MethodsX 16 (2026) 103988

Contents lists available at ScienceDirect

MethodsX

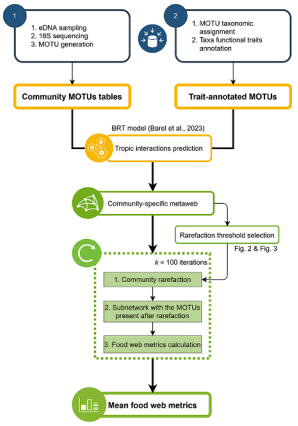
journal homepage: www.elsevier.com/locate/methodsx

Standardizing food web reconstruction from metabarcoding data using iterative rarefaction and trait-based trophic inference model

Víctor Lecegui [a,b,*], Vincent E.J. Jassey [c], Aaron Pérez-Haase [a,b]

[a] Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals (BEECA), Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Spain
[b] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain
[c] Université de Toulouse, Toulouse INP, CNRS, IRD, CRBE, Toulouse, France

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Ecological networks
Sequencing-depth bias
Protists
18S rRNA
Wetlands
Environmental gradients
Machine learning

ABSTRACT

High-throughput metabarcoding combined with trait-based trophic inference models enables the reconstruction of food webs from community eDNA. However, uneven sequencing depth can bias inferred food web structure. Network metrics derived from unstandardized read counts may reflect sampling effort rather than ecological processes. Here, we present an objective and reproducible framework using iterative rarefaction with trait-based trophic inference to standardize food web metrics prior to ecological analysis.

* Corresponding author at: Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain.
E-mail address: lecegui@ub.edu (V. Lecegui).

https://doi.org/10.1016/j.mex.2026.103988
Received 14 April 2026; Accepted 2 June 2026
Available online 3 June 2026
2215-0161/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Our approach links rarefaction depth to network-level properties by generating rarefaction curves based on food web metrics rather than taxonomic richness. An optimal rarefaction threshold is determined using a first derivative criterion, and rarefied networks are reconstructed repeatedly to obtain robust mean metric estimates.

We applied this method to soil protist communities assessed using 18S rRNA metabarcoding. The method reduced sequencing-depth bias and revealed environmental drivers that were masked in non-rarefied analyses.

This framework provides a standardized solution for comparing food webs derived from amplicon datasets across environmental gradients.

Integrates trait-based trophic inference with iterative rarefaction of metabarcoding communities.

Selects rarefaction thresholds using food web network properties rather than species accumulation curve.

Provides reliable estimates of food web structure independent of sequencing effort.

Specifications table

Subject area	Bioinformatics
More specific subject area	Microbial food webs
Name of your method	Pipeline for reconstructing food webs from metabarcoding data
Name and reference of original method	[1]
Resource availability	https://github.com/VictorLecegui/Lecegui2026_MethodsX

Background

Trait-matching models use functional traits of consumers and resources to predict predator–prey interactions, providing a powerful framework for reconstructing food webs [2,3]. When combined with high throughput metabarcoding, these approaches enable the derivation of food web networks directly from eDNA [4]. However, recent work emphasizes that constructing food webs from metabarcoding data requires careful attention to model assumptions, trait completeness, and bioinformatic pre-processing [5]. Additionally, the use of metabarcoding data exposes trait-matching models to a major methodological bottleneck, as uneven sequencing depth strongly biases community composition and, by extension, inferred network properties [6,7].

To solve this gap, rarefaction emerges as the best statistical method to control for uneven sequencing effort in amplicon data [8]. Originally developed for ecological diversity analyses and now widely used in amplicon data, enables unbiased sample comparison by standardizing sequencing depth through random subsampling [9]. To date, this is the only method that has been used for building networks from metabarcoding data, rarefying the community to the lowest sequencing depth [4]. However, this approach is not exempt from limitations. Food web properties (e.g., number of links and modularity) are highly sensitive to the gain or loss of specific taxa as their role in the food web can be very variable [10,11]. Hence, rarefying not to the lowest sequencing depth but to a rarefaction threshold in which the community composition remains stable, with no changes by increasing the sample size, seems more appropriate in a food web perspective. Moreover, networks constructed from a single rarefaction can incorporate or omit taxa randomly, repeated rarefactions would therefore provide a more statistically acceptable technique [9].

This study introduces a method that explicitly links and balances rarefaction depth with network-level metrics. We address existing methodological gaps by integrating a revised metabarcoding pipeline with a trait-based trophic inference model [1] to reconstruct microbial food webs. By utilizing an iterative rarefaction procedure around food web metrics, our framework offers a transparent solution for constructing comparable food webs from high-throughput sequencing data. The aim of this method is to reconstruct food web networks and discern ecological patterns from amplicon data while mitigating the biases introduced by uneven sequencing depth across samples.

To this end, we focussed on soil protists to address this methodological challenge. Protists underpin critical ecosystem processes, such as decomposition, nutrient cycling, and CO2 fixation [12–14], and are involved in myriads of trophic interactions [15]. Despite their important role in ecosystems, the structure of protist-based food webs remains largely under-characterized [12]. Here, we demonstrate the utility of our method by applying iterated rarefaction, calibrated to network stability, to food webs reconstructed from wetland soil protist metabarcoding data.

