 broad diversity of microorganisms across terrestrial systems (Tedersoo et al., 2014; de Vargas et al.,
56 2015; Delgado Baquerizo et al., 2018; Oliverio et al., 2020; Singer et al., 2021; Xiong et al., 2021). In
57 addition to the essential bacterial and fungal diversity, these studies particularly highlighted an
58 unexpected diversity of predators (Oliverio et al., 2020; Petters et al., 2021; Xiong et al., 2021),
59 suggesting the importance of predation in shaping microbial diversity in soils. Predation is indeed
60 increasingly recognized as a dominant driver of bacterial and fungal community structures in soils (Gao
61 et al., 2018; Amacker et al., 2022; Pierce and Dutton, 2022), and hence, of key ecosystem processes
62 such as nutrient and carbon cycling and plant health (Séneca et al., 2021; Amacker et al., 2022).
63 Uncovering microbial food webs to reveal the feeding interactions among organisms and characterizing
64 their specific contribution to ecosystem processes across space and time is thus a clear next step to
65 advance our ability to construct accurate predictive models for global carbon flux under climate change
66 (Geisen, 2021). However, our knowledge of microbial feeding interactions is mostly based on broad
67 functional groups and/or a few model species (Geisen et al., 2016; Mestre et al., 2022). With the
68 growing awareness that even closely related microbial species may differ in their feeding habits (Jassey
69 et al., 2012a, 2013b) comes the question of how to identify and quantify the feeding interactions of the
70 hundreds of species coexisting in a single gram of soil (Pierce and Dutton, 2022). Indeed, microbial
71 consumers exhibit a large variety of morphologies, feeding habits, and life strategies (e.g. mixotrophy)
72 that can influence their feeding interactions with other organisms (Potapov et al., 2022), and their

73 impact on the food web structure and ecosystem functions (Jassey et al., 2013b, 2015; Mieczan et al.,
74 2015).

75 Understanding how microbes interact with each other within a system is challenging. Recently,
76 co-occurrence networks have been increasingly used to explore the spatial or temporal patterns of co-
77 occurrence between pairs of taxa, and relate shifts in their properties to ecosystem changes (Faust and
78 Raes, 2012; de Vries et al., 2018). Although informative and very useful to extract simple patterns from
79 complex datasets, these networks are often misused (Goberna and Verdú, 2022) and cannot in any
80 circumstance reflect trophic interactions between organisms. Because direct observation of microbial
81 feeding interactions is extremely difficult, researchers had to use alternative solutions. For instance,
82 they used correlation-based analyses of microbiome sequencing datasets to infer feeding interactions
83 (Seppey et al., 2017). Other studies used either feeding trials (Amacker et al., 2022), stable isotopic
84 tracing in the microbial biomass (Jassey et al., 2013b; Mieczan et al., 2015) and in the DNA (Morriën et
85 al., 2017), metatranscriptomics (Petters et al., 2021), or gut content analysis (Lo et al., 2022). The latter
86 is extremely challenging for small microbes. While these analyses have been powerful hypothesis-
87 generators, they also have a number of limitations, including biases toward the collection of enough
88 individuals to obtain reliable measurements (stable isotopes, gut content), detection of strong
89 symmetric interactions (correlation-based inferences), and/or inability to distinguish species-specific
90 interactions (stable isotopes). Moreover, with large numbers of species and/or food webs to resolve
91 within studies, these approaches become cumbersome or unfeasible to apply. To overcome these
92 limitations, researchers have pooled microorganisms into broad feeding groups to construct food
93 webs. Although this is an established approach to quantifying energy flows (Holtkamp et al., 2011;
94 Koltz et al., 2018), lumping species together can create large errors particularly when species with
95 divergent behaviours are combined (Buchkowski and Lindo, 2020).

96 As an alternative, several statistical methods based on machine learning could be used to infer
97 microbial feeding interactions (Faisal et al., 2010; Desjardins-Proulx et al., 2017; Pichler et al., 2020).
98 Previous work on fishes and invertebrates showed that predator-prey interactions can be successfully
99 inferred using either phylogenetic relationships (Gray et al., 2015) or the trait-matching approach
100 (Gravel et al., 2013; Laigle et al., 2018a; Pomeranz et al., 2019). Inferences based on foraging traits
101 such as body size have been shown to be successful in predicting feeding links (Gravel et al., 2013;
102 Laigle et al., 2018b; Pecuchet et al., 2020), although machine learning models trained with both traits

103 and phylogenetic predictors should be preferred (Pomeranz et al., 2019). Whilst traits can serve as
104 reliable proxies for the inference of species interactions, phylogeny and/or taxonomy can substitute
105 unmeasured traits (Morales-Castilla et al., 2015), and even recommend links that have not been
106 observed empirically before (Desjardins-Proulx et al., 2017).

107 Here, we propose to use a composite machine learning model trained with both phylogenetic
108 predictors and traits to provide a potential framework for predicting microbial feeding interactions,
109 allowing for a greater understanding of the mechanisms underlying trophic interactions, community
110 structure, and ecosystem functioning. As we have no *a priori* trait-matching and phylogenetic rules, we
111 used a systematic assessment of the performance of different machine learning (ML) techniques to infer
112 microbial feeding interactions. To inspect to what extent the resulting species pairs could have arisen
113 from chance, we further cross-validated the predicted feeding interactions from the best-performing
114 ML model using an array of sensitivity analyses. Then, we applied the best ML model to twelve
115 previously published species lists from peatland (Jassey et al., 2012a; Reczuga et al., 2018) to
116 reconstruct microbial food webs from these empirical data. Following species feeding pairs inferences,
117 we further cross-validated the predicted interactions with direct feeding observations from the case
118 study data (Jassey et al., 2012a). Finally, we developed a new method to estimate the strength of the
119 predicted links according to the abundance and motility of microbial species, going beyond a simple
120 binary network prediction. This newly developed framework permits the analysis of microbial food webs
121 in soils and other ecosystems, which will allow researchers to incorporate microbial food web analyses
122 in their studies and better understand the relationships between microbial food web properties and
123 ecosystem functions.

124

125 **2. Material and methods**

126 **2.1 Training datasets preparation and description**

127 To train our machine learning algorithms, we built an extensive meta food web documenting feeding
128 links between 164 microbial taxa that potentially inhabit peatlands, ranging from bacteria to metazoans
129 (Data available at <https://doi.org/10.6084/m9.figshare.23669019.v1>). Most of the species included in
130 the registry are identified at the species level, with the exception of some biological entities that are
131 described at the functional group level (e.g. bacteria and fungi). The feeding link registry was first

132 documented from peer-reviewed publications, specialized literature and personal communication with
133 experts, and spanned the period 1902 to 2021. In total, 733 trophic links were extracted from tables,
134 figures, the main text, and/or supplementary materials from 40 sources (see Table S1). Of these 40
135 sources, 19 covered soils, 15 freshwater systems and 6 peatlands. Next, we added a list of microbial
136 taxa regularly—but not exclusively—found in peatlands by merging the species lists from our previous
137 work (Jassey et al., 2011, 2015; Reczuga et al., 2018; Sytiuk et al., 2021). Since not all trophic
138 interactions were readily available for these peatland taxa, we retrieved the missing links from the
139 published generic links using the WebBuilder function in *R* (Gray et al., 2015). The outputs were
140 thoroughly checked and, where needed, corrected based on literature, personal observation, and
141 expert opinion (see Table S1 for details). The resulting meta food web comprised 164 taxa, including
142 cyanobacteria (10), microalgae (35), flagellates (12), ciliates (26), testate amoebae (59), rotifers (20), and
143 nematodes (1), and 27,060 rows documenting 3,590 positive links (feeding) and 23,470 negative links
144 (non-feeding). Each resource-consumer pair was further associated with information on 17 potential
145 feeding link predictors for both the consumer and resource (34 predictors in total for each resource-
146 consumer pair; Pecuchet et al., 2020). These predictors included six taxonomic variables (Species,
147 Genus, Family, Order, Phylum, and Kingdom) and eleven traits (Organism type, Nourishment, Feeding
148 group, Morphology, Lifestyle, Locomotion, Movement, Feeding strategy, Body length, Body width,
149 and Biovolume; see Table S2 for details).

150 151 **2.2 Predicting microbial feeding links using machine learning algorithms**

152 We compared six machine learning algorithms (ML) for predicting microbial feeding links based on
153 taxonomic and trait information (Fig. 1; Table S3): Random Forest (RF), Boosted regression trees (BRT),
154 k-nearest-neighbor (kNN), neural networks (NN), Generalized linear model (GLM), and Bayesian GLM
155 (BGLM). These algorithms were selected because they (1) are suitable for classification tasks, (2) are
156 known to perform well in predicting species interactions, and (3) represent various classes of ML
157 algorithms (Table S3). We partitioned our *a priori* meta food web into a Training set (70% of the data,
158 used to train the models) and a Test set (30% of the data, used to test model performance) in which
159 the ratios between feeding and non-feeding links was conserved. We iteratively varied the set of
160 predictors across ten runs for each ML algorithm and evaluated model strength using sensitivity
161 analyses (see below). This approach allowed us to select the best-performing model that had high

162 predictive strength, and optimize predictor selection while taking into account issues related to
163 multicollinearity. For each run, we resampled the predictors from 12 clusters of strongly covarying
164 predictors. Quantitative and qualitative predictors were hierarchically clustered using the *ClustOfVar*
165 package in *R* (Chavent et al., 2012). As we did not have a specific number of clusters defined *a priori*
166 to use as parameters for the clustering procedure, we tested a range of cluster numbers (that is 5 to
167 20) into the *ClustOfVar* function (van den Hoogen et al., 2019). Inspection of the cluster dendrograms
168 revealed that the predictor variables were most homogeneous grouped in 12 clusters (Fig. S1, Table
169 S4). In summary, each run of the six ML algorithm was trained with twelve predictors randomly selected
170 across the 36 predictors grouped into twelve clusters (Table S4). The relevance of the predictors for
171 each run was assessed using the *varImp* function in the *caret* *R* package (scaling 0-100) (Kuhn, 2008).
172 Since *varImp* is model dependent and our approach includes a randomized set of predictors for each
173 run, we ranked predictor importance for each run and each trained model from 1 (most important) to
174 12 (less important) to compare model outputs.

