Received: 29 June 2022 | Revised: 22 November 2022 | Accepted: 29 November 2022

DOI: 10.1002/ecy.3974

ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Sensitivity of bipartite network analyses to incomplete sampling and taxonomic uncertainty

Cristina Llopis-Belenguer[1,2] | **Juan Antonio Balbuena**[3] | **Isabel Blasco-Costa**[4,5] | **Anssi Karvonen**[6] | **Volodimir Sarabeev**[7,8] | **Jukka Jokela**[1,2]

[1]Institute of Integrative Biology, D-USYS, ETH Zürich, Zürich, Switzerland

[2]Department of Aquatic Ecology, EAWAG, Dübendorf, Switzerland

[3]Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain

[4]Department of Invertebrates, Natural History Museum of Geneva, Geneva, Switzerland

[5]Department of Arctic and Marine Biology, UiT The Arctic University of Norway, Tromsø, Norway

[6]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

[7]Department of Biology, Zaporizhzhia National University, Zaporizhzhia, Ukraine

[8]Institute of Parasitology, Slovak Academy of Sciences, Košice, Slovak Republic

Correspondence
Cristina Llopis-Belenguer
Email: cristina.llopis.belenguer@gmail.com

Funding information
Eidgenössische Technische Hochschule Zürich, Grant/Award Number: ETH Postdoctoral Fellowship (20-2 FEL-67); Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A_169211; Slovakian Government for Excellent Researchers Threatened by the War Conflict in Ukraine, Grant/Award Number: 09I03-03-V01-00017

Handling Editor: Nicholas J. Gotelli

Abstract

Bipartite network analysis is a powerful tool to study the processes structuring interactions in ecological communities. In applying the method, it is assumed that the sampled interactions provide an accurate representation of the actual community. However, acquiring a representative sample may be difficult as not all species are equally abundant or easily identifiable. Two potential sampling issues can compromise the conclusions of bipartite network analyses: failure to capture the full range of interactions (sampling completeness) and use of a taxonomic level higher than species to evaluate the network (taxonomic resolution). We asked how commonly used descriptors of bipartite antagonistic communities (modularity, nestedness, connectance, and specialization [H_2']) are affected by reduced host sampling completeness, parasite taxonomic resolution, and their crossed effect, as they are likely to co-occur. We used a quantitative niche model to generate weighted bipartite networks that resembled natural host–parasite communities. The descriptors were more sensitive to uncertainty in parasite taxonomic resolution than to host sampling completeness. When only 10% of parasite taxonomic resolution was retained, modularity and specialization decreased by ~76% and ~12%, respectively, and nestedness and connectance increased by ~114% and ~345% respectively. The loss of taxonomic resolution led to a wide range of possible communities, which made it difficult to predict its effects on a given network. With regards to host sampling completeness, standardized nestedness, connectance, and specialization were robust, whereas modularity was sensitive (~30% decrease). The combination of both sampling issues had an additive effect on modularity. In communities with low effort for both sampling issues (50%–10% of sampling completeness and taxonomic resolution), estimators of modularity, and nestedness could not be distinguished from those of random assemblages. Thus, the categorical description of communities with low sampling effort (e.g., if a community is modular or not) should be done with caution. We

recommend evaluating both sampling completeness and taxonomic certainty when conducting bipartite network analyses. Care should also be exercised when using nonrobust descriptors (the four descriptors for parasite taxonomic resolution; modularity for host sampling completeness) when sampling issues are likely to affect a dataset.

KEYWORDS
bipartite networks, host–parasite interactions, sampling completeness, sampling issues, taxonomic resolution

INTRODUCTION

Species interactions drive ecological and evolutionary processes. In ecological communities, species interactions are diverse, numerous, and often asymmetric due to the unequal dependence between the interacting species (Dormann et al., 2017). These attributes make ecological communities complex, hampering our ability to disentangle ecological and evolutionary dynamics and predict responses in a changing environment. In light of this situation, bipartite network analysis offers a simplified framework to address fundamental research questions and advise on biodiversity management (Dormann et al., 2017; Tylianakis et al., 2007). Bipartite network analysis assesses the distribution of interactions between nodes of different guilds (Strona, 2022). For example, in host–parasite bipartite networks, host and parasite species are nodes. Interactions in bipartite networks can be either unweighted (i.e., presence–absence) or weighted by their strength to give a more accurate description of the processes in the natural environment. For example, host–parasite interactions can be weighted by the mean abundance or prevalence of each parasite species in each host (Cardoso et al., 2021).

To accurately represent the interaction network of an ecological community, we need to record both species composition and interactions (i.e., sampling completeness) (Henriksen et al., 2019). However, representative samples of ecological communities can be difficult to obtain. Samples must be large enough to capture the whole species richness. We also need sufficient individuals of each species to identify a sufficient fraction of the species interaction pattern. Sampling should also consider spatiotemporal variations to capture species and interaction dynamics (Strona, 2022). If such sampling completion is not achieved, teasing apart biological processes from methodological artifacts becomes a difficult task. Sampling issues may, therefore, affect network properties and the conclusions extracted from them (Vizentin-Bugoni et al., 2016).

A sample is considered to be a good representative of the actual community when species richness, interaction richness, or network descriptors reach an asymptote (Henriksen et al., 2019). However, acquiring an asymptotic sample of the interactions in a community requires a higher sampling effort than estimating species richness because, usually, there are more combinations of pairwise interactions than species (Henriksen et al., 2019). In addition, common interactions between abundant species are detected with lower effort, but higher sampling effort is required to record rare interactions of less abundant species (Chacoff et al., 2012; Henriksen et al., 2019).

Inappropriate or heterogenic node resolution can also affect bipartite network analysis. Some community datasets may be affected by heterogenic taxonomic resolution if some nodes are identified as species, whereas other nodes aggregate higher taxonomic ranks. Other studies may suffer from low taxonomic resolution as nodes are homogenously lumped into high taxonomic ranks (Rodrigues & Boscolo, 2020; Thompson & Townsend, 2000). Species over-splitting, or allocating conspecific individuals to different species due to intraspecific variation, can also affect community ecology studies (Isaac et al., 2004). Furthermore, single nodes can involve different entities (e.g., detritus) or include hidden diversity (e.g., cryptic species), which implies that taxonomic uncertainty affects the study of these communities (Thompson & Townsend, 2000). Although node resolution usually decreases when community size increases, small networks may be more affected by the inappropriate resolution of a node as it represents a higher proportion of taxa than in a large network (Renaud et al., 2020). Also, variability in node resolution both within and between community datasets complicates comparative studies.

Low sampling completeness and taxonomic resolution occur frequently, possibly affecting most if not all ecological community datasets to some degree (Rodrigues & Boscolo, 2020; Vizentin-Bugoni et al., 2016). Hence, it is likely that both sampling issues simultaneously affect a dataset. As they are independent, researchers must control all potential sampling issues at the same time to ensure a correct representation of an ecosystem. For instance, a dataset may present poor sampling completeness regardless of the

taxonomic resolution of the few sampled taxa. Even though these sampling issues are known to affect bipartite network descriptors (Rivera-Hutinel et al., 2012; Rodrigues & Boscolo, 2020) and may co-occur in the same survey, we do not yet know how their crossed effect might mislead the interpretation of the structure of ecological communities.