Method details

Metabarcoding data

The proposed framework is designed for metabarcoding datasets intended to characterise and compare food web networks. While the method assumes structured sampling design to accurately capture ecological patterns, it remains agnostic to ecosystem type,

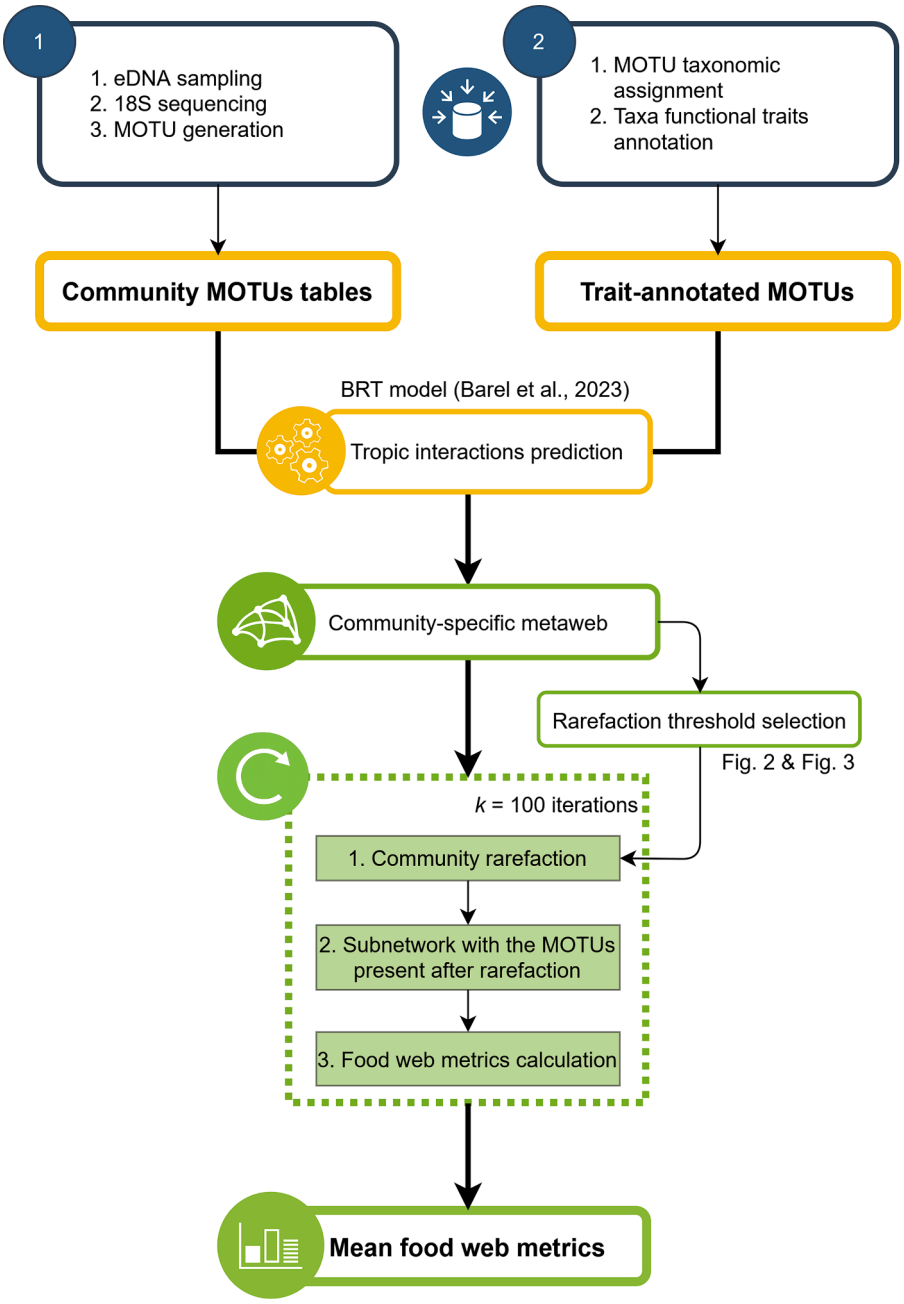


Fig. 1. Workflow for rarefaction-based reconstruction and evaluation of community food webs from eDNA metabarcoding data. The workflow is organized into three color-coded stages. Blue boxes represent data acquisition and annotation steps. Environmental DNA (eDNA) samples were processed through 18S rRNA sequencing to generate Molecular Operational Taxonomic Units (MOTUs) for each community. MOTUs were subsequently taxonomically assigned and annotated with functional traits. Yellow boxes represent trophic interaction inference. A Boosted Regression Tree (BRT) model was applied to the trait and compositional information of MOTUs to predict trophic interactions and reconstruct a community-specific metaweb for each community. Green boxes represent the iterative rarefaction framework proposed in this study. Prior to analysis, an optimal rarefaction threshold was determined by repeating the workflow across multiple sequencing-depth thresholds (see Figs. 2 and 3). For each community and rarefaction threshold, the procedure was repeated $k = 100$ times. In each iteration: (1) the community matrix was rarefied, (2) the induced subnetwork containing MOTUs retained after rarefaction was extracted from the community-specific metaweb, and (3) food web metrics were calculated. Final food web descriptors correspond to the mean metric values across rarefaction iterations, providing standardized and robust estimates of community food web structure. The methodological focus of this framework is the standardization of community-specific metawebs inferred from metabarcoding data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

taxonomic group, or genetic marker.

In the present study, we applied this workflow to an empirical dataset obtained from replicated soil samples collected across multiple wetlands in the Iberian Peninsula, spanning environmental gradients. We followed standardized metabarcoding protocols described in Guardiola et al., [16], to the samples obtained from the field. We extracted their DNA, amplified the V7 region of the 18S rRNA amplicon and sequenced it with Next Generation Sequencing (more details in S1). Environmental DNA (eDNA) metabarcoding employing the 18S rRNA is the standard methodological approach for documenting protist communities across diverse habitats [7,17]. Simultaneously, the rapid expansion of functional trait databases has enabled the integration of trait-based ecology with metabarcoding-derived taxon lists, facilitating the estimation of trophic interactions [6,18,19].

The bioinformatic pipeline involved clustering and denoising to identify molecular units that represent biological taxa (i.e. MOTUs) (see details in S1). These molecular units were then assigned to taxa using a customized reference database. This is a crucial step for building food web networks, as these networks will be built using the functional traits associated with the taxa [20]. Thus, taxa should be filtered carefully to exclude contaminants and taxa that may not be part of the free-living food web, such as parasites, whose free-living status cannot be resolved from metabarcoding data. We also excluded miscellaneous fungi and macroscopic organisms.