175 To assess and compare the ML model's predictive performance, we calculated three
176 performance indices using the Test set. The area under the receiver operating characteristic curve
177 (AUC, range 0 to 1) was used as an indication of the model's ability to distinguish between feeding and
178 non-feeding links regardless of the classification method (AUC ≥ 0.5 has a $\geq 50\%$ chance of being
179 correct) (Manel et al., 2001). The proportion of correct predictions was used as a measure of Accuracy,
180 while the true skill statistic (TSS, range -1 to +1) normalizes the overall accuracy of the model for the
181 accuracy occurring by chance (Allouche et al., 2006). A TSS of -1 is expected if predictions are random,
182 1 if predictions are perfect, and 0 if they are the opposite of observations. We considered the
183 performance of an algorithm to be good when AUC, Accuracy, and TSS values were close to one.
184 Related indices, including the receiver operating characteristic (ROC), sensitivity (rate of true positives
185 among all positive predictions), and specificity (rate of true negatives among all negative predictions),
186 were also calculated and used for selecting the best ML algorithm. Machine learning analyses were
187 performed in *R* (R Core Team, 2020) using the *caret* suite packages (Kuhn, 2008), and custom *R* codes.

188
189
190
191

192 2.3 Sensitivity analyses of microbial feeding link predictions

193 Following ML model selection, we subjected the 'best-performing model among the ten runs of the
194 best-performing ML algorithm to further sensitivity analyses in order to inspect the quality of the
195 predictions and their sensitivity to false predictors and poor species taxonomy (Fig. 1) (Jassey et al.,
196 2022). In other words, the 'best-performing model' was our best identified ML algorithm trained with
197 the twelve best predictors identified from the ten runs with randomized sets of predictors of that
198 particular ML algorithm.

199 The first sensitivity analysis (i.e. k -fold 'Predictor Variable Shuffling (PVS) SA') tested the
200 influence of poorly identified predictors on feeding link predictions, a plausible situation especially
201 when working on microbial traits (Escalas et al., 2019). To do so, we iteratively and randomly shuffled
202 1, 3, and 6 predictors to break any structure in the relationship between feeding links and some
203 taxonomic and/or trait predictors before training our best-performing ML model. PVS SA was run 10
204 times ($k = 10$) for each variable shuffling condition (a total of 30 runs) to cover as many as possible
205 combinations of predictor variables shuffling.

206 The second sensitivity analysis covered the influence of species that are new or poorly
207 characterized at the taxonomic level (i.e. k -fold 'Poorly Taxonomic' (PT) SA). For instance, such an effect
208 is likely to happen if the empirical dataset for which to predict feeding links contains species that are
209 'new' and not represented in the Training set used to train the ML model. To simulate this potential
210 bias, we randomly selected 1%, 2.5%, and 5% of the Training set in which 75% of the taxa (resource or
211 consumer) at the species level were set as 'Unknown' before training the best-performing ML model.
212 Among this 75% of species, 50% of them were further set as 'Unknown' at the genus level, and 25% at
213 the family level. Each condition (1%, 2.5% and 5% of the Training set) was run 10 times ($k = 10$; in total
214 30 runs). One run with the conditions 10%, 25% and 50% was further computed to identify at which
215 percentage our best ML model is no longer trustworthy.

216 For each sensitivity analysis, we assessed the predictive performance of the models for each
217 run using the Test set and generated sensitivity statistics, namely Accuracy, AUC, and TSS. All sensitivity
218 analyses were performed in *R* using the *caret* suite packages and custom *R* codes.

219
220
221

222 2.4 Case study — Inferring microbial feeding links in a *Sphagnum*-dominated peatland

223 As a case study for inferring microbial feeding links in ecosystems, we used a well-studied empirical
224 dataset from a *Sphagnum*-dominated peatland situated in the Jura Mountains, France (The Forbonnet
225 peatland, 46°49'35" N, 6°10'20" E; Jassey *et al.* 2011a, 2012b, 2013a). The dataset comprises twelve
226 microbial species lists, including microbial abundances, from twelve independent plots spread across
227 the peatland (more details can be found in Jassey *et al.* 2011). Because of the wet conditions, peat
228 mosses support diverse microbial communities (Gilbert and Mitchell, 2006) including bacteria,
229 cyanobacteria, fungi, protists and metazoan covering diverse feeding behaviours (saprotrophy,
230 autotrophy, mixotrophy and heterotrophy). In summary, the case study dataset contained a total of 108
231 species: bacteria, (1 group level), fungi (2 groups level), microalgae (18 taxa), cyanobacteria (10 taxa),
232 flagellates (9 taxa), ciliates (24 taxa), testate amoebae (28 taxa), rotifers (11 taxa), and nematode (1
233 taxon). Of these 108 species, 15 were not listed in the Training set used to train the ML models. Among
234 these 15 taxa, 50% were 'new' at the genus level when referring to the Training set, and 25% of them
235 were 'new' at the family level. These new taxa were classified as 'Unknown' to allow the predictions
236 with the selected ML model, while such bias in predictions has been tested beforehand in sensitivity
237 analyses (see above). The same traits were available in the case study data as in the Training set used
238 to train the ML models. More details about the methods used to identify microbial species and quantify
239 their abundance and morphological traits can be found in the Supplementary method.

240 We used the best-performing ML model to predict the binary food web from each species list
241 (i.e. presence/absence of feeding links between species and/or generic group, hereafter 'binary food
242 web'). Next, we accounted for population dynamics that could lead to significant deviations from the
243 binary food web. For example, the abundance of a consumer may greatly exceed that of its potential
244 resource, making the occurrence of that feeding link unlikely. We took such potential bias into account
245 using an abductive rule and abundance data of each taxon. Every predicted feeding link where the
246 abundance of the predator was higher than the abundance of the prey was considered unlikely and
247 removed. The pruned binary food webs were further quantified, with links strengths expressed as the
248 probability of a consumer capturing a resource based on its swimming speed and size, and the
249 abundance of the prey (hereafter 'link-strength food web'). All details about link strength calculations
250 are given in the supplementary method and Fig. S6 and Table S6.

251

252 2.5 Validating feeding habit predictions from the case study data

253 We cross-validated the feeding link predictions from ML in the case study data using the feeding habit
254 data from two testate amoebae species— *Hyalosphenia papilio* and *Nebela tincta*. We used these two
255 species in particular because we had detailed information of their feeding habit in the same plots at
256 the same sampling date (see Jassey *et al.* 2012 for details). These two focal species were chosen
257 because they are dominant microbial predators in peatlands and their high abundance allowed robust
258 and reliable feeding habit quantification. 1,240 individuals have been observed to establish feeding
259 habits (see details Jassey *et al.*, 2012), thus giving reliable feeding habits.

260 To validate the feeding predictions for these two species with their respective feeding habits,
261 we used two approaches. First, we calculated sensitivity and specificity indices; two metrics based on
262 a confusion matrix which evaluate model performance to properly assign trophic links. Sensitivity
263 measures the capacity of the model to predict observed feeding links well (trues positives), while
264 specificity measures the capacity of the model to predict non-observed feeding links (false positives).
265 Second, we reconstructed the diet of the two focal testate amoebae species quantitatively using the
266 predicted feeding links. To do so, we made the comparison based on the predicted link-strength food
267 web as our binary food webs cannot count the number of times a predator feeds on a prey item. For
268 each focal species and in each food web, we extracted the feeding links predicted by the ML model
269 and calculated the proportion of link strength by main prey type (i.e. microalgae, cyanobacteria,
270 ciliates, ...). For the observed feeding links, we calculated link strength, as explained above, as well as
271 the proportion of link strength by main prey type. Finally, we compared the predicted link strength
272 proportion per prey type to those observed. We removed the feeding link towards bacteria beforehand
273 as they could not be directly observed under the microscope.

274

275 2.6 Food web metrics

276 To assess the structure of the predicted food webs and quantify the benefits of link strength, we
277 measured seven unweighted, topology-based, food web metrics to characterize the structure of the
278 food webs, including the number of species (S), connectance (C), generality (G), vulnerability (V),
279 shortest path length (SPL), short-weighted trophic level (TL), the degree of omnivory (OI), averaged
280 food chain length (FCL), and the proportion of basal (Ba), intermediate (Int) and top species (Pred), as
281 well as vulnerability SD and generality SD (Kortsch *et al.*, 2019, 2021). The selected food web metrics

282 were chosen because they cover diverse and ecologically relevant aspects of the food web structure
283 (Ledger et al., 2013; Kortsch et al., 2015, 2021) (Table S7). In particular, they relate to the vertical
284 (trophic level, omnivory) and the horizontal (generality, vulnerability) dimensions of food webs and to
285 their complexity (species richness, connectance). Although these unweighted metrics can unveil the
286 topological structure of the food webs, they are also limited by the fact that in real food webs the links
287 may have different capacities or intensities or flows of information or strengths (Barrat et al., 2004).
288 Therefore, we calculated weighted metrics using link strength, namely the average strength of links (st),
289 the average strength distribution among links (ps), the weighted shortest path length (dist_w), the
290 transitivity (trans), the average strength of each node's connection (neigh), the modularity (mod), the
291 sum of link weights (flow), and the global efficiency (eff) (Table S7).

292

293 **2.7 Statistical analyses**

294 We tested the differences in algorithm performance and sensitivity analyses by running generalized
295 linear models (GLM) per indicator with the algorithm as the explanatory variable, using the *nlme*
296 package (Pinheiro et al., 2016). VarIdent variance structure was applied to account for
297 heteroscedasticity (Zuur et al., 2010). Assumptions on normality and homogeneity of variances were
298 verified using diagnostic plots. Pairwise differences between the algorithms were tested with Tukey
299 post-hoc test. We performed a multivariate principal component analysis (PCA) on all food web metrics
300 to investigate the main differences in food web structure and function between binary and link-strength
301 food webs, and to assess whether unweighted and weighted food web approaches (calculated for the
302 link-strength food webs) highlight different (or similar) aspects of the microbial food web in peatlands.
303 *vegan* R package was used to perform the PCAs (Oksanen, 2011). Analysis of similarities (ANOSIM) was
304 used to test statistically whether there are significant differences between binary and link-strength food
305 web topologies, as well as weighted and unweighted link-strength food web topologies. We further
306 calculated the network dissimilarity between PA and LS food webs using the *betalink* R package (Poisot
307 et al., 2012). Differences in species feeding interactions between networks (WN) was calculated. We
308 further calculated whether these differences originated from differences in species composition of food
309 webs (S), differences in feeding interactions in food webs as response to species turnover (ST) or
310 because shared species between food webs had different feeding interactions (OS). To compare the
311 observed and predicted feeding habits of both testate amoeba species, we used a PCAs to compare

312 the distribution and variability of feeding habits and ANOSIMs to test statistically whether observed
313 and predicted feeding habits were similar.

314 All computations and statistical analyses were performed in *R* (R Core Team, 2020) using custom-
315 written codes and available *R* packages. Several of the food web metrics including the food web graph
316 were performed using the *igraph* package (Csardi et al., 2006).