Host–parasite associations are among the most common types of interactions in ecological communities (Lafferty et al., 2006), owing to the central role of parasite species richness, abundance, and biomass in many communities (Carlson et al., 2020). Hence, bipartite network analysis is key to investigating processes in host–parasite communities with many species and interactions (Runghen et al., 2021). Host individuals are typically the sampling units and, commonly, data from host individuals of the same species are pooled together to obtain the parasite community of each host species (Poulin, 2007) (Figure 1a,b). As more host individuals are sampled in a community, the probability of finding an unrecorded parasite species or a new host–parasite interaction reduces, and parasite species or interaction richness approach an asymptote (Henriksen et al., 2019). Here, we refer to host sampling completeness as the set of host individuals of the same species that is sampled to record the host–parasite interactions in that particular host species. At the same time, studies listing parasite species often suffer from poor taxonomic resolution at least for certain taxa. Such inaccurate assessments of parasite diversity seriously limit our understanding of host–parasite dynamics (Poulin & Presswell, 2022). Indeed, the loss of taxonomic resolution had a higher impact on the predicted structure of antagonistic rather than mutualistic insect–plant networks due to the stronger dependence of the consumers on their resources in antagonistic communities (Rodrigues & Boscolo, 2020).

Our goal was to understand how decreasing gradients of host sampling completeness, parasite taxonomic resolution, and their crossed effect affect four commonly used descriptors of host–parasite communities: modularity, nestedness, connectance and specialization (H_2') (Bellay et al., 2015) (Table 1). We generated replicates of a simulated host–parasite community and resampled the replicates to evaluate the effect of the sampling issues to different degrees. We based our simulations on fish–metazoan parasite communities. There is an increasing number of studies using them as a model system for antagonistic bipartite networks (references in Runghen et al., 2021) as they are usually species rich and comprise many trophic levels. However, the effect of sampling issues on these networks has not yet been evaluated.

First, we assessed the effect of decreasing host sampling completeness on the four descriptors by gradually reducing the number of host–parasite interactions, simulating a loss of host individuals for each host species while all the species are preserved. Second, we evaluated how the four descriptors were influenced by decreasing resolution of parasite identification. We gradually reduced the number of parasite species in the communities by hierarchically lumping the nodes and their interactions according to their overlap in host use. This simulated the aggregation of phylogenetically close species with ecologically similar requirements into higher homogeneous taxonomic ranks (i.e., species into genera, genera into families, etc.). In other words, it mimicked a homogeneous reduction in taxonomic resolution. We finally evaluated the crossed effect of the two sampling issues as they are likely to occur simultaneously. Our research extends former studies since (1) we assessed the effects of sampling issues on antagonistic networks, contrary to the majority of studies that evaluated mutualistic systems; (2) we used weighted interactions to realistically represent the actual ecological process, whereas unweighted data are more often used; (3) to our knowledge, earlier studies have not explicitly considered the crossed effect of two simultaneous sampling issues on ecological networks.

We hypothesized that the descriptors would be more robust to host sampling completeness than to parasite taxonomic resolution, at least in communities that are not severely affected by sampling issues - Hypothesis 1. Host individuals of the same species commonly sustain similar parasite communities and represent biological replicates of the same system (Llopis-Belenguer et al., 2020). When decreasing host sampling completeness, the network structure gradually changes until it no longer supports conclusions. We expected the reduction in parasite taxonomic resolution to have a greater impact on the community descriptors. Usually, parasite species are able to infect a limited number of host species (Llopis-Belenguer et al., 2020). Parasite specialization in their hosts drives host–parasite network structure (Krasnov et al., 2012). Therefore, the reduction in parasite taxonomic resolution suppresses parasite specialization, and they appear as generalist parasites, which affects the overall network structure. In addition, we expected the crossed effect of both sampling issues to cause an additive effect in the descriptors since former studies showed similar patterns for both sampling issues (Henriksen et al., 2019; Rodrigues & Boscolo, 2020) - Hypothesis 2.

We also hypothesized that the four network descriptors would differ in sensitivity - Hypothesis 3. Based on evidence from other ecological systems (Blüthgen et al., 2006; Vizentin-Bugoni et al., 2016), redundant interactions would make the descriptors considerably robust against reduced sampling completeness. Modularity, nestedness and specialization would be reasonably robust to loss of taxonomic resolution (Rodrigues & Boscolo, 2020; Thompson & Townsend, 2000). However, communities in

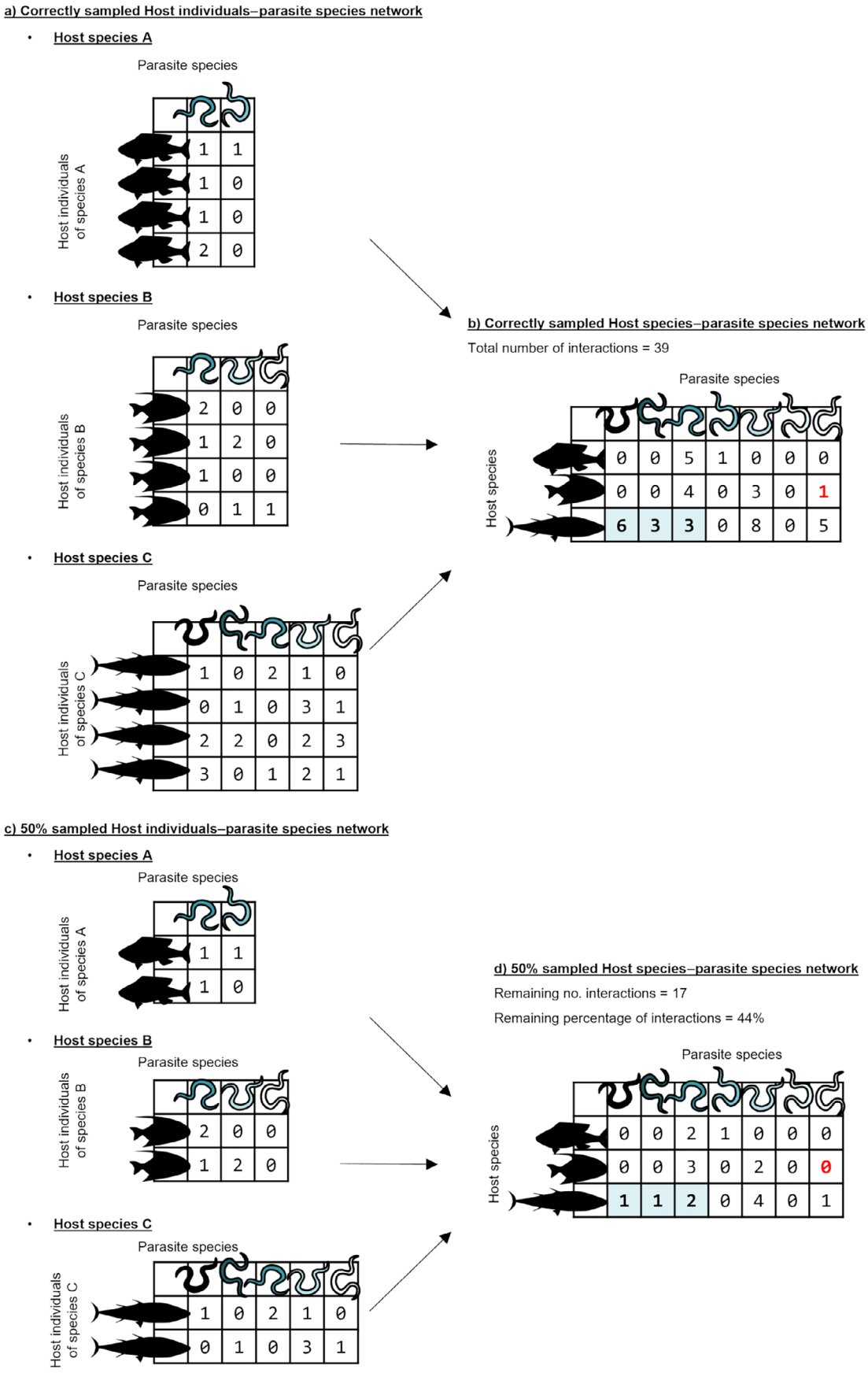

FIGURE 1 Legend on next page.