The resulting database comprises 5804 MOTUs assigned to 646 taxa (233 microalgae, 136 ciliates, 78 nematodes, 77 naked amoebae, 65 flagellates, 51 testate amoebae, 10 rotifers and 6 fungi) and 4807,994 reads distributed in 448 samples.

Functional traits and trophic interactions assignment

To infer trophic interactions among MOTUs, we used the approach described in [1]. To ensure consistency, we selected the same set of traits: type of organism, nourishment, feeding group, lifestyle, feeding strategy, morphology, locomotion, movement type, body width, body length, and biovolume (trait details and categories in [1] Table S2). We built a trait database for our taxa list. For the shared taxa between our study and [1] we used their trait values, and for the rest of taxa we looked for them into several specialized databases and literature (e.g. [6]). For MOTUs classified at higher taxonomic levels (e.g. genus - *Stentor*) we assigned categorical traits using the most common category and quantitative traits using the mean value of the species belonging to that group (e.g. mean values of *Stentor muelleri* and *Stentor amethystinus*). Then, we assessed the traits of the MOTUs according to the taxa to which they were assigned.

Finally, for all communities we run the Boosted Regression Tree (BRT) model, following [1], to our dataset. We identified all potential trophic interactions between the MOTUs present at each community and thereby reconstructed the corresponding community-specific metaweb.

Food web metrics

To characterize the topological structure and emergent properties of the food webs, we selected a set of food web metrics used as quantitative descriptors of different food web properties: 1) number of nodes, 2) number of links, 3) link density, 4) mean degree, 5) mean generality, 6) mean vulnerability, 7) mean trophic level, 8) fraction of basal MOTUs, 9) fraction of top MOTUs [20–22] (full descriptions in S3).

Rarefaction framework

To minimize bias associated with unequal sequencing depth, we accounted for variation in read counts among samples prior to food web construction. Differences in sequencing effort can artificially inflate observed MOTU richness, as higher-depth samples are more likely to detect both common and rare taxa. Because food web metrics are highly sensitive to node presence, uncorrected variation in sequencing depth may distort estimates of connectance, linkage density, modularity, and nestedness by reflecting sampling intensity rather than ecological structure [23]. This bias is particularly pronounced for rare MOTUs, whose detection probability is strongly dependent on sequencing depth and whose inclusion or exclusion can substantially alter network topology. Controlling for sequencing depth is therefore necessary to ensure that differences in inferred food web architecture reflected ecological patterns rather than methodological artifacts.

To account for the stochasticity inherent in subsampling-based standardization, we calculated food web metrics within an iterative rarefaction framework analogous to that implemented in the *avgdist* approach for community dissimilarities [24]. For each iteration ($k = 100$), all communities were rarefied to a common sequencing depth to ensure comparability. Then, food webs were predicted using the BRT model and subsetted to include only the MOTUs retained in the rarefied communities. A complete set of food web metrics was finally calculated from each subsetted network. This procedure was repeated 100 times (k) to generate independent realizations of network structure under rarefaction. The final food web metrics for each community was defined as the mean value across all k iterations, thereby capturing solid network-level properties rather than single, stochastic outcomes (Fig. 1). We used the mean because metric values across iterations showed low variability and highly similar distributions, with mean and median values being nearly identical (Pearson correlation > 0.99), whereas in cases of skewed distributions we would recommend the use of the median instead.

To ensure that the randomly generated rarefied communities represented plausible realizations of the original data in terms of richness (number of MOTUs), an initial rarefaction at each target sequencing depth was performed. Then each iteration was tested and if the richness fell within the 95% confidence interval it was retained, thereby excluding low-probability subsampling outcomes and ensuring that food web metrics are derived from statistically consistent community realizations.

Potential rarefaction thresholds

To determine an optimal rarefaction threshold, we performed rarefaction across a range of sequencing depths and constructed rarefaction curves. Unlike traditional approaches that depict species accumulation on the y-axis, our curves displayed the food web metrics to assess how network properties responded to changes in sequencing depth.

Selecting the rarefaction threshold represents a critical trade-off between retaining the maximum number of samples and preserving the richness of MOTUs within each sample. Higher thresholds provide community structures that better reflect the underlying reality; however, they result in the exclusion of samples with read counts below the chosen threshold, thereby reducing the overall sample representation.

Calculating food web metrics across numerous iterations (i.e., 100) and sequencing depths can be computationally demanding. Therefore, we conducted a preliminary analysis to identify a range of potential rarefaction thresholds. These potential rarefaction thresholds were selected by doing rarefactions at different cut-off values and we evaluated: (i) the proportion of retained plots relative to the original dataset, (ii) the mean proportion of MOTUs retained per plot, and (iii) the mean slope of the rarefaction curves.

To ensure comparability across thresholds in subsequent analyses, the same set of samples was used; therefore, this exploratory assessment was conducted using samples with read counts higher than the maximum value of reads within the tested threshold range. In our study, we did this test by rarefying from 500 to 8000 reads every 500 and we decided to perform the food web metrics rarefaction curves from 500 to 4500 reads. This decision was made because over 4500 reads 50% of the samples were lost even though 80% of the MOTUs per sample were already retained. (Fig. 2). To ensure comparability, we selected the communities with >4500 reads to perform the exploration of the rarefaction threshold based on food web metrics.