317

318 **3. Results**

319 **3.1 Predictive performance and sensitivity**

320 All ML models achieved high AUC (> 0.80), TSS (> 0.50) and Accuracy (> 0.90) values. More particularly,
321 Tree-based (BRT, RF) and neural networks (NN) achieved higher AUC (0.85-0.96), TSS (0.60-0.93), and
322 Accuracy (0.94-0.98) than parametric (GLM), distance-based (kNN) and probabilistic classifiers (BGLM)
323 models (Fig. 2a, b, and c; Table S8). Among the three best ML algorithms, NN performance was
324 comparable to BRT and RF with regard to AUC and TSS, but its overall Accuracy was lower. BRT showed
325 a lower variability in its performance than RF with lower standard deviations in regard to AUC, TSS, and
326 Accuracy (Fig. 2a, b, and c). Feeding link predictions were driven by the same clusters of predictive
327 variables in most ML models (Fig. 2d), with clusters 1 (i.e. variables related to the body size of the prey),
328 4 (i.e. Feeding types of the prey) and 7 (i.e. body size of the consumers) being the most important. As
329 BRT overall performed better and provided robust predictions using the same clusters of predictors as
330 other ML models, we selected BRT to create predictive microbial food webs.

331 Rigorous sensitivity analyses further revealed that feeding link predictions from BRT were robust
332 to the falsity of certain predictors (PVS-SA), and/or the inclusion of 'new' species in the taxa list (PT-SA;
333 Fig. S3). BRT predictions remained relatively good (AUC > 0.80 , TSS > 0.80 , Accuracy > 0.90) when
334 one to three predictors were wrongly referenced to species but strongly decreased when six predictors
335 did not correctly represent species (PVS-SA: AUC < 0.80 , TSS < 0.80 , Accuracy < 0.90 ; Fig. S3a). Finally,
336 the inclusion of up to 50% 'new' species in the database did not impair feeding link predictions through
337 BRT, with AUC, TSS, and Accuracy values remaining >0.90 in all cases (PT-SA; Fig. S3b, Table S9).

338

339 **3.2 Inferring the microbial food web for *Sphagnum*-dominated peatland**

340 We inferred the feeding links and calculated link strength for the twelve peatland plots (Fig. S4). The
341 binary metaweb (i.e. summary of all species interactions across all plots) had a total of 847 links, for a

342 connectance of 0.10 (Fig. 3a). The binary metaweb is also highly nested (nestedness = 0.96), indicating
343 that specialist species feed on a subset of prey of the most generalist species. Approximately 30% of
344 the pairwise interactions across the twelve food webs had a link strength <0.1 , indicating that some
345 link inferences are possibly weak. We, therefore, pruned the binary food webs according to link
346 strength. We iteratively removed feeding links in each food web with a strength of >0.1 to >0.99 and
347 found two thresholds, at 0.1 and 0.99, respectively (Fig. S5). The number of links strongly decreased
348 when links with a strength <0.1 were dropped (-82 links on average; Fig. S5a) while the mean link
349 strength of the food webs increased from 0.62 to 0.86 on average (Fig. S5b). The removal of links with
350 a strength <0.1 also influenced food web properties, which diverged from food webs with all links (Fig.
351 S5c, d). Removing links with a strength between 0.2 and 0.9 only slightly influenced the food web's
352 species interactions (food web β -diversity ranged between 0.22 and 0.28; Fig. S5c), whilst keeping only
353 very strong links (>0.99) in the food webs significantly reduced the number of links by 65% (Fig. S5a)
354 and drastically changed food web properties (Fig. S5c, d). Hence, we refined our food webs with a link
355 strength >0.1 .

356 Overall, the refined food webs had 627 links and a connectance of 0.07 (Fig. 3a). Refined food
357 webs based on link strength differed from binary food webs, especially in terms of species interactions
358 (food web β -diversity OS ~ 0.14 ; Fig. 3b). The PCA analysis further showed that refined link-strength
359 food webs had higher stability (SPL) and higher mean number of predators per prey (V , sdV , and $Pred$)
360 compared to binary food webs (ANOSIM, $P = 0.018$, $R = 0.17$; Fig. 3c).

361 The inclusion of link strength in the food webs clearly highlighted the main flows within the food
362 webs, highlighting the main predation routes in the food webs and/or the most important predators in
363 the network (Fig. 3a, Fig. S4). The inclusion of weighted metrics in addition to unweighted metrics in a
364 PCA analysis extended our comprehension of the food web topological structures (ANOSIM, $P = 0.001$,
365 $R = 0.38$; Fig. 3d). Weighted link-strength food webs showed a higher connectance (C), efficiency (eff),
366 and average connectivity ($neigh$) than unweighted link-strength food webs. Food chains (FCL) also
367 lengthened in weighted LS food webs, while the importance of omnivory (OI) and intermediate species
368 (Int) was more prevalent than in unweighted food webs (Fig. 3d). Finally, weighted link-strength food
369 webs better defined the importance of top predators in the network as shown by higher vulnerability
370 (V , sdV) and proportion of predators ($Pred$).

371 Our model predicted well the feeding and non-feeding interactions of the two testate amoeba
372 species. Both species showed high sensitivity values (*H. papilio* = 1; *N. tincta* = 0.87 on average; Fig.
373 4a) but null specificity values (specificity = 0 for both species). This shows that our model predicted
374 non-observed feeding links for these two species. Moreover, PCAs showed that the feeding habits of
375 the two testate amoeba species were relatively similar between microscopic observations and ML
376 predictions, though predicted feeding habits were less variable than those observed (Fig. 4b, d). Both
377 microscopic observations and ML predictions showed that *H. papilio* primarily fed on microalgae, fungi,
378 ciliates, and testate amoebae (ANOSIM, $P = 0.3$, $R = 0.033$; Fig. 4c). On the opposite, observed feeding
379 habits of *N. tincta* slightly differed from predicted one, mostly because ML predicted predation on
380 testate amoebae and rotifers in all plots while those feeding interactions were not always observed
381 (ANOSIM, $P = 0.02$, $R = 0.26$; Fig. 4d). Nevertheless, the overall feeding habits of *N. tincta* was very
382 similar between observed and predicted conditions, with microalgae, cyanobacteria, and fungi as the
383 main prey resources (Fig. 4e).

384

385 **Discussion**

386 Understanding the structure and functioning of microbial food webs is of both fundamental and applied
387 interest considering their role in ecosystems (Geisen, 2021). This paper shows that it is possible to infer
388 microbial feeding links with satisfying accuracy using a robust and repeatable machine learning
389 approach based on microbial traits and taxonomy. The data required to perform this methodology are
390 relatively simple to obtain, though more effort is required to acquire the absolute abundance data of
391 species to extend the reconstruction of the basic food web by incorporating feeding links strength.
392 Besides the calculation of link strength, our method can help generate baseline expectations about the
393 microbial food web structure in ecosystems (binary food webs) and thereby overcoming laborious
394 sampling and laboratory efforts required for traditional food web construction. In addition, we show
395 that the use of link strength provides an estimation of the main routes in the food web, which could be
396 relevant for testing basic microbial ecology theory (van Altena et al., 2016) and biodiversity-function
397 relationships (Thompson et al., 2012; Geisen, 2021; Berlinches de Gea et al., 2022; Gaüzère et al.,
398 2022). In summary, this approach will help guide new empirical research aimed at developing
399 mechanistically informed predictions about the magnitudes of microbial effects on ecosystem functions
400 (Hunt et al., 1987; De Ruiter et al., 1993; Thompson and Townsend, 2005; Grass et al., 2020). Recent

401 studies demonstrated the importance of food web's topology-based metrics and energy fluxes
402 between species in driving ecosystem functions, while species richness had only weak effects (Kortsch
403 et al., 2021; Jassey et al., 2023; Wu et al., 2023). These results clearly show that the assessments of
404 microbial feeding interactions in soils using our approach will be key for understanding the
405 consequences of soil biodiversity for the maintenance of ecosystem functions and services that are
406 critical for ecosystem sustainability.

407 Using recently described approaches for inference of species interactions in ecological networks
408 (Laigle et al., 2018a; Pichler et al., 2019; Pontarp et al., 2019; Pecuchet et al., 2020), we assessed the
409 ability of six ML models to predict microbial feeding interactions based on taxonomy and traits. We
410 found that the performance of the tree- and neuron-based models, namely BRT, RF, and NN, exceeded
411 the distance- and parametric-based models (kNN, BGLM, and GLM respectively) performances for
412 predicting microbial feeding links (Fig. 2). BRT was the best-performing ML algorithm with a low
413 sensitivity to model and predictor stochasticity (Fig. 2). It predicted 98% of the microbial feeding links
414 correctly, with a narrow variability between runs (Fig. 2). This is a very encouraging result and supports
415 previous findings on the inference of trophic interactions between soil fauna (Laigle et al., 2018a).
416 Another important point is that BRT (AUC >0.9) by far out-performed naïve random models, for
417 example, when feeding interactions were not tied to traits (AUC of 0.8; Fig. S3). Moreover, we
418 acknowledge that BRT can exhibit noticeable overfitting on some datasets, especially in case of
419 overlapping classes (Vezhnevets and Barinova, 2007). However, this is never the case of RF in which
420 collinearity decreases performance rather than leading to an overfit (Lesmeister, 2019). Because RF
421 performed almost as good as BRT (Fig. 2, Table S8), this clearly shows that there were no
422 multicollinearity nor overfitting problems in our model training and resulting feeding predictions.
423 Further, these results are in line with earlier findings on species interactions in other ecosystems (e.g.
424 Pichler et al. 2019; Pomeranz et al. 2019) and confirm the importance of the trait-matching approach
425 to infer biotic interactions between species (Gravel et al., 2013; Olito and Fox, 2015; Laigle et al.,
426 2018a; Pichler et al., 2019; Pomeranz et al., 2019; Pecuchet et al., 2020).

427 Our results showed that inferences of feeding interactions between microorganisms were
428 strongly determined by traits related to body size (i.e., cluster 7; Fig. 2, Table S4). This is in agreement
429 with the food web theory (Brose et al., 2019; Brose, 2020) and previous findings on soil fauna (Laigle
430 et al., 2018a) and aquatic food webs (Pecuchet et al., 2020). Notable exceptions to the average ratio

431 of predator-to-prey body mass can exist when microorganisms cooperate to feed on larger prey (e.g.
432 pack hunters; Geisen *et al.* 2015). However, the use of additional traits overcomes such potential bias
433 in the predictions (Laigle *et al.*, 2018a), as shown by our results. Traits related to feeding habits
434 (nourishment) were the second most important cluster of predictors in our model training, followed by
435 taxonomy (i.e. kingdom; Fig. 2). Indeed, the kingdom of the prey (resource) and the general diet of the
436 consumers is a general way to approximate the backbone of microbial food webs (Geisen *et al.*, 2016).
437 More particularly, kingdom and diet help to identify trivially prohibited interactions (Morales-Castilla *et al.*
438 *et al.*, 2015), such as microalgae not feeding on algivores. Based on the growing interest in applying trait-
439 based concepts to predict the microbial mechanisms driving global biogeochemical cycles (Malik *et al.*
440 *et al.*, 2020), one could expect rapid progress in increasing the quality of microbial trait information further
441 improving predictions of microbial feeding interactions using the trait-matching approach.