TABLE 1 Effect of sampling issues on community descriptors.

Definition and sampling issues	Standardized modularity	Standardized nestedness	Connectance [0,1]	H_2' [0,1]
Definition	Pattern in which host–parasite communities are organized in subsets of species that interact more frequently among themselves than with other members of the community (Beckett, 2016). Higher values indicate higher modularity	Pattern in which less rich parasite assemblages are subsets of richer parasite assemblages (Almeida-Neto & Ulrich, 2011). Higher values indicate higher nestedness	Relationship between linkage density and no. species (Tylianakis et al., 2007). Values closer to 1 indicate higher connectance	Specialization (Blüthgen et al., 2006). Higher values of H_2' indicate higher specialization
Host sampling completeness	Decrease	Robust	Robust	Robust
Parasite taxonomic resolution	Decrease	Increment	Increment	Decreasing tendency
Crossed effect	Additive	No evidence of a crossed effect	No evidence of a crossed effect	No evidence of a crossed effect

the gradient of parasite taxonomic resolution varied in size. Connectance expresses the proportion of realized interactions out of all possible interactions (Table 1). A realized interaction represents a higher proportion in a small network than in a larger network. Consequently, for the same number of interactions, communities with low parasite taxonomic resolution (small networks) should present higher connectance than communities with correct taxonomic resolution (large networks). Thus, we expected connectance to be sensitive to parasite taxonomic resolution.

MATERIALS AND METHODS

Building simulated communities

Our main dataset consists of 10 replicate simulated networks (hereafter, "full communities") that were constructed using host–parasite community parameters extracted from published fish species–metazoan parasite species community data ($n = 6$, "natural communities") (Figure 2a; Appendix S1). These communities are host species-level summaries of interactions at the individual level. That is, they do not report host individual–parasite interactions. These six natural communities are not to be confounded with 51 host individual-level datasets used below (see "*Simulating sampling completeness and taxonomic resolution issues*"). To build the full communities, we used a quantitative niche model (see below) that was initiated with the mean number of host and parasite species (nhost = 13; npara = 42) observed in the six natural communities. The mean number of interactions per parasite species (i.e., mean parasite species abundance) in the full communities (maxobs.rf = 2067), was the mean overall number of interactions (or parasite individuals) in the six natural communities (ni = 86,794) divided by npara. In other words, these 10 replicate full communities are a generalization of the different natural host–parasite communities with respect to their number of species and interactions (Figure 2a). We performed all the analyses in R (R Core Team, 2021). If not specified

FIGURE 1 Parasite communities of host individuals of the same species (a, c) are aggregated to obtain the parasite community of each host species and the host species–parasite species network (b, d). When enough host individuals of each host species are collected (a), the host species–parasite species network is a fair representation of the natural community (b). However, when only a few host individuals of each host species are collected (c) the host species–parasite species network does not correctly represent the pattern of interactions of the natural community (d). For example (d), some host–parasite interactions are missed (red value in second row and last column) and others are not presented in the same proportion (shaded cells last row and first three columns). Additionally, although 50% of the host individuals are sampled, ~44% of the interactions are recorded (d).

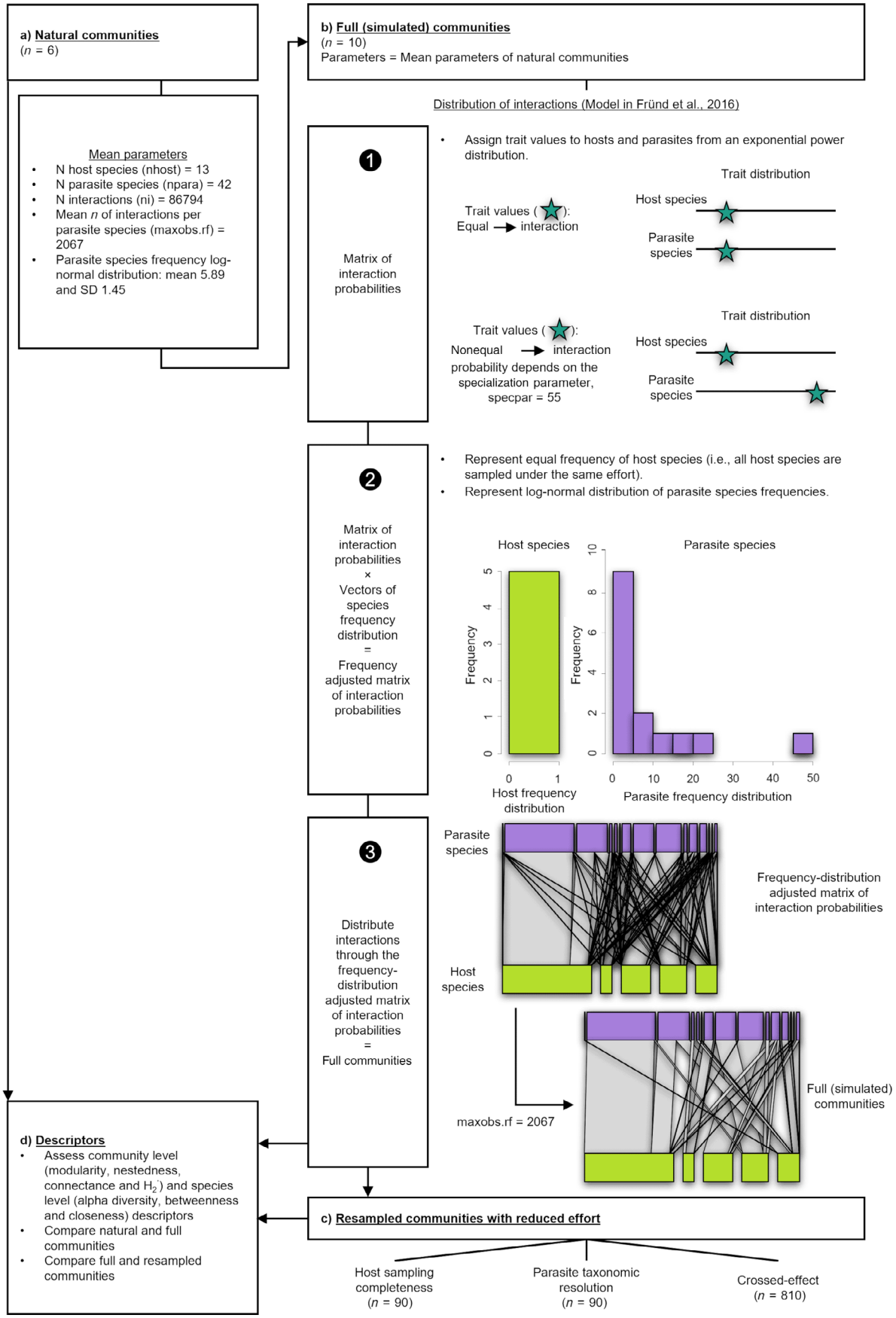

FIGURE 2 Legend on next page.

otherwise, the functions mentioned are available in a bipartite package (Dormann et al., 2008).