Rarefaction curves based on food web metrics

For each community and candidate rarefaction threshold, the expected number of MOTUs (S) and its 95% confidence interval was first estimated using the *rarefy* function from the *vegan* R package. Rarefied communities were then generated using *rrarefy*. Each rarefied realization was evaluated by comparing its observed number of MOTUs to the confidence interval derived at the target rarefaction threshold, and only realizations whose richness fall within this interval was retained. If a realization falls outside the confidence interval, rarefaction is repeated until a statistically plausible subsample is obtained. This procedure was repeated until (k =

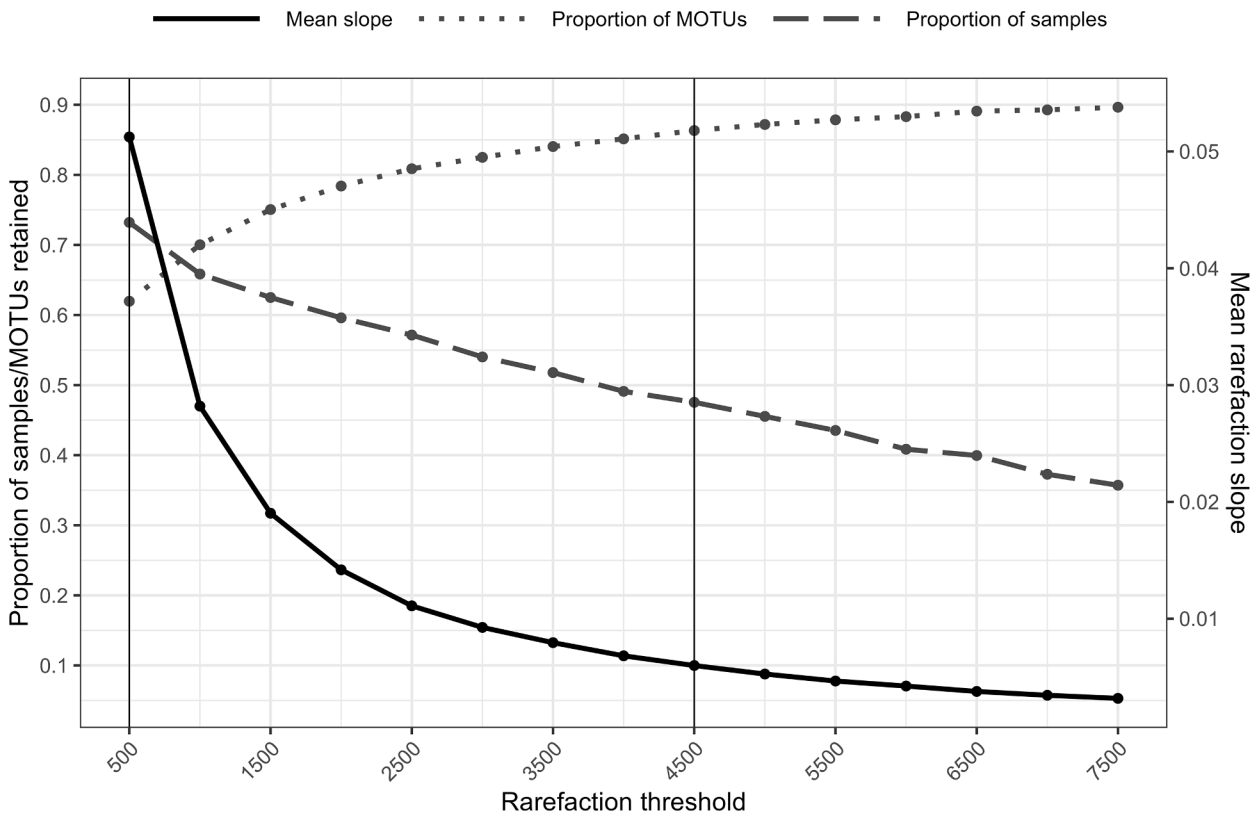


Fig. 2. Effect of rarefaction threshold on data retention. The dotted grey line represents the mean proportion of retained MOTUs across samples, while the dashed grey line represents the proportion of samples retained after rarefaction. The black line is the mean of the richness rarefaction slopes. The vertical lines indicate candidate rarefaction thresholds from 500 to 4500 reads.

100) valid rarefied communities were obtained for each community and threshold.

For each accepted realization, the corresponding community metaweb was subsetted to include only the MOTUs present in the rarefied community, and the set of food web metrics was computed from the resulting network. Final values for each metric were obtained by averaging across the k iterations, and these mean values were used as the food web metrics at the corresponding sequencing depth. This process was repeated for each candidate threshold and then the rarefaction curves based on food web metrics were drawn.

To determine the optimal threshold we used the Dynamical First Derivate Threshold (DFDT) [25], which selects the knee or elbow point of a curve by partitioning it into regions of high and low first-derivative (slope) values. The DFDT was calculated separately for each rarefaction curve (i.e., each community) and for each food web metric. This produced one threshold value per community and metric combination. For each metric, we then calculated the mean threshold across all communities. Finally, to obtain a single rarefaction threshold applicable to all communities and metrics, we selected the highest of these metric-specific mean thresholds. This