442 By applying our method to case study data, we inferred microbial food webs with 232 to 388
443 feeding links for peatland microbial communities of 41 to 53 taxa. Cross-validation of feeding
444 inferences with feeding habit observations of two testate amoeba species performed in the same plots
445 was satisfying. While our model predicted well the observed feeding and non-feeding links of the two
446 testate amoeba species, it also predicted non-observed feeding links for these two species. For
447 example, all inferred food webs systematically predicted feeding interactions between *N. tinctoria* and
448 other testate amoebae, while such feeding interactions were not observed in every plot. This result is
449 not surprising. It is extremely difficult to document all realized feeding links at a specific site even when
450 sampling effort is high (Woodward *et al.*, 2005). This is particularly true for microorganisms where
451 identifying feeding interactions is always a challenge (Antoniewicz, 2020). Furthermore, the inference
452 of predicted but not observed empirically links is encouraging as it may refine the current
453 understanding of microbial interactions in ecosystems.

454 Approximately 27% of the inferred links had a very low strength (<0.1), suggesting that these
455 links were weak and unlikely given the abundance of resources and the moving speed of the
456 consumers. This indicates that the link strength calculation is an effective approach to discriminating
457 against links that may not occur empirically while retaining those that do. Another potential advantage
458 of calculating link strength is the ability to identify the main routes in the food web, rather than inferring
459 a simple binary network (Laigle *et al.*, 2018a; Pomeranz *et al.*, 2019; Pecuchet *et al.*, 2020). Unweighted
460 (binary) food webs treat all links as if they are equally important to the food web. However, it is very

461 common in food webs that feeding rates vary considerably in magnitude between predators and/or
462 among prey of the same predator (Jassey et al., 2012a; Mieczan et al., 2015). Our results showed that
463 approximately 65% of the inferred links were of weak to intermediate strength. This pattern
464 corroborates previous findings on interaction strengths in natural food webs (Paine, 1992; McCann et
465 al., 1998; Neutel et al., 2002) showing that interaction strengths are usually characterized by many weak
466 interactions and a few strong ones. This is an encouraging result. Furthermore, weighted food web
467 properties based on link strength significantly differed from unweighted food webs, especially in terms
468 of connectance. Hence, link strength can be used to refine our predictive food webs, aiding in our
469 ability to infer microbial food web properties. Connectance based on weighted links captures well food
470 web stability (van Altena et al., 2016), suggesting that our method, in addition to predicting food web
471 structures, could be used to measure food web stability in ecosystems. Estimating microbial food web
472 stability is relevant because it is often at this level of multitrophic microbial communities that questions
473 regarding the fragility of ecosystem structure and functioning under environmental change are most
474 pressing (Naylor et al., 2020; Geisen, 2021).

475 Our approach has some limitations and therefore paves the way for future research. First, the
476 taxonomic resolution of basal species or groups (i.e. bacteria and fungi) was low and consequently
477 influenced the number of species and links in each food web. This limitation does not impact the
478 analysis of trophic interactions between basal species and predators, but certainly hides some
479 important routes in the food web as well as some variability among sites. For instance, competition
480 between fungi and bacteria shapes their community structures (Bahram et al., 2018), which in return
481 most likely influence feeding interactions and food web structure (Morriën et al., 2017). Moreover,
482 predation pressure on basal species is likely unequal between the various bacterial and fungal species
483 as a result of a combination of differential prey defence strategies and/or predator traits (Gao et al.,
484 2018; Amacker et al., 2022). Some bacteria can, for instance, produce antimicrobial compounds to
485 repel predators (Mazzola et al., 2009; Jousset and Bonkowski, 2010; Hamard et al., 2019). Therefore,
486 increasing our knowledge of feeding interactions between basal species and their respective predators
487 will certainly improve our understanding of the microbial food web structure and functioning. Second,
488 feeding link inferences are somewhat limited to taxa present in the training data set. Our training, test
489 and Case study data sets were comparable in terms of traits distribution, although significantly different
490 (Fig. S7). It may limit feeding link predictions if the number of taxa not documented in the training set

491 increases drastically. However, our sensitivity analysis revealed that the inclusion of 'new' taxa (up to
492 50% at the species level) does not impair feeding link inferences (Fig. S3, Table S9). This shows that
493 our model is robust and could easily be applied in other systems. Nevertheless, continuing efforts to
494 document microbial traits and feeding interactions is imperative for the reliable application of
495 predictive modelling.

496 So far, microbial food webs have been mostly considered at broad functional group levels
497 (Mestre et al., 2022). The concepts of integrating taxonomic diversity in food webs have been
498 overlooked (Geisen et al., 2016; Glibert and Mitra, 2022), mostly because of methodological limitations
499 (Geisen and Bonkowski, 2018). We offer a new approach to constructing highly resolved qualitatively
500 and quantitatively microbial food webs using species lists, taxonomy, and traits. Despite some
501 limitations, we showed how microbial trait composition and values, and taxonomy, determine both the
502 distribution of feeding interactions in a microbial community and identify the main routes in the food
503 web. The growing interest in multitrophic microbial interactions in soils (Geisen et al., 2016; Amacker
504 et al., 2022; Potapov et al., 2022) will likely improve the documentation of microbial feeding
505 interactions and traits, and thereby the precision of future inferences of microbial food webs using our
506 approach. The next challenge will be to relate microbial food web functioning and properties to
507 ecosystem functioning such as carbon and nutrient cycling (Geisen, 2021). Future integrated work
508 combining traditional methods (trait and feeding habit documentation), molecular techniques (DNA-
509 based high-throughput species identification) and food web inference in different environmental
510 settings offer the opportunity of upscaling soil ecological studies.

511 **Acknowledgements**

512 This work has been supported by the MIXOPEAT project funded by the French National Research
513 Agency (Grant No. ANR-17-CE01-0007 to VEJJ). JMB was kindly supported by the Dutch Science
514 Foundation (grant no. ENW-M OCENW.M20.339) during the final writing stages of this paper.

515

516 **Data and codes accessibility**

517 The microbial feeding link registry used to train the machine learning models is available from FigShare
518 in .csv format (<https://doi.org/10.6084/m9.figshare.23669019.v1>). The species lists, traits and inferred
519 microbial feeding links from the Case Study are available from FigShare in .zip format
520 (<https://doi.org/10.6084/m9.figshare.24081102.v1>). The feeding habit data on the two testate amoeba
521 species used to validate feeding inferences are available from FigShare in .csv format
522 (<https://doi.org/10.6084/m9.figshare.24081435.v1>). The R codes used in this article are available from
523 FigShare (<https://doi.org/10.6084/m9.figshare.24081480>). All Data and R codes related to this paper
524 are available from the Collection *Data and R codes from Uncovering microbial food webs using*
525 *machine learning* <https://doi.org/10.6084/m9.figshare.c.6814071>.

526

527 **References**

- 528 Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models:
529 Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43, 1223–1232.
530 doi:10.1111/j.1365-2664.2006.01214.x
- 531 Amacker, N., Gao, Z., Hu, J., Jousset, A.L.C., Kowalchuk, G.A., Geisen, S., 2022. Protist feeding
532 patterns and growth rate are related to their predatory impacts on soil bacterial communities.
533 *FEMS Microbiology Ecology* 98, 1–11. doi:10.1093/FEMSEC/FIAC057
- 534 Antoniewicz, M.R., 2020. A guide to deciphering microbial interactions and metabolic fluxes in
535 microbiome communities. *Current Opinion in Biotechnology* 64, 230–237.
536 doi:10.1016/J.COPBIO.2020.07.001
- 537 Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M.,
538 Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H.,
539 Maltz, M.R., Mundra, S., Olsson, P.A., Pent, M., Pölme, S., Sunagawa, S., Ryberg, M., Tedersoo,
540 L., Bork, P., 2018. Structure and function of the global topsoil microbiome. *Nature* 560, 233+.
- 541 Barrat, A., Barthélemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex
542 weighted networks. *Proceedings of the National Academy of Sciences of the United States of*
543 *America* 101, 3747–3752. doi:10.1073/PNAS.0400087101/ASSET/7970A6FD-0C68-49D2-AB37-
544 938124DCEAFA/ASSETS/GRAPHIC/ZPQ0080439150007.JPEG
- 545 Berlinches de Gea, A., Hautier, Y., Geisen, S., 2022. Interactive effects of global change drivers as
546 determinants of the link between soil biodiversity and ecosystem functioning. *Global Change*
547 *Biology* 29, 296–307. doi:10.1111/GCB.16471
- 548 Brose, U., 2020. Trait-based models of complex ecological networks. *Theoretical Ecology* 134–142.
549 doi:10.1093/OSO/9780198824282.003.0009
- 550 Brose, U., Archambault, P., Barnes, A.D., Bersier, L.-F., Boy, T., Canning-Clode, J., Conti, E., Dias, M.,
551 Digel, C., Dissanayake, A., Flores, A.A. V, Fussmann, K., Gauzens, B., Gray, C., Häussler, J., Hirt,
552 M.R., Jacob, U., Jochum, M., Kefi, S., McLaughlin, O., MacPherson, M.M., Latz, E., Layer-Dobra,
553 K., Legagneux, P., Li, Y., Madeira, C., Martinez, N.D., Mendonça, V., Mulder, C., Navarrete, S.A.,
554 O’Gorman, E.J., Ott, D., Paula, J., Perkins, D., Piechnik, D., Pokrovsky, I., Raffaelli, D., Rall, B.C.,
555 Rosenbaum, B., Ryser, R., Silva, A., Sohlström, E.H., Sokolova, N., Thompson, M.S.A.,
556 Thompson, R.M., Vermandele, F., Vinagre, C., Wang, S., Wefer, J.M., Williams, R.J., Wieters, E.,

557 Woodward, G., Iles, A.C., 2019. Predator traits determine food-web architecture across
558 ecosystems. *Nature Ecology & Evolution* 1–11.

559 Buchkowski, R.W., Lindo, Z., 2020. Stoichiometric and structural uncertainty in soil food web models.
560 *Functional Ecology* 56, 113–122.

561 Chavent, M., Kuentz-Simonet, V., Lique, B., Saracco, J., 2012. ClustOfVar: An R package for the
562 clustering of variables. *Journal of Statistical Software* 50, 1–16. doi:10.18637/jss.v050.i13

563 Csardi, G., InterJournal, T.N., Systems, C., 2006, 2006. The igraph software package for complex
564 network research. *InterJournal, Complex Systems*, 1695, 1–9.