The full communities were created with a quantitative niche model that generates weighted bipartite networks reflecting a chosen specialization parameter (Fründ et al., 2016) (Figure 2b). First, the model creates a matrix of interaction probabilities based on quantitative trait values of each host and parasite species (Figure 2b: 1. Matrix of interaction probabilities). Equal trait values of a host and a parasite represent a realized interaction. Nonequal trait values receive an interaction probability depending on the specialization parameter, which defines the shape (width) of the Gaussian niche function. In our case, we assigned trait values to host and parasite species from an exponential power distribution, as the expected distribution of key traits (e.g., body mass) for interaction establishment is usually skewed (Poulin & Morand, 1997). We used the highest specialization parameter used in Fründ et al. (2016) (specpar = 55) (Appendix S2) because specialization is the general trend in metazoan host–parasite interactions (Poulin, 2007). We implemented this procedure with the function *makeweb* (Fründ et al., 2016). Second, we adjusted the interaction probability matrix according to the relative frequency of each species with the function *make_trueweb* (Fründ et al., 2016). That is, by using the relative frequency we consider that it is more likely to record interactions between two frequent species than between two rare ones (Figure 2b: 2. Frequency adjusted matrix of interaction probabilities). We assumed an even frequency distribution of the host species, which corresponds to sampling procedures where the same number of individuals of each host species are captured (Poulin, 2007). With the function *get_skewedabuns* (Fründ et al., 2016), we represented a log-normal distribution with mean 5.89 and SD 1.45 both in the log scale for parasite species' frequencies, which are similar to the mean and SD of parasite species in the natural communities. Therefore, only the distribution of parasite species frequencies adjusted the interaction probability. Finally, we weighted the frequency adjusted interaction probability matrix with parasite abundances with the function *sampleweb* (Fründ et al., 2016) (Figure 2b: 3. Full communities). We assumed a mean abundance or mean number of interactions per parasite species equal to maxobs.rf. This gave us the full communities.

Simulating sampling completeness and taxonomic resolution issues

We resampled the full communities with a reduced effort to simulate communities along decreasing gradients of host sampling completeness, parasite taxonomic resolution, and their crossed effect (hereafter, "resampled communities") (Figure 2c). Each full community was resampled in 10% steps from 90% to 10% of host sampling completeness and parasite taxonomic resolution, thus simulating a situation in which researchers do not have a priori knowledge of the true size and taxonomic structure of the communities they are sampling.

The decreasing gradient of host sampling completeness showed how ecological communities are affected by reducing the number of sampled host individuals of each host species (Figure 1). The full communities only comprised host species–parasite species information (Figure 1b). Parasites follow an aggregated distribution in host individuals: many host individuals carry few parasites, whereas few host individuals have many parasites. Then, the relationship between the percentage of preserved hosts and the percentage of preserved interactions is not perfectly linear (Figure 1d). To realistically simulate a decreasing gradient of sampled host individuals of each species on the full communities, we resampled 51 natural host individuals–parasite species datasets (Figure 1a,c; Appendix S1) (3258 fish host individuals from 17 locations and of 41 species; 63.9 ± 71.4 fish individuals/species). Only host individuals–parasite species datasets were available for one of the natural communities (Valtonen et al., 2001), these were 22 out of the 51 natural host individuals–parasite species datasets. We resampled each of those 51 datasets by preserving from 90% to 10% of its original host individuals and calculated the remaining percentage of interactions in the community at each step (Figure 1c,d). We then calculated the mean remaining percentage of interactions across the 51 datasets at each host sampling completeness step (Table 2a). For example, when we preserved 90% of the host individuals, 85.1% of the interactions remained (Table 2a). These percentages were applied to the mean number of interactions per parasite species (maxobs.rf = 2067) to fill the matrices of interaction probabilities (Figure 2b). Hence, we simulated the decreasing number of sampled host individuals by reducing the mean

FIGURE 2 Analysis steps. (a) Obtain mean parameters of the natural communities to fit the model. (b) Create host–parasite simulated communities (full communities). (1) Build a matrix of interaction probabilities by trait matching and according to a specialization parameter. (2) Adjust the matrix of interactions probabilities by species frequency distributions. (3) Distribute interactions through the frequency adjusted matrix of interaction probabilities (Fründ et al., 2016). (c) Simulate nine levels of sampling biases for host sampling completeness and parasite taxonomic resolution, and their crossed effect. (d) Assess and compare network descriptors between (1) natural and full communities and (2) full and resampled communities.

TABLE 2 Decreasing gradients of (a) host sampling completeness and (b) parasite taxonomic resolution.

Sampling issue	Full communities	Resampled communities								
(a) Host sampling completeness										
% of remaining host individuals	100	90	80	70	60	50	40	30	20	10
% of remaining interactions	100	85.1	78.5	69.3	57.2	50.5	36.8	27.7	17.9	8.2
Mean number interactions per parasite species	2067	1759	1622	1433	1182	1043	761	574	371	170
(b) Parasite taxonomic resolution										
% of parasite taxa	100	90	80	70	60	50	40	30	20	10
Remaining no. parasite taxa	42	37	33	29	25	21	16	12	8	4

number of interactions per parasite species in the full communities. We derived 90 resampled communities affected by host sampling completeness.

To simulate the effect of the decreasing gradient of parasite taxonomic resolution on the full communities, we clustered parasite species based on the similarity of their host range, which resembled common taxonomic resolution issues in parasitology given that closely related species may share hosts. We calculated pairwise Euclidean distances between all the parasite species based on their host-range overlap. These pairwise distances were used to distribute the species to a number of groups for each decreasing resolution class (Table 2b). We implemented the "around medoids" clustering, which is a more robust version of "k-means," with the function *pam* in the "cluster" package (Maechler et al., 2021). For example, for "50%" of parasite taxonomic resolution, we asked the function to assign the parasite species to 21 groups, which represents 50% of the original number of parasite species in the full communities (npara). The overall number of interactions (ni) did not change in this regrouping of parasite species into fewer groups. We obtained 90 resampled communities to examine the effect of parasite taxonomic resolution.

Finally, we simulated the crossed effect of host sampling completeness and parasite taxonomic resolution. A gradient of parasite taxonomic resolution was created for each resampled community in the gradient of host sampling completeness. This produced 810 resampled communities, which were biased for both sampling issues to varying degrees.

Community- and species-level descriptors

We assessed four weighted community-level descriptors for each of the 1000 full and resampled communities: modularity, nestedness, connectance and specialization (H_2') (Table 1). We used the Becket algorithm (Beckett, 2016) to calculate modularity with the function *computeModules*. The function *networklevel* was used to measure the weighted versions of the algorithms: NODF (Nestedness metric based on Overlap and Decreasing Fill, Almeida-Neto & Ulrich, 2011), connectance (Tylianakis et al., 2007) and H_2' (Blüthgen et al., 2006). We standardized the modularity and nestedness of each web since the raw values of these descriptors are not directly comparable. We created 1000 null communities for each of the 1000 full and resampled communities with the function *swap.web*. This restrictive algorithm constrains both connectance and marginal totals (Dormann et al., 2008) (Appendix S3). We then calculated the mean value and the standard deviation of the modularity and nestedness of each set of 1000 null communities. Finally, we standardized the modularity and nestedness of each full and resampled community following the equation of the standardized effect size (SES) (Gotelli & Rohde, 2002) (Equation 1):

$$\mathrm{SES} = \frac{\text{value of a full or resampled community} - \text{mean of its 1000 null communities}}{\text{SD of its 1000 null communities}}. \quad (1)$$

Finally, to assess the effect of each sampling issue and their crossed effect on the community descriptors, four two-way analyses of variance (ANOVAs) were used with host sampling completeness and parasite taxonomic resolution as fixed factors. If the two fixed factors are significant, a significant interaction term indicates a synergistic effect of both sampling issues, otherwise additive (Ferguson & Stiling, 1996).