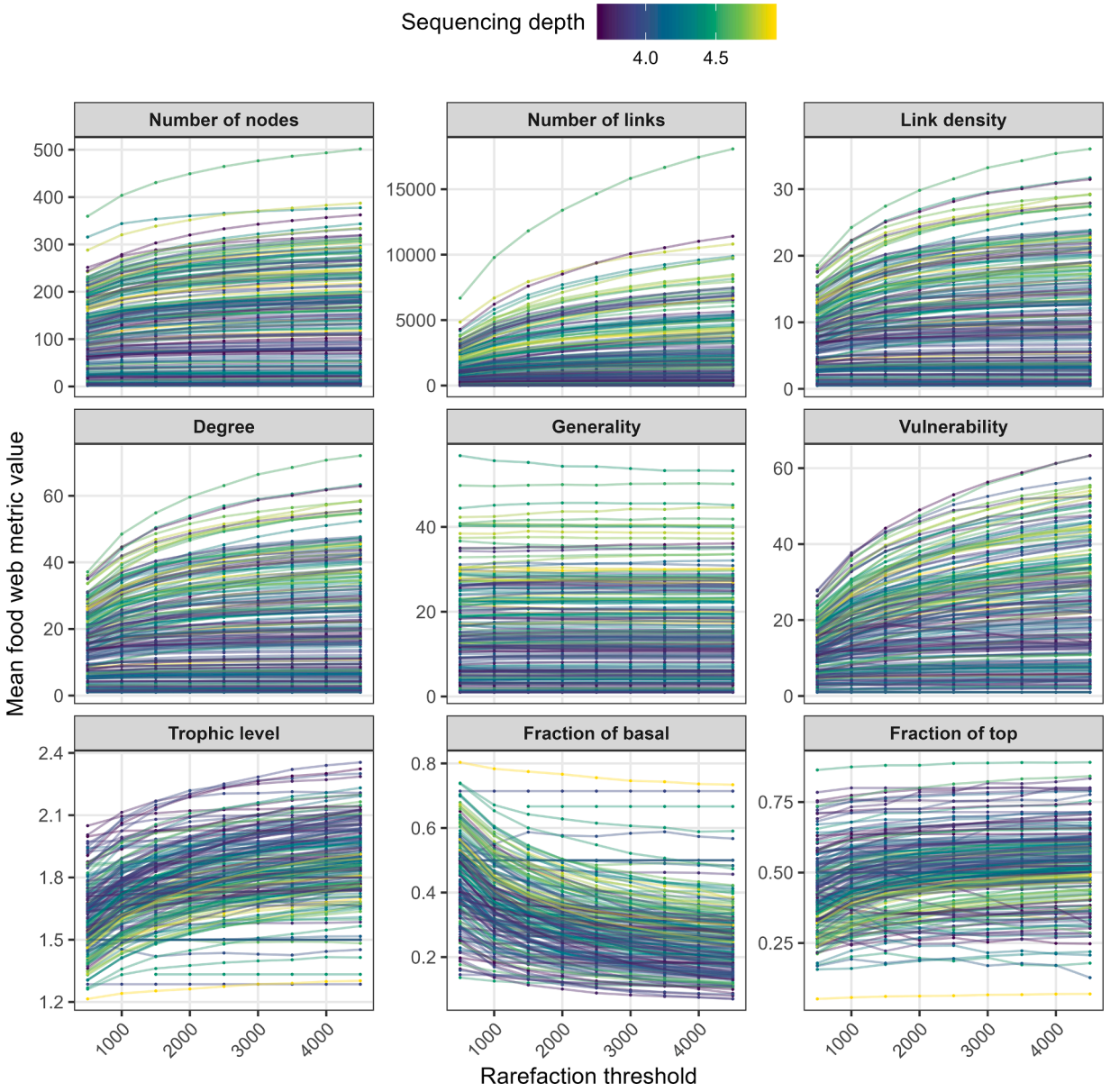


Fig. 3. Mean values of nine food-webs metrics across rarefaction thresholds for soil protist communities. Each panel shows the mean metric value after the iterative method (*y*-axis; Number of nodes, Number of links, Link density, Degree, Generality, Vulnerability, Trophic level, Faction of basal, Fraction of top) plotted against rarefaction threshold (*x*-axis; read count per sample) for individual networks. Colors indicate the log10-transformed original sequencing depth of each sample prior to rarefaction (for better visualization). Curves show the stability of the metrics at high rarefaction thresholds. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

conservative approach ensured that the beyond the final threshold most of the communities remained stable for all metrics.

In our study, we followed this pipeline and obtained rarefaction curves for our communities for each food web metric (Fig. 3). Applying the DFDT procedure to these rarefaction curves yielded a threshold value for each metric. Among them, generality produced the highest mean threshold across communities and was therefore selected as the final rarefaction threshold according to our conservative criterion. This resulted in an optimal threshold at 2348 reads (Fig. 4). Finally, to support this choice, we calculated the Good's coverage *C*, a sample completeness estimator used in rarefaction to estimate how much diversity was captured for the remaining samples:

$$C = \frac{1 - F_1}{N}$$

Where F_1 is the number of singletons and N the total number of reads. Thresholds are considered acceptable when samples coverage