565 De Ruiter, P.C., Van Veen, J.A., Moore, J.C., Brussaard, L., Hunt, H.W., 1993. Calculation of nitrogen
566 mineralization in soil food webs. *Plant and Soil* 157, 263–273. doi:10.1007/bf00011055

567 de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot,
568 N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron,
569 S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horák, A., Jaillon, O., Lima-Mendez, G.,
570 Lukeš, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A.,
571 Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Coordinators, T.O., Acinas, S.G., Bork, P.,
572 Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S.,
573 Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P.,
574 Karsenti, E., 2015. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science*
575 348, 1261605.

576 de Vries, F.T., Griffiths, R.I., Bailey, M., Craig, H., Girlanda, M., Gweon, H.S., Hallin, S., Kaisermann,
577 A., Keith, A.M., Kretzschmar, M., Lemanceau, P., Lumini, E., Mason, K.E., Oliver, A., Ostle, N.,
578 Prosser, J.I., Thion, C., Thomson, B., Bardgett, R.D., 2018. Soil bacterial networks are less stable
579 under drought than fungal networks. *Nature Communications* 9, 912–913.

580 Delgado Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-Gonzalez, A., Eldridge, D.J., Bardgett,
581 R.D., Maestre, F.T., Singh, B.K., Fierer, N., 2018. A global atlas of the dominant bacteria found
582 in soil. *Science* 359, 320–325.

583 Desjardins-Proulx, P., Laigle, I., Poisot, T., Gravel, D., 2017. Ecological interactions and the Netflix
584 problem. *PeerJ* 2017, e3644. doi:10.7717/PEERJ.3644/FIG-3

585 Escalas, A., Hale, L., Voordeckers, J.W., Yang, Y., Firestone, M.K., Alvarez-Cohen, L., Zhou, J., 2019.
586 Microbial functional diversity: From concepts to applications. *Ecology and Evolution* 9, 12000.

587 doi:10.1002/ECE3.5670

588 Faisal, A., Dondelinger, F., Husmeier, D., Beale, C.M., 2010. Inferring species interaction networks
589 from species abundance data: A comparative evaluation of various statistical and machine
590 learning methods. *Ecological Informatics* 5, 451–464. doi:10.1016/j.ecoinf.2010.06.005

591 Falkowski, P.G., Fenchel, T., Delong, E.F., 2008. The microbial engines that drive earth’s
592 biogeochemical cycles. *Science* 320, 1034–1039.
593 doi:10.1126/SCIENCE.1153213/ASSET/6B08817F-E1AD-4130-A430-
594 2F3CBF3177F0/ASSETS/GRAPHIC/320_1034_F3.JPEG

595 Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nature Reviews*.
596 *Microbiology* 10, 538–550.

597 Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G., Jousset, A., 2018. Protists: Puppet Masters of the
598 Rhizosphere Microbiome. *Trends in Plant Science*.

599 Gaüzère, P., O’Connor, L., Botella, C., Poggiato, G., Münkemüller, T., Pollock, L.J., Brose, U.,
600 Maiorano, L., Harfoot, M., Thuiller, W., 2022. The diversity of biotic interactions complements
601 functional and phylogenetic facets of biodiversity. *Current Biology* 32, 2093-2100.e3.
602 doi:10.1016/j.cub.2022.03.009

603 Geisen, S., 2021. The Future of (Soil) Microbiome Studies: Current Limitations, Integration, and
604 Perspectives. *MSystems* 6, e00613-21.

605 Geisen, S., Bonkowski, M., 2018. Methodological advances to study the diversity of soil protists and
606 their functioning in soil food webs. *Applied Soil Ecology* 123, 328–333.
607 doi:10.1016/J.APSOIL.2017.05.021

608 Geisen, S., Koller, R., Huenninghaus, M., Dumack, K., Urich, T., Bonkowski, M., 2016. The soil food
609 web revisited: Diverse and widespread mycophagous soil protists. *Soil Biology and Biochemistry*
610 94, 10–18.

611 Geisen, S., Rosengarten, J., Koller, R., Mulder, C., Urich, T., Bonkowski, M., 2015. Pack hunting by a
612 common soil amoeba on nematodes. *Environmental Microbiology* 17, 4538–4546.

613 Gilbert, D., Mitchell, E.A.D., 2006. Chapter 13 Microbial diversity in Sphagnum peatlands.
614 *Developments in Earth Surface Processes* 9, 287–318. doi:10.1016/S0928-2025(06)09013-4

615 Glibert, P.M., Mitra, A., 2022. From webs, loops, shunts, and pumps to microbial multitasking:
616 Evolving concepts of marine microbial ecology, the mixoplankton paradigm, and implications for

617 a future ocean. *Limnology and Oceanography* 67, 585–597. doi:10.1002/LNO.12018

618 Goberna, M., Verdú, M., 2022. Cautionary notes on the use of co-occurrence networks in soil
619 ecology. *Soil Biology and Biochemistry* 166, 108534. doi:10.1016/J.SOILBIO.2021.108534

620 Grass, I., Kubitzka, C., Krishna, V. V., Corre, M.D., Mußhoff, O., Pütz, P., Drescher, J., Rembold, K.,
621 Ariyanti, E.S., Barnes, A.D., Brinkmann, N., Brose, U., Brümmer, B., Buchori, D., Daniel, R.,
622 Darras, K.F.A., Faust, H., Fehrmann, L., Hein, J., Hennings, N., Hidayat, P., Hölscher, D., Jochum,
623 M., Knohl, A., Kotowska, M.M., Krashevskaya, V., Kreft, H., Leuschner, C., Lobite, N.J.S., Panjaitan,
624 R., Polle, A., Potapov, A.M., Purnama, E., Qaim, M., Röhl, A., Scheu, S., Schneider, D., Tjoa, A.,
625 Tschardt, T., Veldkamp, E., Wollni, M., 2020. Trade-offs between multifunctionality and profit
626 in tropical smallholder landscapes. *Nature Communications* 11, 1113–1186.

627 Gravel, D., Poisot, T., Albouy, C., Velez, L., Mouillot, D., 2013. Inferring food web structure from
628 predator-prey body size relationships. *Methods in Ecology and Evolution* 4, 1083–1090.

629 Gray, C., Figueroa, D.H., Hudson, L.N., Ma, A., Perkins, D., Woodward, G., 2015. Joining the dots:
630 An automated method for constructing food webs from compendia of published interactions.
631 *Food Webs* 5, 11–20. doi:10.1016/J.FOOWEB.2015.09.001

632 Guerra, C.A., Berdugo, M., Eldridge, D.J., Eisenhauer, N., Singh, B.K., Cui, H., Abades, S., Alfaro,
633 F.D., Bamigboye, A.R., Bastida, F., Blanco-Pastor, J.L., de los Ríos, A., Durán, J., Grebenc, T.,
634 Illán, J.G., Liu, Y.-R., Makhalyane, T.P., Mamet, S., Molina-Montenegro, M.A., Moreno, J.L.,
635 Mukherjee, A., Nahberger, T.U., Peñaloza-Bojacá, G.F., Plaza, C., Picó, S., Verma, J.P., Rey, A.,
636 Rodríguez, A., Tedersoo, L., Teixido, A.L., Torres-Díaz, C., Trivedi, P., Wang, Juntao, Wang, L.,
637 Wang, Jianyong, Zaady, E., Zhou, X., Zhou, X.-Q., Delgado-Baquerizo, M., 2022. Global
638 hotspots for soil nature conservation. *Nature* 2022 1–6. doi:10.1038/s41586-022-05292-x

639 Guo, S., Tao, C., Jousset, A., Xiong, W., Wang, Z., Shen, Z., Wang, B., Xu, Z., Gao, Z., Liu, S., Li, R.,
640 Ruan, Y., Shen, Q., Kowalchuk, G.A., Geisen, S., 2022. Trophic interactions between predatory
641 protists and pathogen-suppressive bacteria impact plant health. *The ISME Journal* 2022 16:8 16,
642 1932–1943. doi:10.1038/s41396-022-01244-5

643 Hamard, S., Robroek, B.J.M., Allard, P.-M., Signarbieux, C., Zhou, S., Saesong, T., de Baaker, F.,
644 Buttler, A., Chiapusio, G., Wolfender, J.-L., Bragazza, L., Jasse, V.E.J., 2019. Effects of
645 Sphagnum Leachate on Competitive Sphagnum Microbiome Depend on Species and Time.
646 *Frontiers in Microbiology* 10, 3317.

647 Holtkamp, R., van der Wal, A., Kardol, P., van der Putten, W.H., de Ruiter, P.C., Dekker, S.C., 2011.
648 Modelling C and N mineralisation in soil food webs during secondary succession on ex-arable
649 land. *Soil Biology and Biochemistry* 43, 251–260. doi:10.1016/J.SOILBIO.2010.10.004

650 Hunt, H.W., Coleman, D.C., Ingham, E.R., Ingham, R.E., Elliott, E.T., Moore, J.C., Rose, S.L., Reid,
651 C.P.P., Morley, C.R., 1987. The detrital food web in a shortgrass prairie. *Biology and Fertility of*
652 *Soils* 3, 57–68.

653 Jansson, J.K., Hofmockel, K.S., 2020. Soil microbiomes and climate change. *Nature Reviews*
654 *Microbiology* 18, 35–46. doi:10.1038/s41579

655 Jassey, V., Petchey, O., Binet, P., Buttler, A., Chiapusio, G., Delarue, F., Laggoun-Défarge, F., Gilbert,
656 D., Mitchell, E.A., Barel, J., 2023. Food Web Structure and Energy Flux Dynamics, But Not
657 Taxonomic Richness, Influence Microbial Ecosystem Functions in A Sphagnum-Dominated
658 Peatland. SSRN under revision. doi:10.2139/SSRN.4459208

659 Jassey, V.E.J., Chiapusio, G., Binet, P., Buttler, A., Laggoun-Défarge, F., Delarue, F., Bernard, N.,
660 Mitchell, E.A.D., Toussaint, M.-L., Francez, A.-J., Gilbert, D., 2013a. Above- and belowground
661 linkages in Sphagnum peatland: climate warming affects plant-microbial interactions. *Global*
662 *Change Biology* 19, 811–823.

663 Jassey, V.E.J., Chiapusio, G., Mitchell, E.A.D., Binet, P., Toussaint, M.-L., Gilbert, D., 2011. Fine-scale
664 horizontal and vertical micro-distribution patterns of testate amoebae along a narrow Fen/Bog
665 gradient. *Microbial Ecology* 61, 374–385.

666 Jassey, V.E.J., Meyer, C., Dupuy, C., Bernard, N., Mitchell, E.A.D., Toussaint, M.-L., Metian, M.,
667 Chatelain, A.P., Gilbert, D., 2013b. To What Extent Do Food Preferences Explain the Trophic
668 Position of Heterotrophic and Mixotrophic Microbial Consumers in a Sphagnum Peatland?
669 *Microbial Ecology* 66, 571–580.