We ran *t*-tests (Kruskal–Wallis test for standardized modularity) to establish whether network descriptors of the full communities differed significantly from those of natural communities. Additionally, to know whether the full communities reproduced the pattern of interactions of the natural communities at the species level, we computed three common species descriptors for each full and natural community (Dallas et al., 2019;

Llopis-Belenguer et al., 2020; Poulin et al., 2013). Rao alpha diversity measures the richness and abundance (Pavoine et al., 2004) of the parasite community of a host species. We then measured two weighted descriptors of centrality for both host and parasite species. Betweenness is the extent to which a species funnels the interactions among all other species in the community. Closeness is the average distance of a species to all other species in the community (Newman, 2001). We then statistically tested the differences between sample-size corrected species descriptors of the full and natural communities (Appendix S3). Finally, to reassure the resemblance between the full and natural antagonistic communities, we compared them with natural mutualistic communities (Appendix S4).

RESULTS

Community- and species-level descriptors of the full communities did not significantly differ from descriptors of the natural communities, except for modularity and closeness (Table 3; see Appendix S4 for justification of the parameter choice used in the quantitative niche model).

With respect to host sampling completeness, standardized modularity decreased up to ~30% (Figure 3a darkest purple points and lines; Table 4). Standardized modularity of resampled communities capturing 40% or less of host sampling completeness differed from those values of the full communities (Figure 3a, error bars do not overlap). In contrast, standardized nestedness, connectance and H_2' were robust to host sampling completeness as they showed little difference between the resampled and full communities (Figure 3, Table 4).

The four indices were affected by parasite taxonomic resolution (Figure 3 100% of host sampling completeness; Table 4). Standardized modularity decreased ~76%, and values of resampled communities with 70% or less of parasite taxonomic resolution differed from standardized modularity of the full communities (Figure 3a, error bars do not overlap). Both standardized nestedness and connectance increased ~114% and ~345%, respectively, as parasite taxonomic resolution decreased. Standardized nestedness of resampled communities with 90% of parasite taxonomic resolution and below differed from nestedness of the full communities (Figure 3b, error bars do not overlap). We found moderate to strong differences between the connectance of the full communities and the connectance of the resampled communities with 90% of parasite taxonomic resolution or less (Figure 3c, error bars do not overlap). H_2' decreased ~15% with parasite taxonomic resolution. We found differences between H_2' of full communities and resampled communities with 50% or less of parasite taxonomic resolution (Figure 3d).

The crossed effect of host sampling completeness and parasite taxonomic resolution was additive for standardized modularity, as both effects were significant and no evidence of an interaction was found (Figure 3, Table 4). Parasite taxonomic resolution had a higher influence on network metrics than host sampling completeness. The standard error of the mean tended to increase with sampling bias, showing proportionally higher variation for connectance than for the other three indices (Figure 3, Table 4).

TABLE 3 Comparisons of the weighted descriptors between the full and natural communities.

		Community level				Species level				
Communities and tests	**Paramenters**	**M**	**N**	**C**	**H_2'**	**Bp**	**Bh**	**Clp**	**Clh**	**Rao alpha diversity**
Full communities ($n = 10$)	Minimum	11.08	−8.68	0.05	0.67	0	0	0.001	0.02	1.08
	Maximum	20.4	−6.14	0.07	0.84	0.46	0.46	0.13	0.22	7.9
	Mean	18.07	−7.36	0.05	0.76	0.02	0.08	0.02	0.09	3.22
	Standardized error	0.83	0.26	0.00	0.01	0.003	0.01	0.001	0.004	0.12
Natural communities ($n = 6$)	Minimum	3.32	−9.34	0.04	0.44	0	0	10^{-5}	0	1
	Maximum	24.55	−3.79	0.07	0.88	0.77	0.72	0.08	0.7	9.54
	Mean	10.94	−6.43	0.05	0.66	0.02	0.08	0.005	0.06	2.73
	Standardized error	3.07	0.75	0.00	0.06	0.005	0.02	6×10^{-4}	0.01	0.19
Kruskal–Wallis (M)/*t*-test (N, C, H_2')/ ANOVA (species level)	*p*-value	0.04	0.28	0.92	0.16	0.33	0.88	0.00	0.00	0.24

Abbreviations: B, betweenness; C, connectance; Cl, closeness; h, hosts; H_2', specialization; M, standardized modularity; N, standardized nestedness; p, parasites.