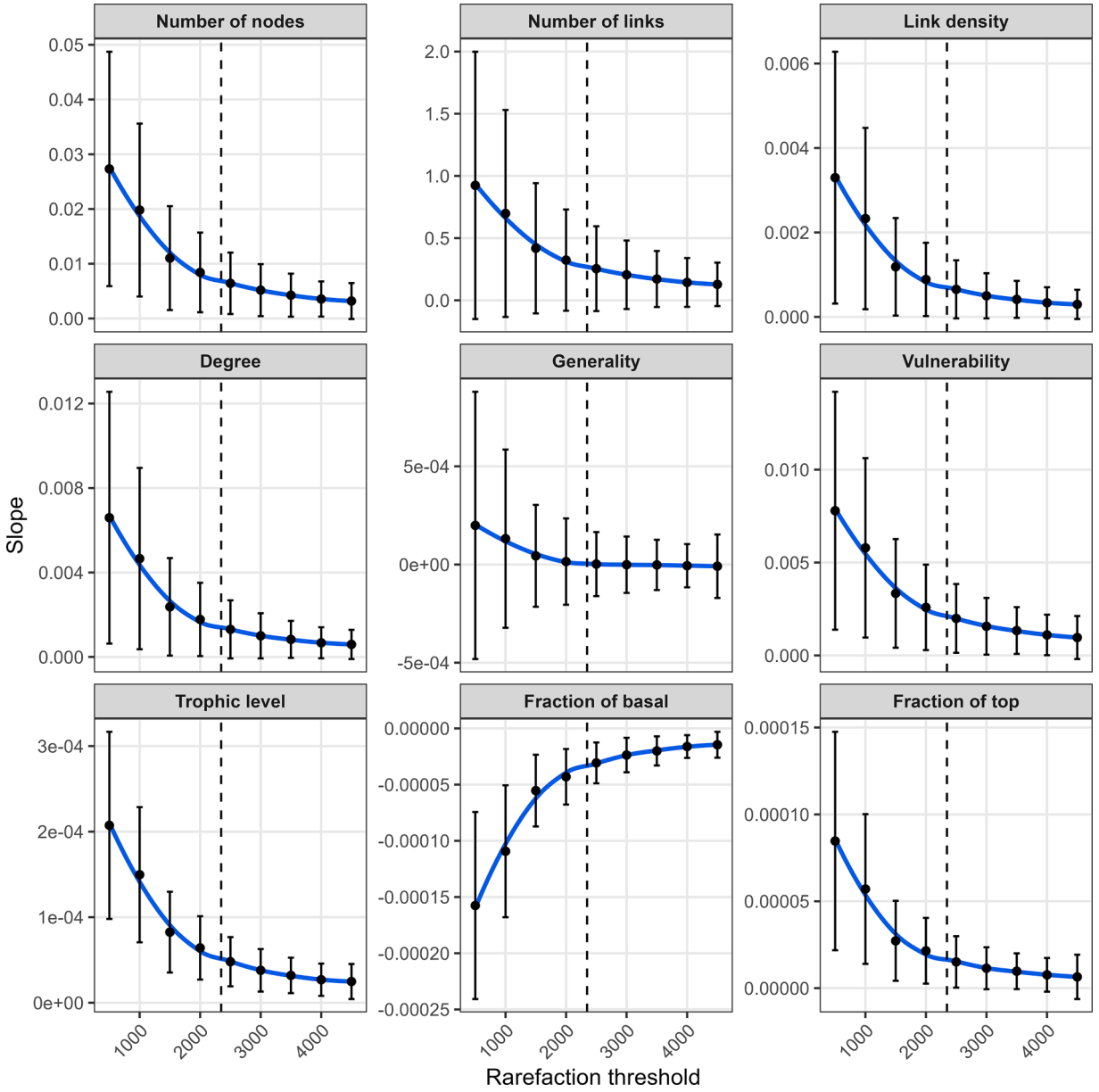


Fig. 4. Slopes of rarefaction curves for food web metrics across rarefaction thresholds. Points represent the mean slope values (first derivative) of the rarefaction curves shown in Fig. 3, and error bars indicate standard deviations across communities. The DFDT was applied to each metric to identify the rarefaction threshold at which the metric stabilized. The dashed vertical line at 2348 indicates the most conservative (highest) DFDT, the threshold applied to all food web metrics. Solid blue loess lines (span = 0.75) illustrate the stabilization of slopes beyond the selected threshold.

values fall between 0.95 and 0.99, indicating that approximately 95–99% of the community's diversity has been sampled and undersampling bias is minimal. The samples over 2348 reads achieved C $\geq$ 0.97, confirming adequate sequencing completeness.

With this pipeline we established an objective and reproducible way to select the rarefaction threshold for rarefaction of food webs.

Final rarefactions

Once the rarefaction threshold was decided, communities without predicted feeding interactions and above the rarefaction threshold were filtered. Then, the iterative rarefaction procedure was applied, standardizing all communities at the same sequencing depth and calculating the mean value of the food web metrics.

In our study, we discarded 19 communities for which no trophic interactions were found and we kept all communities with sequencing depths exceeding 2348 reads (n = 238 communities). We performed rarefaction across 100 iterations per community, calculating mean values for the 9 food web metrics and used them as descriptors of network structure.

Consequently, these mean metrics now serve as reliable proxies for food web structure, enabling valid statistical comparisons across sites. This approach facilitates robust ecological conclusions, particularly when examining how network structure varies along environmental gradients.

Method validation

We empirically evaluated the method by performing the entire analytical procedure without rarefying the data and assessing the influence of sequencing depth on the results. To enable comparison with rarefied data, we focused on communities containing >2348 reads and directly calculated food web metrics for these communities. We then examined the correlation between each food web metric and sequencing depth (Table 1), identifying generality, vulnerability, link density, degree, number of links, and number of nodes as the metrics exhibiting the strongest correlations (correlation coefficient > 0.35).

To assess the influence of sequencing depth on the overall combination of food web metrics and its potential effect on the interpretation of results, we performed a redundancy analysis (RDA). As potential explanatory variables describing the food web structure across sampled wetlands, we included water table depth, pH, electric conductivity, vegetation height, vascular plant cover, community-type and diversity, bryophyte cover, community-type and diversity, air mean annual temperature, mean annual precipitation, annual potential evapotranspiration and disturbance degree. In addition, we incorporated the number of reads as a potential explanatory variable. We applied a bidirectional stepwise selection procedure using the ordistep function from the vegan package, with food web metrics as response variables and environmental factors along with sequencing depth as explanatory variables. The analysis revealed that the number of reads was among the variables selected by the model and remained statistically significant after conducting an ANOVA on the final RDA model.

We rarefied the previously analyzed communities using our proposed method and subsequently performed a redundancy analysis (RDA) with bidirectional variable selection. Comparison of the resulting models revealed a hiding effect in variable selection by the number of reads. Three variables were retained in the non-rarefied model, whereas seven (including the ones in the non rarefaction model) variables were selected in the rarefied model (Table 2).

These results highlight the critical importance of accounting for sequencing depth when computing food web metrics from metabarcoding data. Failure to control for the number of reads can introduce systematic bias into network properties, which in turn may generate spurious correlations with environmental variables and obscure genuine ecological signals. Consequently, removing or statistically correcting for the effect of sequencing depth is essential to avoid misleading interpretations of environmental drivers and to prevent drawing erroneous conclusions about the underlying ecological processes and patterns.

Although the framework presented here was specifically developed and validated for metabarcoding-derived food webs, but the general principles underlying the approach are transferable to other ecological applications affected by uneven sampling effort or incomplete standardization. However, such generalization should be evaluated cautiously and would require dedicated validation in each specific context. We therefore encourage future studies to explore the applicability of these ideas across other disciplines beyond metabarcoding-derived food webs.