670 Jassey, V.E.J., Shimano, S., Dupuy, C., Toussaint, M.-L., Gilbert, D., 2012a. Characterizing the
671 Feeding Habits of the Testate Amoebae *Hyalosphenia papilio* and *Nebela tinctoria* along a Narrow
672 “Fen-Bog” Gradient Using Digestive Vacuole Content and ¹³C and ¹⁵N Isotopic Analyses.
673 *Protist* 163, 451–464. doi:https://doi.org/10.1016/j.protis.2011.07.006

674 Jassey, V.E.J., Shimano, S., Dupuy, C., Toussaint, M.-L., Gilbert, D., 2012b. Characterizing the
675 feeding habits of the testate amoebae *Hyalosphenia papilio* and *Nebela tinctoria* along a narrow
676 “fen-bog” gradient using digestive vacuole content and ¹³C and ¹⁵N isotopic analyses. *Protist*

677 163, 451–464.

678 Jassey, V.E.J., Signarbieux, C., Hättenschwiler, S., Bragazza, L., Buttler, A., Delarue, F., Fournier, B.,
679 Gilbert, D., Laggoun-Défarge, F., Lara, E., Mills, R.T.E., Mitchell, E.A.D., Payne, R.J., Robroek,
680 B.J.M., 2015. An unexpected role for mixotrophs in the response of peatland carbon cycling to
681 climate warming. *Scientific Reports* 5, 16910–16931.

682 Jassey, V.E.J., Walcker, R., Kardol, P., Geisen, S., Heger, T., Lamentowicz, M., Hamard, S., Lara, E.,
683 2022. Contribution of soil algae to the global carbon cycle. *New Phytologist* 234, 64–76.
684 doi:10.1111/nph.17950

685 Jousset, A., Bonkowski, M., 2010. The model predator *Acanthamoeba castellanii* induces the
686 production of 2,4, DAPG by the biocontrol strain *Pseudomonas fluorescens* Q2-87. *Soil Biology*
687 and *Biochemistry* 42, 1647–1649. doi:10.1016/J.SOILBIO.2010.05.018

688 Koltz, A.M., Asmus, A., Gough, L., Pressler, Y., Moore, J.C., 2018. The detritus-based microbial-
689 invertebrate food web contributes disproportionately to carbon and nitrogen cycling in the
690 Arctic. *Polar Biology* 41, 1531–1545. doi:10.1007/s00300-017-2201-5

691 Kortsch, S., Frelat, R., Pecuchet, L., Olivier, P., Putnis, I., Bonsdorff, E., Ojaveer, H., Jurgensone, I.,
692 Strāķe, S., Rubene, G., Krūze, Ē., Nordström, M.C., 2021. Disentangling temporal food web
693 dynamics facilitates understanding of ecosystem functioning. *Journal of Animal Ecology* 90,
694 1205–1216. doi:10.1111/1365-2656.13447

695 Kortsch, S., Primicerio, R., Aschan, M., Lind, S., Dolgov, A. V., Planque, B., 2019. Food-web structure
696 varies along environmental gradients in a high-latitude marine ecosystem. *Ecography* 42, 295–
697 308. doi:10.1111/ECOG.03443

698 Kortsch, S., Primicerio, R., Fossheim, M., Dolgov, A. V., Aschan, M., 2015. Climate change alters the
699 structure of arctic marine food webs due to poleward shifts of boreal generalists. *Proceedings of*
700 *the Royal Society B: Biological Sciences* 282. doi:10.1098/RSPB.2015.1546

701 Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of Statistical*
702 *Software* 28, 1–26.

703 Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I., Gravel, D., 2018a. Species traits as drivers of
704 food web structure. *Oikos* 127, 316–326.

705 Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I., Gravel, D., 2018b. Species traits as drivers of
706 food web structure. *Oikos* 127, 316–326. doi:10.1111/oik.04712

707 Ledger, M.E., Brown, L.E., Edwards, F.K., Milner, A.M., Woodward, G., 2013. Drought alters the
708 structure and functioning of complex food webs. *Nature Climate Change* 3, 223–227.
709 doi:10.1038/nclimate1684

710 Lesmeister, C., 2019. *Mastering machine learning with R: advanced machine learning techniques for*
711 *building smart applications with R 3.5*.

712 Liang, C., Schimel, J.P., Jastrow, J.D., 2017. The importance of anabolism in microbial control over
713 soil carbon storage. *Nature Microbiology* 2, 17105–17106.

714 Lo, W.S., Han, Z., Witte, H., Röseler, W., Sommer, R.J., 2022. Synergistic interaction of gut microbiota
715 enhances the growth of nematode through neuroendocrine signaling. *Current Biology* 32, 2037-
716 2050.e4. doi:10.1016/J.CUB.2022.03.056

717 Malik, A.A., Martiny, J.B.H., Brodie, E.L., Martiny, A.C., Treseder, K.K., Allison, S.D., 2020. Defining
718 trait-based microbial strategies with consequences for soil carbon cycling under climate change.
719 *The ISME Journal* 14, 1–9.

720 Manel, S., Ceri Williams, H., Ormerod, S.J., 2001. Evaluating presence–absence models in ecology:
721 the need to account for prevalence. *Journal of Applied Ecology* 38, 921–931.
722 doi:10.1046/J.1365-2664.2001.00647.X

723 Mazzola, M., De Bruijn, I., Cohen, M.F., Raaijmakers, J.M., 2009. Protozoan-induced regulation of
724 cyclic lipopeptide biosynthesis is an effective predation defense mechanism for *Pseudomonas*
725 *fluorescens*. *Applied and Environmental Microbiology* 75, 6804–6811. doi:10.1128/AEM.01272-
726 09/ASSET/87B6EB73-E8F8-43E6-87C5-
727 B2A934EFC244/ASSETS/GRAPHIC/ZAM0210903710005.JPEG

728 McCann, K., Hastings, A., Huxel, G.R., 1998. Weak trophic interactions and the balance of nature.
729 *Nature* 1998 395:6704 395, 794–798. doi:10.1038/27427

730 Mestre, F., Gravel, D., García-Callejas, D., Pinto-Cruz, C., Matias, M.G., Araújo, M.B., 2022.
731 Disentangling food-web environment relationships: A review with guidelines. *Basic and Applied*
732 *Ecology* 61, 102–115. doi:10.1016/J.BAAE.2022.03.011

733 Mieczan, T., Michał, N., Adamczuk, M., Bielańska Grajner, I., 2015. Stable isotope analyses revealed
734 high seasonal dynamics in the food web structure of a peatbog. *International Review of*
735 *Hydrobiology* n/a-n/a.

736 Morales-Castilla, I., Matias, M.G., Gravel, D., Araújo, M.B., 2015. Inferring biotic interactions from

737 proxies. *Trends in Ecology & Evolution* 30, 347–356. doi:10.1016/J.TREE.2015.03.014

738 Morriën, E., Hannula, S.E., Snoek, L.B., Helmsing, N.R., Zweers, H., de Hollander, M., Soto, R.L.,
739 Bouffaud, M.-L., Buée, M., Dimmers, W., Duyts, H., Geisen, S., Girlanda, M., Griffiths, R.I.,
740 Jørgensen, H.B., Jensen, J., Plassart, P., Redecker, D., Schmelz, R.M., Schmidt, O., Thomson,
741 B.C., Tisserant, E., Uroz, S., Winding, A., Bailey, M.J., Bonkowski, M., Faber, J.H., Martin, F.,
742 Lemanceau, P., de Boer, W., van Veen, J.A., van der Putten, W.H., 2017. Soil networks become
743 more connected and take up more carbon as nature restoration progresses. *Nature*
744 *Communications* 8, 14349.

745 Naylor, D., Sadler, N., Bhattacharjee, A., Graham, E.B., Anderton, C.R., McClure, R., Lipton, M.,
746 Hofmockel, K.S., Jansson, J.K., 2020. Soil microbiomes under climate change and implications
747 for carbon cycling. *Annual Review of Environment and Resources* 45, 29–59.
748 doi:10.1146/ANNUREV-ENVIRON-012320-082720

749 Neutel, A.M., Heesterbeek, J.A.P., De Ruiter, P.C., 2002. Stability in real food webs: Weak links in
750 long loops. *Science* 296, 1120–1123.
751 doi:10.1126/SCIENCE.1068326/SUPPL_FILE/NEUTELWEB.PDF

752 Oksanen, J., 2011. Multivariate analysis of ecological communities in R: vegan tutorial. R Package
753 Version.

754 Olito, C., Fox, J.W., 2015. Species traits and abundances predict metrics of plant–pollinator network
755 structure, but not pairwise interactions. *Oikos* 124, 428–436. doi:10.1111/OIK.01439

756 Oliverio, A.M., Geisen, S., Delgado Baquerizo, M., Maestre, F.T., Turner, B.L., Fierer, N., 2020. The
757 global-scale distributions of soil protists and their contributions to belowground systems.
758 *Science Advances* 6, eaax8787.

759 Paine, R.T., 1992. Food-web analysis through field measurement of per capita interaction strength.
760 *Nature* 1992 355:6355 355, 73–75. doi:10.1038/355073a0

761 Pecuchet, L., Blanchet, M.A., Frainer, A., Husson, B., Jørgensen, L.L., Kortsch, S., Primicerio, R., 2020.
762 Novel feeding interactions amplify the impact of species redistribution on an Arctic food web.
763 *Global Change Biology* 26, 4894–4906. doi:10.1111/GCB.15196

764 Petters, S., Groß, V., Söllinger, A., Pichler, M., Reinhard, A., Bengtsson, M.M., Urich, T., 2021. The soil
765 microbial food web revisited: Predatory myxobacteria as keystone taxa? *The ISME Journal* 2021
766 15:9 15, 2665–2675. doi:10.1038/s41396-021-00958-2

767 Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M., Hartig, F., 2019. Machine learning algorithms to
768 infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology*
769 *and Evolution* 11, 281–293.

770 Pichler, M., Boreux, V., Klein, A., Schleuning, M., Hartig, F., 2020. Machine learning algorithms to
771 infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology*
772 *and Evolution* 11, 281–293. doi:10.1111/2041-210X.13329

773 Pierce, E.C., Dutton, R.J., 2022. Putting microbial interactions back into community contexts. *Current*
774 *Opinion in Microbiology* 65, 56–63. doi:10.1016/J.MIB.2021.10.008

775 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2016. *nlme: Linear and Nonlinear Mixed*
776 *Effects Models*.

777 Poisot, T., Canard, E., Mouillot, D., Mouquet, N., Gravel, D., Jordan, F., 2012. The dissimilarity of
778 species interaction networks. *Ecology Letters* 15, 1353–1361.