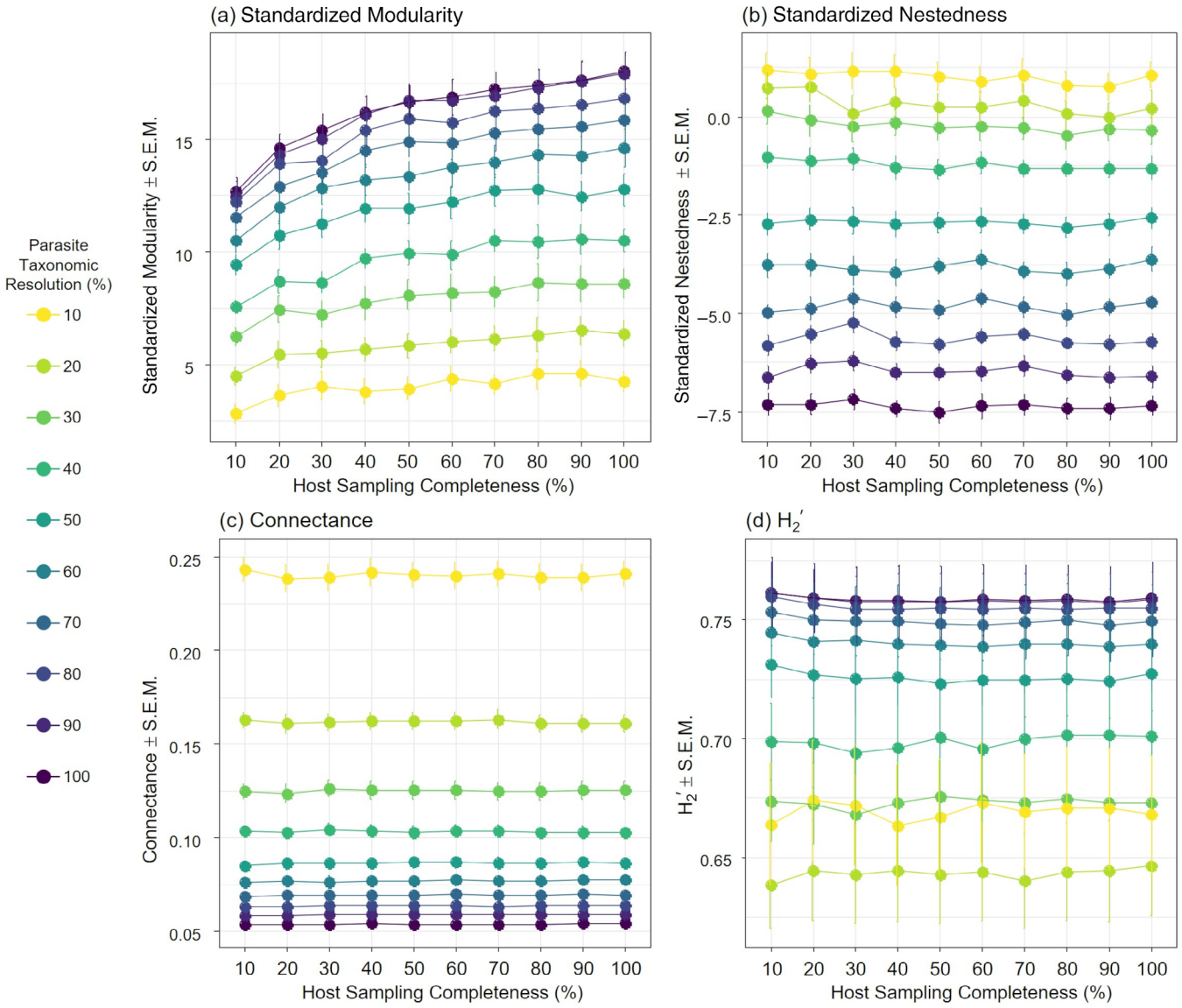

FIGURE 3 Effect of the decreasing gradient of host sampling completeness, parasite taxonomic resolution, and their crossed effect on host–parasite community descriptors: (a) Standardized Modularity; (b) Standardized Nestedness; (c) Connectance; and (d) H_2'.

Finally, for some resampled communities, modularity (20%–10% of parasite taxonomic resolution and 40% or less of host sampling completeness) and nestedness (all levels of parasite taxonomic resolution capturing 50% host sampling completeness or less) did not differ from the modularity and nestedness of 1000 random assemblages (results not shown).

DISCUSSION

Low sampling completeness and low taxonomic resolution are common sources of uncertainty in community ecology and influence the interpretation of bipartite interactions (Chacoff et al., 2012; Rodrigues & Boscolo, 2020). In the present study, community descriptors were more sensitive to parasite taxonomic resolution than to host sampling completeness, in accordance with Hypothesis 1. The crossed effect of both sampling issues was additive for modularity, supporting Hypothesis 2 (Table 4). Additionally, the descriptors differed in their sensitivity to both sampling issues. This partially concurs with our Hypothesis 3 as modularity was not as robust as in previous studies. Acknowledging that sampling issues are inevitable to some extent, studies should (1) evaluate both sampling completeness and taxonomic resolution when conducting bipartite network analyses; (2) avoid applying bipartite network analyses to communities with low sampling effort or taxonomic resolution; (3) use the most robust measures to evaluate community structure

TABLE 4 Two-way ANOVA of host sampling completeness and parasite taxonomic resolution for (a) standardized modularity, (b) standardized nestedness, (c) connectance, and (d) specialization (H_2').

(a) Standardized modularity	***df***	**SS**	**MS**	***F***	***p*-value**
Sampling completeness	9	1151.4	127.93	27.96	<0.001
Taxonomic resolution	9	17502.9	1944.76	425.07	<0.001
Interaction	81	167.5	2.07	0.45	1
Residuals	900	4117.7	4.58		
(b) Standardized nestedness	***df***	**SS**	**MS**	***F***	***p*-value**
Sampling completeness	9	7.6	0.84	0.91	0.52
Taxonomic resolution	9	8009.5	889.94	956.57	<0.001
Interaction	81	14.1	0.17	0.19	1
Residuals	900	837.3	0.93		
(c) Connectance	***df***	**SS**	**MS**	***F***	***p*-value**
Sampling completeness	9	0.00	0.00	0.05	1
Taxonomic resolution	9	3.08	0.34	2359.22	<0.001
Interaction	81	0.00	0.00	0.03	1
Residuals	900	0.13	0.00		
(d) H_2'	***df***	**SS**	**MS**	***F***	***p*-value**
Sampling completeness	9	0.00	0.00	0.02	1
Taxonomic resolution	9	1.66	0.18	60.75	<0.001
Interaction	81	0.00	0.00	0.01	1
Residuals	900	2.74	0.00		

Abbreviations: *df*, degrees of freedom; *F*, *F* statistic; MS, mean square; SS, sum-of-squares.

in communities severely affected by sampling issues (nestedness, connectance and H_2' were robust even in case of severely under-sampled communities); (4) pay attention to the conclusions relying on more sensitive metrics (all descriptors were sensitive to taxonomic resolution); and (5) compare interaction patterns over time and space of communities with comparable and adequate sampling efforts, especially for taxonomic resolution.

Antagonistic communities are often highly modular (Runghen et al., 2021), most likely due to parasite specialization on its host resource (Krasnov et al., 2012). We found that modularity was sensitive to the reduction in both host sampling completeness and parasite taxonomic resolution. Lower sampling efforts left less frequent interactions undetected, which decreased resolution in module uniqueness. Contrary to our results, modularity was found to be a robust descriptor in mutualistic communities. Moreover, modularity increased in mutualistic communities with lower sampling efforts due to increased module identity through the removal of between-module interactions (Rivera-Hutinel et al., 2012; Vizentin-Bugoni et al., 2016). Low taxonomic resolution can decrease modularity, as it was found here and in other plant–insect mutualistic and antagonistic systems (Renaud et al., 2020; Rodrigues & Boscolo, 2020). We suggest that the aggregation of parasite species along the gradient of parasite taxonomic resolution makes host species with truly different parasite communities members of the same module. However, these host species can be additionally connected to other modules where the rest of their parasite species are placed. Hence, the loss of parasite taxonomic resolution makes between-module interactions more frequent, which decreases modularity.

Nestedness has been shown to be reasonably robust to low sampling completeness (Fründ et al., 2016; Henriksen et al., 2019; Nielsen & Bascompte, 2007; Vizentin-Bugoni et al., 2016) and low taxonomic resolution (Renaud et al., 2020; Rodrigues & Boscolo, 2020) in both mutualistic and antagonistic communities. However, in our study, nestedness was robust to the loss of sampling completeness but sensitive to the loss of taxonomic resolution. Host–parasite communities typically show low values of nestedness (Runghen et al., 2021), possibly resulting from coevolution leading to trade-offs in parasite transmission (McQuaid & Britton, 2013). The full communities had low values of nestedness (Figure 3b and Table 3: negative values) that increased with decreasing taxonomic resolution. In this gradient, aggregated parasite nodes appeared as a single generalist parasite able to infect shared and

nonshared hosts of their foundational parasite nodes, whereas specialist parasites were not grouped with any other nodes. Then, the network structure became more nested because the host spectrum of specialist parasites was a subset of the host species used by generalist parasites.