Table 1
Spearman's rank correlation between the food web metrics and sequencing depth.

Food web metric	Correlation coefficient
Generality	0.418
Vulnerability	0.411
Link density	0.404
Degree	0.404
Number of links	0.396
Number of links	0.386
Top fraction	−0.049
Trophic level	−0.039
Basal fraction	−0.035

Table 2
p-values from ANOVA-like permutation tests (999 permutations) on RDA marginal effects, comparing environmental variables explaining protist trophic networks structure. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; NS, not significant ($p \geq 0.05$). Rarefaction applied iterative threshold selection; variables excluded from RDA if non-significant pre-selection.

Variable	NOT Rarefied	Rarefied
Number of reads	0.001 ***	Not included
Water table depth	0.001 ***	0.001 ***
Vegetation height	0.007 **	0.005 **
Bryophyte cover	0.005 **	0.001 ***
Vascular plant cover	NS	0.001 ***
Vascular plant community-type	NS	0.003 **
Annual precipitation	NS	0.008 **
Annual potential evapotranspiration	NS	0.018 *

Limitations

One of the main limitations of inferring trophic interactions from metabarcoding data is that trait-matching algorithms depend critically on the taxonomic assignment of MOTUs, which is itself subject to several sources of bias. First, some taxonomic groups remain underrepresented in reference databases, particularly rare or poorly studied lineages, leading to higher rates of misassignment or non-assignment [20]. Second, the choice of amplicon introduces additional bias; for example, the 18S rSSU marker used here amplifies a broad spectrum of eukaryotic taxa, whereas more taxon-specific markers may overrepresent lineages and artificially inflate their importance in the inferred food web. Finally, different reference databases can yield divergent taxonomic labels for the same MOTU, making the selection of the reference database a critical decision that directly influences downstream network inference [26].

Moreover, our method may be particularly useful for reconstructing food webs from metabarcoding data. Its application relies on two essential components: a comprehensive trait database and a trait-based inference model. Trait information is still incomplete for many taxa, and although it may often be found in species descriptions, standardized trait databases are crucial for making it usable. Likewise, developing a robust inference model requires documented trophic interactions, typically gathered through literature searches or direct observations, and translating these data into a formal model remains a significant challenge.

Related research article

None.

Ethics statements

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used occasionally PerplexityAI Pro (model Sonar 2) among other conventional sources in order to identify relevant bibliography to fill the trait matrix, and ChatGPT (model 5.5) to improve the clarity and correctness of the English language. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the published article.

CRediT authorship contribution statement

Víctor Lecegui: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft. **Vincent E.J. Jassey:** Conceptualization, Methodology, Writing – review & editing. **Aaron Pérez-Haase:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Organismo Autónomo de Parques Nacionales (OAPN) [BIOOCULT (2413/2017)] and the LIFE Programme of the European Union [LIFE—Resque Alpyr LIFE20 NAT/ES/000369]. We gratefully acknowledge the staff and collaborators of the National Parks of Aigüestortes i Estany de Sant Maurici, Ordesa y Monte Perdid, Picos de Europa, Sierra Nevada, and Sierra de Guadarrama for their assistance with field sampling.

We also thank all individuals who contributed to field sampling, laboratory assistance, and the bioinformatic processing of the data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2026.103988.

Data availability

Data will be made available on request.

References

[1] J.M. Barel, O.L. Petchey, A. Ghaffouli, V.E.J. Jassey, Uncovering microbial food webs using machine learning, Soil Biol. Biochem. 186 (2023) 109174, https://doi.org/10.1016/j.soilbio.2023.109174.

[2] D. Caron, U. Brose, M. Lurgi, F.G. Blanchet, D. Gravel, L.J. Pollock, Trait-matching models predict pairwise interactions across regions, not food web properties, Glob. Ecol. Biogeogr. 33 (4) (2024), https://doi.org/10.1111/geb.13807.

[3] J.P.F. Pomeranz, R.M. Thompson, T. Poisot, J.S. Harding, Inferring predator–prey interactions in food webs, Methods Ecol. Evol. 10 (3) (2019) 356–367, https://doi.org/10.1111/2041-210X.13125.

[4] M. Seymour, F.K. Edwards, B.J. Cosby, M.G. Kelly, M. de Bruyn, G.R. Carvalho, S. Creer, Executing multi-taxa eDNA ecological assessment via traditional metrics and interactive networks, Sci. Total Environ. 729 (2020) 138801, https://doi.org/10.1016/j.scitotenv.2020.138801.

[5] C.L. Pereira, Z. Ersoy, M.P. Thomas Gilbert, D. Gravel, M.B. Araújo, M.G. Matias, Future-proofing environmental DNA and trait-based predictions of food webs, BioScience 73 (12) (2023) 862–878, https://doi.org/10.1093/biosci/biad089.

[6] S. Giachello, I. Cantera, A. Carteron, S. Marta, C. Cipriano, A. Guerrieri, A. Bonin, W. Thuiller, G.F. Ficetola, Toward a common set of functional traits for soil protists, Soil Biol. Biochem. 187 (2023) 109207, https://doi.org/10.1016/j.soilbio.2023.109207.

[7] R.L. Mau, M. Hayer, A.M. Purcell, S. Geisen, B.A. Hungate, E. Schwartz, Measurements of soil protist richness and community composition are influencedby primer pair, annealing temperature, and bioinformatics choices, Appl. Environ. Microbiol. 90 (7) (2024), https://doi.org/10.1128/aem.00800-24.

[8] P.D. Schloss, Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses, mSphere 9 (2) (2024), https://doi.org/10.1128/msphere.00354-23.

[9] E.S. Cameron, P.J. Schmidt, B.J.M. Tremblay, M.B. Emelko, K.M. Müller, Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities, Sci. Rep. 11 (1) (2021), https://doi.org/10.1038/s41598-021-01636-1.