779 Pomeranz, J.P.F., Thompson, R.M., Poisot, T., Harding, J.S., 2019. Inferring predator–prey
780 interactions in food webs. *Methods in Ecology and Evolution* 10, 356–367. doi:10.1111/2041-
781 210X.13125

782 Pontarp, M., Brannstrom, A., Petchey, O.L., 2019. Inferring community assembly processes from
783 macroscopic patterns using dynamic eco-evolutionary models and Approximate Bayesian
784 Computation (ABC). *Methods in Ecology and Evolution* 10, 450–460.

785 Potapov, A.M., Beaulieu, F., Birkhofer, K., Bluhm, S.L., Degtyarev, M.I., Devetter, M., Goncharov,
786 A.A., Gongalsky, K.B., Klärner, B., Korobushkin, D.I., Liebke, D.F., Maraun, M., Mc Donnell, R.J.,
787 Pollierer, M.M., Schaefer, I., Shrubovych, J., Semenyuk, I.I., Sendra, A., Tuma, J., Tũ Mov A, M.,
788 Vassilieva, A.B., Chen, T.-W., Geisen, S., Schmidt, O., Tiunov, A. V, Scheu, S., Blumenbach, J.F.,
789 2022. Feeding habits and multifunctional classification of soil-associated consumers from protists
790 to vertebrates. *Biol. Rev* 0–000. doi:10.1111/brv.12832

791 R Core Team, 2020. *R: A language and environment for statistical computing*. R Foundation for
792 *Statistical Computing*, Vienna, Austria.

793 Reczuga, M.K., Lamentowicz, M., Mulot, M., Mitchell, E.A.D., Buttler, A., Chojnicki, B., Słowiński, M.,
794 Binet, P., Chiapusio, G., Gilbert, D., Słowińska, S., Jassey, V.E.J., 2018. Predator–prey mass ratio
795 drives microbial activity under dry conditions in Sphagnum peatlands. *Ecology and Evolution* 8,
796 5752–5764.

797 Schimel, J.P., Schaeffer, S.M., 2012. Microbial control over carbon cycling in soil. *Frontiers in*
798 *Microbiology* 3, 348.

799 Séneca, J., Söllinger, A., Herbold, C.W., Pjevac, P., Prommer, J., Verbruggen, E., Sigurdsson, B.D.,
800 Peñuelas, J., Janssens, I.A., Urich, T., Tveit, A.T., Richter, A., 2021. Increased microbial
801 expression of organic nitrogen cycling genes in long-term warmed grassland soils. *ISME*
802 *Communications* 2021 1:1 1, 1–9. doi:10.1038/s43705-021-00073-5

803 Seppey, C.V.W., Singer, D., Dumack, K., Fournier, B., Belbahri, L.L., Mitchell, E.A.D., Lara, E., 2017.
804 Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic
805 protists as an alternative pathway for nutrient cycling. *Soil Biology and Biochemistry* 112, 68–76.

806 Singer, D., Seppey, C.V.W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., Blandenier, Q.,
807 Debross, D., de Groot, G.A., de Vargas, C., Domaizon, I., Duckert, C., Izaguirre, I., Koenig, I.,
808 Mataloni, G., Schiaffino, M.R., Mitchell, E.A.D., Geisen, S., Lara, E., 2021. Protist taxonomic and
809 functional diversity in soil, freshwater and marine ecosystems. *Environment International* 146,
810 106262. doi:10.1016/J.ENVINT.2020.106262

811 Singh, B.K., Bardgett, R.D., Smith, P., Reay, D.S., 2010. Microorganisms and climate change:
812 terrestrial feedbacks and mitigation options. *Nature Reviews. Microbiology* 8, 779–790.

813 Sytiuk, A., Céréghino, R., Hamard, S., Delarue, F., Guittet, A., Barel, J.M., Dorrepaal, E., Küttim, M.,
814 Lamentowicz, M., Pourrut, B., Robroek, B.J.M., Tuittila, E., Jassey, V.E.J., 2021. Predicting the
815 structure and functions of peatland microbial communities from Sphagnum phylogeny,
816 anatomical and morphological traits and metabolites. *Journal of Ecology* 1365-2745.13728.
817 doi:10.1111/1365-2745.13728

818 Tedersoo, L., Bahram, M., Pölme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Villarreal Ruiz, L.,
819 Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas,
820 M., Riit, T., Ratkowsky, D., Pritsch, K., Pöldmaa, K., Piepenbring, M., Phosri, C., Peterson, M.,
821 Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A.L., Nilsson, R.H., Morgado, L.N.,
822 Mayor, J., May, T.W., Majuakim, L., Lodge, D.J., Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K.,
823 Hiiesalu, I., Henkel, T.W., Harend, H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G.,
824 Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., De Kesel, A., Dang, T., Chen, X., Buegger,
825 F., Brearley, F.Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K., 2014. Fungal biogeography.
826 Global diversity and geography of soil fungi. *Science* 346, 1256688.

827 Thompson, R.M., Brose, U., Dunne, J.A., Hall, R.O., Hladyz, S., Kitching, R.L., Martinez, N.D., Rantala,
828 H., Romanuk, T.N., Stouffer, D.B., Tylianakis, J.M., 2012. Food webs: reconciling the structure
829 and function of biodiversity. *Trends in Ecology & Evolution* 27, 689–697.

830 Thompson, R.M., Townsend, C.R., 2005. Energy availability, spatial heterogeneity and ecosystem size
831 predict food-web structure in streams. *Oikos* 108, 137–148. doi:10.1111/J.0030-
832 1299.2005.11600.X

833 van Altena, C., Hemerik, L., de Ruiter, P.C., 2016. Food web stability and weighted connectance: the
834 complexity-stability debate revisited. *Theoretical Ecology* 9, 49–58. doi:10.1007/S12080-015-
835 0291-7/TABLES/3

836 van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D.A., de Goede,
837 R.G.M., Adams, B.J., Ahmad, W., Andriuzzi, W.S., Bardgett, R.D., Bonkowski, M., Campos-
838 Herrera, R., Cares, J.E., Caruso, T., de Brito Caixeta, L., Chen, X., Costa, S.R., Creamer, R.,
839 Mauro da Cunha Castro, J., Dam, M., Djigal, D., Escuer, M., Griffiths, B.S., Gutiérrez, C.,
840 Hohberg, K., Kalinkina, D., Kardol, P., Kergunteuil, A., Korthals, G., Krashevskaya, V., Kudrin, A.A.,
841 Li, Q., Liang, W., Magilton, M., Marais, M., Martín, J.A.R., Matveeva, E., Mayad, E.H., Mulder, C.,
842 Mullin, P., Neilson, R., Nguyen, T.A.D., Nielsen, U.N., Okada, H., Rius, J.E.P., Pan, K., Peneva,
843 V., Pellissier, L., Carlos Pereira da Silva, J., Pitteloud, C., Powers, T.O., Powers, K., Quist, C.W.,
844 Rasmann, S., Moreno, S.S., Scheu, S., Setälä, H., Sushchuk, A., Tiunov, A. V, Trap, J., van der
845 Putten, W., Vestergård, M., Villenave, C., Waeyenberge, L., Wall, D.H., Wilschut, R., Wright,
846 D.G., Yang, J.-I., Crowther, T.W., 2019. Soil nematode abundance and functional group
847 composition at a global scale. *Nature* 2, 1042.

848 Vezhnevets, A., Barinova, O., 2007. Avoiding boosting overfitting by removing confusing samples.
849 *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence*
850 *and Lecture Notes in Bioinformatics)* 4701 LNAI, 430–441. doi:10.1007/978-3-540-74958-
851 5_40/COVER

852 Woodward, G., Speirs, D.C., Hildrew, A.G., 2005. Quantification and Resolution of a Complex, Size-
853 Structured Food Web. *Advances in Ecological Research* 36, 85–135. doi:10.1016/S0065-
854 2504(05)36002-8

855 Wu, D., Xu, C., Wang, S., Zhang, L., Kortsch, S., 2023. Why are biodiversity—ecosystem functioning
856 relationships so elusive? Trophic interactions may amplify ecosystem function variability. *Journal*

857 of Animal Ecology 92, 367–376. doi:10.1111/1365-2656.13808
858 Xiong, W., Jousset, A., Li, R., Delgado-Baquerizo, M., Bahram, M., Logares, R., Wilden, B., de Groot,
859 G.A., Amacker, N., Kowalchuk, G.A., Shen, Q., Geisen, S., 2021. A global overview of the
860 trophic structure within microbiomes across ecosystems. Environment International 151.
861 doi:10.1016/j.envint.2021.106438
862 Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M., 2010. Mixed Effects Models and
863 Extensions in Ecology with R. Springer-Verlag, New York.
864
865
866

867 **Figure captions**

868 **Figure 1.** Conceptual diagram showing the workflow followed to infer microbial food webs.

869

870 **Figure 2.** Identification of the best-performing ML modelling to infer microbial feeding interactions. (a)
871 Performance of six algorithms in predicting microbial feeding links quantified as (a) Area under the
872 receiver operating characteristic curve (AUC), (b) true skill statistic (TSS), and (c) overall accuracy.
873 Different letters indicate significant differences ($P < 0.05$) between algorithms based on GLMs and
874 Tukey-HSD post hoc test. (d) Importance of clusters of predictors based on variable importance across
875 all models (see Fig. S2 and Table S4 for details on the clusters), (e) across all Boosted Regression Tree
876 (BRT) models. Predictors included in the best-performing BRT model are given in (f). These predictors
877 were used for sensitivity analyses and application to case study data. For the ML algorithm's
878 abbreviations see Table S3.

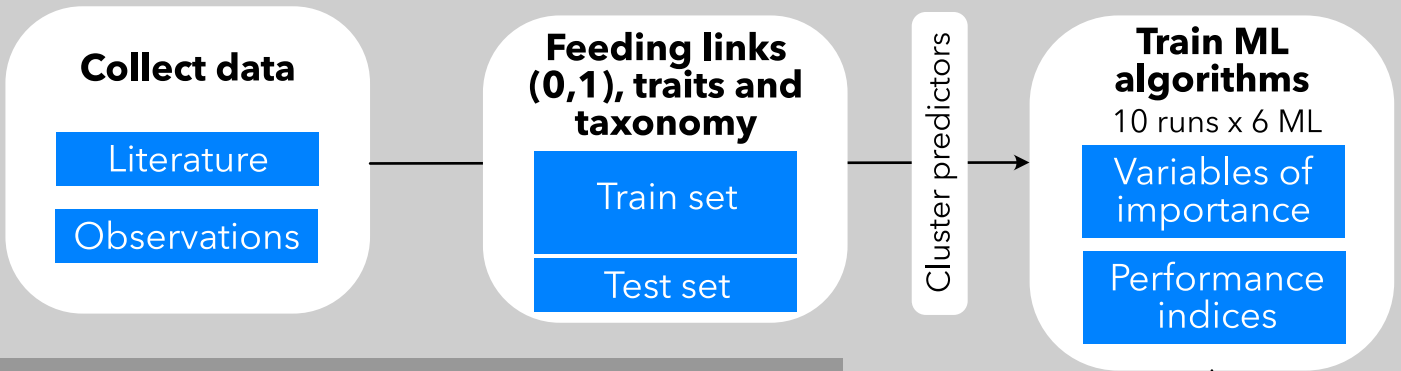
879

880 **Figure 3.** Microbial feeding inferences and food web properties of Case study data. (a) Meta-webs
881 (binary and link-strength based) summarizing the feeding interactions within the 12 microbial
882 communities from the Forbonnet peatland. (b) Network beta-diversity indices of the pairwise
883 comparison between the P/A food webs and those pruned according to link strength. OS = Dissimilarity
884 between of feeding interactions established between species common to both food webs; S =
885 dissimilarity in the species composition of the food webs; ST = dissimilarity of feeding interactions due
886 to species turnover in food webs; WN = dissimilarity of feeding interactions between food webs. (c)
887 Principal component analysis (PCA) comparing the food web properties of simple binary food webs
888 (PA) and those pruned according to link strength (AB). (d) PCA comparing of unweighted (UnW) and
889 weighted (W) food web properties for the inferred link-strength food webs.