Connectance and specialization (H_2') were robust to decreasing host sampling completeness, but not to parasite taxonomic resolution. These descriptors showed opposite patterns, as expected by definition (Table 1). Our results were consistent with previous studies reporting the robustness of these metrics to the loss of interactions (Fründ et al., 2016, Henriksen et al., 2019, Nielsen & Bascompte, 2007, Vizentin-Bugoni et al., 2016), but not to the loss of species (Fründ et al., 2016; Henriksen et al., 2019; Renaud et al., 2020; Rodrigues & Boscolo, 2020). The strong dependence of connectance on network size hindered the interpretation of many biological processes (Blüthgen et al., 2006). This pattern was evident along our crossed-effect gradient, where decreasing parasite taxonomic resolution increased the relative contribution of the interactions. H_2' was developed as an alternative index of specialization to overcome the scale dependence issue of connectance (Blüthgen et al., 2006). Although we found H_2' to be sensitive to the loss of parasite taxonomic resolution, it was much less affected than connectance (~16% decrease for H_2' against ~350% increment for connectance in communities with the lowest efforts, i.e., 10% host sampling completeness × 10% parasite taxonomic resolution) (Figure 3).

The loss of parasite taxonomic resolution led to a wide range of possible community outcomes (Figure 3 wider error bars), which made it difficult to predict its effect on a given network. Communities in the gradient of parasite taxonomic resolution vary in size. Large and small communities are biased to different degrees (Henriksen et al., 2019; Shvydka et al., 2018). The full communities represented the largest communities, whereas resampled communities with the lower parasite taxonomic resolution were the smallest communities (Table 2b). The latter aggregated most of the parasite species with similar infection patterns, losing redundant interactions. Typically, large communities showed a higher overlap in their species' interaction patterns compared with small communities, where redundancy is low (Henriksen et al., 2019). Hence, sampling issues usually have a limited impact on the network structure of large communities because, if a species or interaction is not recorded, the overall infection patterns are nevertheless recorded in the redundant interactions. However, the exclusion of species or interactions represents a proportionally greater change in the network structure of small communities. Therefore, the redundancy–size relationship could explain the increasing variance observed in the resampled communities along the parasite taxonomic resolution gradient. Furthermore, the categorical description of the resampled communities should be viewed with caution (e.g., if a community is modular or not). This is because the modularity and nestedness of resampled communities with the lower efforts could not be distinguished from those values of random assemblages (last paragraph in "*Results*").

The combination of both sampling issues was additive to modularity. Predictive models potentially overcome the limitation of incomplete interaction richness. These models identify where interactions are most likely to be missed in a sampled community and eventually include them in the dataset to improve the study of ecological networks (Terry & Lewis, 2020). Similarly, information on permitted and forbidden interactions (e.g., a feeding interaction from a low to a high trophic taxon is forbidden) helps to build a more realistic representation of the community (Strona & Veech, 2017). Despite its greater impact on community descriptors, solutions to taxonomic limitations may require, for example, the collaboration between ecologists and taxonomists (Poulin & Presswell, 2022).

Community- and species-level descriptors of the full and natural communities were not significantly different, except for modularity and closeness. Despite the significant differences between the standardized modularity of the full and natural communities, those values of the full communities fell within the range of the standardized modularity of the natural communities. Therefore, standardized modularities of the full communities represent realistic values for host–parasite communities (Appendix S4). Earlier studies evaluated species centrality descriptors, such as closeness, of both parasites (Poulin et al., 2013) and hosts (Dallas et al., 2019). Taxonomic identification at the family level was important in explaining the centrality of parasite and host species, suggesting that phylogeny could help to predict the centrality of species (Poulin et al., 2013). We did not account for the phylogenetic structure of the guilds in the full communities. Instead, each interaction acquired a probability according to the specialization parameter in the model, but independently of the phylogenetic distance among the members of each guild. The consideration of the phylogenetic structure of parasites and hosts in our model could improve our representation of host–parasite communities and predictions. Nonetheless, we considered our approach representative of natural communities as most of the community- and species-level descriptors were effectively captured in the full communities.

The latest methods available in open-access software facilitate the use of network analysis in parasitology (Runghen et al., 2021). The increasing availability of host–parasite interaction datasets also favors their

comparison to address macroecological questions (Doherty et al., 2021) or ecosystem long-term dynamics and trends (Carlson et al., 2020). However, if communities in comparative studies notably differ in, or do not include, sufficient completeness and resolution, the conclusions extracted from the network analyses of such data will be of limited use, if not defective, as our study shows.

AUTHOR CONTRIBUTIONS

Cristina Llopis-Belenguer and Jukka Jokela conceived the ideas; Juan Antonio Balbuena, Isabel Blasco-Costa, and Volodimir Sarabeev collected some of the natural communities; Anssi Karvonen complied one of the datasets; Cristina Llopis-Belenguer designed methodology, arranged and analyzed the data; Cristina Llopis-Belenguer led the writing. All authors contributed to the drafts and gave approval for publication.

ACKNOWLEDGMENTS

We thank Dr. S. Dennis (EAWAG) for his computation assistance. Cristina Llopis-Belenguer acknowledges the support of an ETH Postdoctoral Fellowship (20-2 FEL-67). Volodimir Sarabeev thanks the support from the Slovakian Government for Excellent Researchers Threatened by the War Conflict in Ukraine (09I03-03-V01-00017). Isabel Blasco-Costa acknowledges the funding from the Swiss National Science Foundation (SNSF grant 31003A_169211). Open access funding provided by Eidgenossische Technische Hochschule Zurich.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Replication data (Llopis-Belenguer et al., 2022a) are available in Zenodo at https://doi.org/10.5281/zenodo.7386012. Data for parasite communities of *Coregonus* spp. from Swiss and Norwegian lakes (Brabec et al., 2022) are available in Zenodo at https://doi.org/10.5281/zenodo.7411957. Data for parasite communities of fishes from the Northeastern Baltic Sea (Llopis-Belenguer et al., 2022b) are available in Zenodo at https://doi.org/10.5281/zenodo.7390536. Additional data sources are listed in Appendix S1.

ORCID

Cristina Llopis-Belenguer https://orcid.org/0000-0001-7672-0088
Juan Antonio Balbuena https://orcid.org/0000-0003-4006-1353
Isabel Blasco-Costa https://orcid.org/0000-0002-5001-5986
Anssi Karvonen https://orcid.org/0000-0002-2731-3344
Volodimir Sarabeev https://orcid.org/0000-0003-4724-3141
Jukka Jokela https://orcid.org/0000-0002-1731-727X

REFERENCES

Almeida-Neto, M., and W. Ulrich. 2011. "A Straightforward Computational Approach for Measuring Nestedness Using Quantitative Matrices." *Environmental Modelling & Software* 26: 173–8.

Beckett, S. J. 2016. "Improved Community Detection in Weighted Bipartite Networks." *Royal Society Open Science* 3: 140536.

Bellay, S., E. F. de Oliveira, M. Almeida-Neto, V. D. Abdallah, R. K. de Azevedo, R. M. Takemoto, and J. L. Luque. 2015. "The Patterns of Organisation and Structure of Interactions in a Fish-Parasite Network of a Neotropical River." *International Journal for Parasitology* 45: 549–57.

Blüthgen, N., F. Menzel, and N. Blüthgen. 2006. "Measuring Specialization in Species Interaction Networks." *BMC Ecology* 6: 9.

Brabec, J., O. Selz, R. Knudsen, and I. Blasco-Costa. 2022. "Parasite communities of *Coregonus* spp. from Swiss and Norwegian Lakes." Zenodo. https://doi.org/10.5281/zenodo.7411957.

Cardoso, T. D. S., C. S. Andreazzi, A. Maldonado Junior, and R. Gentile. 2021. "Functional Traits Shape Small Mammal-Helminth Network: Patterns and Processes in Species Interactions." *Parasitology* 148: 947–55.