[10] M. Lurgi, N. Galiana, B.C. López, L.N. Joppa, J.M. Montoya, Network complexity and species traits mediate the effects of biological invasions on dynamic food webs, Front. Ecol. Evol. 2 (JUL) (2014), https://doi.org/10.3389/fevo.2014.00036.

[11] M. Reji Chacko, C. Albouy, F. Altermatt, V. Boussange, M. Brändle, N. Farwig, M.M. Gossner, H.C. Ho, A. Joss, F. Neff, L. Pellissier, Species loss in key habitats accelerates regional food web disruption, Commun. Biol. 8 (1) (2025), https://doi.org/10.1038/s42003-025-08396-y.

[12] S. Geisen, E.A.D. Mitchell, S. Adl, M. Bonkowski, M. Dunthorn, F. Ekelund, L.D. Fernández, A. Jousset, V. Krashevska, D. Singer, F.W. Spiegel, J. Walochnik, E. Lara, Soil protists: a fertile frontier in soil biology research, FEMS Microbiol. Rev. 42 (3) (2018) 293–323, https://doi.org/10.1093/femsre/fuy006.

[13] V.E.J. Jassey, S. Hamard, C. Lepère, R. Céréghino, B. Corbara, M. Küttim, J. Leflaive, C. Leroy, J.-F. Carrias, Photosynthetic microorganisms effectively contribute to bryophyte CO2 fixation in boreal and tropical regions, ISME Commun. 2 (1) (2022), https://doi.org/10.1038/s43705-022-00149-w.

[14] F. Maillard, F. Klinghammer, V.E.J. Jassey, B. Zhang, P.G. Kennedy, E. Lara, S. Geisen, L. Tranvik, E. Hammer, A. Tunlid, Hidden decomposers: revisiting saprotrophy among soil protists and its potential impact on carbon cycling, Soil Biol. Biochem. 205 (2025) 109786, https://doi.org/10.1016/j.soilbio.2025.109786.

[15] V.E.J. Jassey, O.L. Petchey, P. Binet, A. Buttler, G. Chiapusio, F. Delarue, F. Laggoun-Défarge, D. Gilbert, E.A.D. Mitchell, J.M. Barel, Food web structure and energy flux dynamics, but not taxonomic richness, influence microbial ecosystem functions in a Sphagnum-dominated peatland, Eur. J. Soil. Biol. 118 (2023) 103532, https://doi.org/10.1016/j.ejsobi.2023.103532.

[16] O.S. Wangensteen, C. Palacín, M. Guardiola, X. Turon, DNA metabarcoding of littoral hardbottom communities: High diversity and database gaps revealed by two molecular markers, PeerJ 2018 (5) (2018), https://doi.org/10.7717/peerj.4705.

[17] F. Burki, M.M. Sandin, M. Jamy, Diversity and ecology of protists revealed by metabarcoding, Curr. Biol. 31 (19) (2021) R1267–R1280, https://doi.org/10.1016/j.cub.2021.07.066.

[18] M. Jamy, P. Ramond, D. Bass, J. del Campo, M. Dunthorn, E. Lara, A. Mitra, D. Vaulot, L. Santoferrara, Towards a trait-based framework for protist ecology and evolution, Trends Microbiol. (2026), https://doi.org/10.1016/j.tim.2025.08.008.

[19] G. Lentendu, D. Singer, S. Agatha, M. Bahram, S.E. Hannula, J. Helder, L. Tedersoo, W. Traunspurger, S. Geisen, E. Lara, EukFunc: a holistic eukaryotic functional reference for automated profiling of soil eukaryotes, Mol. Ecol. Resour. 25 (2025), https://doi.org/10.1111/1755-0998.14118.

[20] F. Keck, M. Couton, F. Altermatt, Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses, Mol. Ecol. Resour. 23 (4) (2023) 742–755, https://doi.org/10.1111/1755-0998.13746.

[21] S. Kortsch, R. Frelat, L. Pecuchet, P. Olivier, I. Putnis, E. Bonsdorff, H. Ojaveer, I. Jurgensone, S. Strāķe, G. Rubene, Ē. Krüze, M.C. Nordström, Disentangling temporal food web dynamics facilitates understanding of ecosystem functioning, J. Anim. Ecol. 90 (5) (2021) 1205–1216, https://doi.org/10.1111/1365-2656.13447.

[22] S. Kortsch, R. Primicerio, M. Aschan, S. Lind, A.V. Dolgov, B. Planque, Food-web structure varies along environmental gradients in a high-latitude marine ecosystem, Ecography 42 (2) (2019) 295–308, https://doi.org/10.1111/ecog.03443.

[23] B. Gauzens, S. Legendre, X. Lazzaro, G. Lacroix, Food-web aggregation, methodological and functional issues, Oikos 122 (11) (2013) 1606–1615, https://doi.org/10.1111/j.1600-0706.2013.00266.x.

[24] J. Oksanen, G.L. Simpson, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, P. Solymos, M.H.H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, T. Borman, G. Carvalho, M. Chirico, M. De Caceres, J. Weedon, vegan: Community Ecology Package, 2026. https://vegandevs.github.io/vegan/.

[25] M. Antunes, D. Gomes, R.L. Aguiar, Knee/elbow estimation based on first derivative threshold, in: Proceedings - IEEE 4th International Conference on Big Data Computing Service and Applications, BigDataService 2018, 2018, pp. 237–240, https://doi.org/10.1109/BigDataService.2018.00042.

[26] F. Mugnai, F. Costantini, A. Chenuil, M. Leduc, J.M.G. Ortega, E. Meglécz, Be positive: customized reference databases and new, local barcodes balance false taxonomic assignments in metabarcoding studies, PeerJ 11 (2023) e14616, https://doi.org/10.7717/peerj.14616.