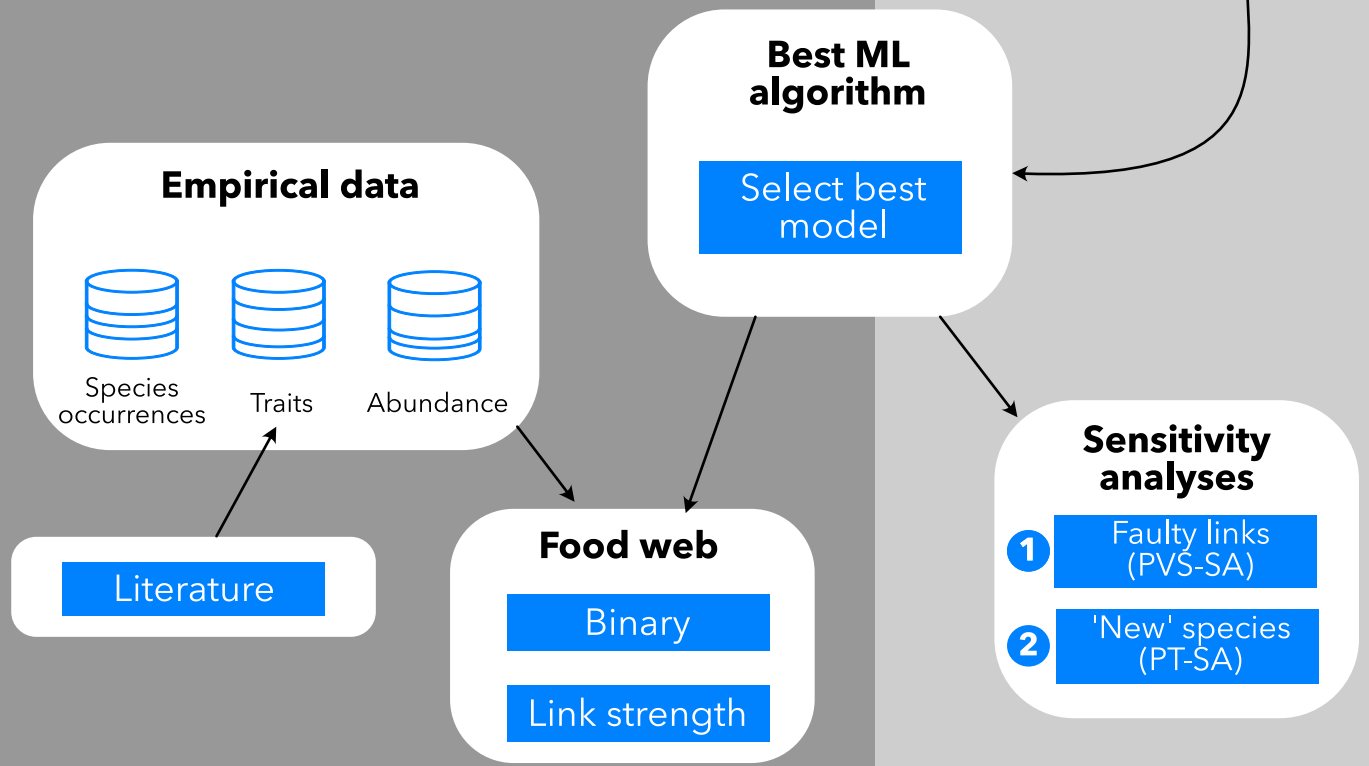
890

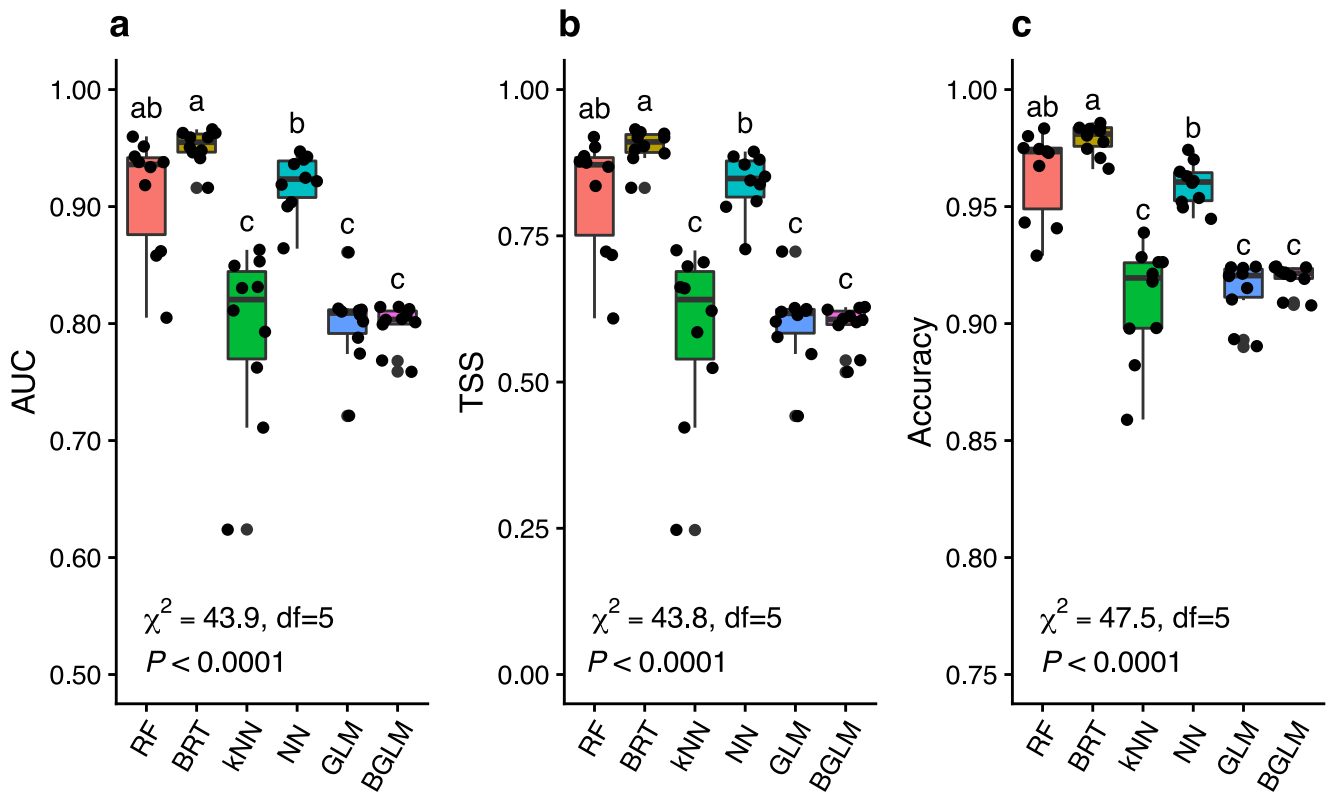
891 **Figure 4.** Cross-validation of feeding link inferences for two testate amoeba species. (a) Sensitivity
892 values evaluating the ML model's ability to predict true feeding links for the two testate amoeba species
893 (Hp: *Hyalosphenia papilio*; Nt: *Nebela tincta*). (b, d) PCAs comparing the observed and predicted
894 feeding habits of the two testate amoeba species. (c, e) Overall comparison of the general feeding
895 habit of *Hyalosphenia papilio* and *Nebela tincta* across the 12 food webs following direct microscopic
896 observations (Observed) and ML predictions (Predicted).

Workflow of development



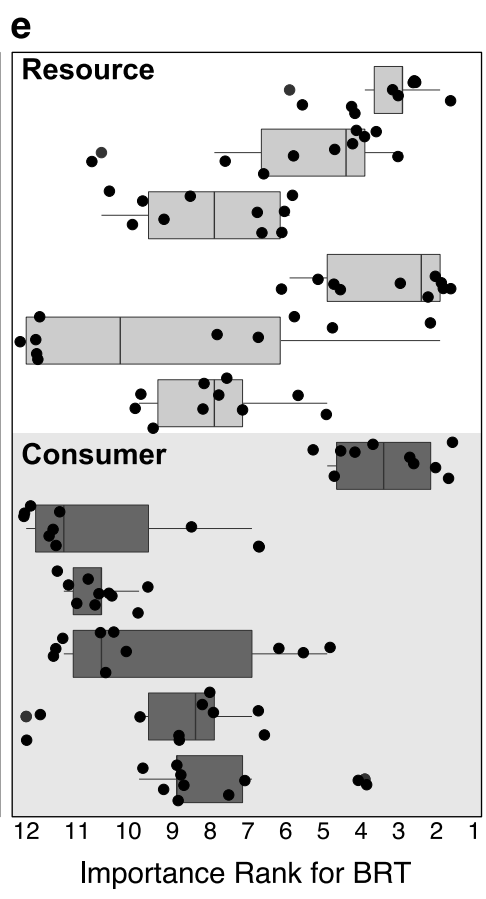
Implementation





d

Cluster of predictors	RF	BRT	KNN	NN	GLM	BGLM
1	4.9	2.4	2.9	3.2	6.2	2.8
2	6	4.6	5.5	4.7	5.5	4.8
3	8	7.1	7.8	7.9	3.8	6.2
4	9.6	2.4	2.5	2.6	5	2.7
5	4.9	8.3	7.4	8.1	4.1	9.2
6	7.2	7	6.3	5.8	4.9	6.1
7	2.3	2.5	1.8	1.8	7.1	1.5
8	9.6	10	9.8	8.7	9.2	8.7
9	8.9	10.1	10.3	9.4	6.9	9.8
10	3.6	8.6	8.5	9.7	8.3	9.5
11	5.9	8.2	8.5	9.1	6.1	8.1
12	7.1	6.8	6.7	7	8	8.6



f