Carlson, C. J., S. Hopkins, K. C. Bell, J. Doña, S. S. Godfrey, M. L. Kwak, K. D. Lafferty, et al. 2020. "A Global Parasite Conservation Plan." *Biological Conservation* 250: 108596.

Chacoff, N. P., D. P. Vázquez, S. B. Lomáscolo, E. L. Stevani, J. Dorado, and B. Padrón. 2012. "Evaluating Sampling Completeness in a Desert Plant–Pollinator Network." *Journal of Animal Ecology* 81: 190–200.

Dallas, T. A., B. A. Han, C. L. Nunn, A. W. Park, P. R. Stephens, and J. M. Drake. 2019. "Host Traits Associated with Species Roles in Parasite Sharing Networks." *Oikos* 128: 23–32.

Doherty, J. F., X. Chai, L. E. Cope, D. de Angeli Dutra, M. Milotic, S. Ni, E. Park, and A. Filion. 2021. "The Rise of Big Data in Disease Ecology." *Trends in Parasitology* 37: 1034–7.

Dormann, C. F., J. Fründ, and H. M. Schaefer. 2017. "Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations." *Annual Review of Ecology, Evolution, and Systematics* 48: 559–84.

Dormann, C. F., B. Gruber, and J. Fründ. 2008. "Introducing the bipartite Package: Analysing Ecological Networks." *R News* 8: 8–11.

Ferguson, K. I., and P. Stiling. 1996. "Non-additive Effects of Multiple Natural Enemies on Aphid Populations." *Oecologia* 108: 375–9.

Fründ, J., K. S. McCann, and N. M. Williams. 2016. "Sampling Bias is a Challenge for Quantifying Specialization and Network Structure: Lessons from a Quantitative Niche Model." *Oikos* 125: 502–13.

Gotelli, N. J., and K. Rohde. 2002. "Co-Occurrence of Ectoparasites of Marine Fishes: A Null Model Analysis." *Ecology Letters* 5: 86–94.

Henriksen, M. V., D. G. Chapple, S. L. Chown, and M. A. McGeoch. 2019. "The Effect of Network Size and Sampling Completeness in Depauperate Networks." *Journal of Animal Ecology* 88: 211–22.

Isaac, N. J. B., J. Mallet, and G. M. Mace. 2004. "Taxonomic Inflation: Its Influence on Macroecology and Conservation." *Trends in Ecology & Evolution* 19: 464–9.

Krasnov, B. R., M. A. Fortuna, D. Mouillot, I. S. Khokhlova, G. I. Senbrot, and R. Poulin. 2012. "Phylogenetic Signal in Module Composition and Species Connectivity in Compartmentalized Host-Parasite Networks." *The American Naturalist* 179: 501–11.

Lafferty, K. D., A. P. Dobson, and A. M. Kuris. 2006. "Parasites Dominate Food Web Links." *Proceedings of the National Academy of Sciences* 103: 11211–6.

Llopis-Belenguer, C., J. A. Balbuena, I. Blasco-Costa, A. Karvonen, V. Sarabeev, and J. Jokela. 2022a. "Replication Data for: Sensitivity of Bipartite Network Analyses to Incomplete Sampling and Taxonomic Uncertainty." Zenodo. https://doi.org/10.5281/zenodo.7386012

Llopis-Belenguer, C., J. A. Balbuena, I. Blasco-Costa, A. Karvonen, V. Sarabeev, and J. Jokela. 2022b. "Parasite Communities of Fishes from Northeastern Baltic Sea (Data in Valtonen, et al., 2001)." Zenodo https://doi.org/10.5281/zenodo.7390536.

Llopis-Belenguer, C., S. Pavoine, I. Blasco-Costa, and J. A. Balbuena. 2020. "Assembly Rules of Helminth Parasite Communities in Grey Mullets: Combining Components of Diversity." *International Journal for Parasitology* 50: 1089–98.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and R. K. Hornik. 2021. "cluster: Cluster Analysis Basics and Extensions." https://cran.r-project.org/web/packages/cluster/index.html.

McQuaid, C. F., and N. F. Britton. 2013. "Host-Parasite Nestedness: A Result of co-Evolving Trait-Values." *Ecological Complexity* 13: 53–9.

Newman, M. E. J. 2001. "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality." *Physical Review E* 64: 016132.

Nielsen, A., and J. Bascompte. 2007. "Ecological Networks, Nestedness and Sampling Effort." *Journal of Ecology* 95: 1134–41.

Pavoine, S., A.-B. Dufour, and D. Chessel. 2004. "From Dissimilarities among Species to Dissimilarities among Communities: A Double Principal Coordinate Analysis." *Journal of Theoretical Biology* 228: 523–37.

Poulin, R. 2007. *Evolutionary Ecology of Parasites*. Princeton, NJ: Princeton University Press.

Poulin, R., B. R. Krasnov, S. Pilosof, and D. W. Thieltges. 2013. "Phylogeny Determines the Role of Helminth Parasites in Intertidal Food Webs." *Journal of Animal Ecology* 82: 1265–75.

Poulin, R., and S. Morand. 1997. "Parasite Body Size Distributions: Interpreting Patterns of Skewness." *International Journal for Parasitology* 27: 959–64.

Poulin, R., and B. Presswell. 2022. "Is Parasite Taxonomy Really in Trouble? A Quantitative Analysis." *International Journal for Parasitology* 52: 469–74.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Renaud, E., E. Baudry, and C. Bessa-Gomes. 2020. "Influence of Taxonomic Resolution on Mutualistic Network Properties." *Ecology and Evolution* 10: 3248–59.

Rivera-Hutinel, A., R. O. Bustamante, V. H. Marín, and R. Medel. 2012. "Effects of Sampling Completeness on the Structure of Plant–Pollinator Networks." *Ecology* 93: 1593–603.

Rodrigues, B. N., and D. Boscolo. 2020. "Do Bipartite Binary Antagonistic and Mutualistic Networks Have Different Responses to the Taxonomic Resolution of Nodes?" *Ecological Entomology* 45: 709–17.

Runghen, R., R. Poulin, C. Monlleó-Borrull, and C. Llopis-Belenguer. 2021. "Network Analysis: Ten Years Shining Light on Host–Parasite Interactions." *Trends in Parasitology* 37: 445–55.

Shvydka, S., V. Sarabeev, V. D. Estruch, and C. Cadarso-Suárez. 2018. "Optimum Sample Size to Estimate Mean Parasite Abundance in Fish Parasite Surveys." *Helminthologia* 55: 52–9.

Strona, G., ed. 2022. "Ecological Networks." In *Hidden Pathways to Extinction*, 41–55. Cham: Springer.

Strona, G., and J. A. Veech. 2017. "Forbidden versus Permitted Interactions: Disentangling Processes from Patterns in Ecological Network Analysis." *Ecology and Evolution* 7: 5476–81.

Terry, J. C. D., and O. T. Lewis. 2020. "Finding Missing Links in Interaction Networks." *Ecology* 101: e03047.

Thompson, R. M., and C. R. Townsend. 2000. "Is Resolution the Solution?: The Effect of Taxonomic Resolution on the Calculated Properties of Three Stream Food Webs." *Freshwater Biology* 44: 413–22.

Tylianakis, J. M., T. Tscharntke, and O. T. Lewis. 2007. "Habitat Modification Alters the Structure of Tropical Host–Parasitoid Food Webs." *Nature* 445: 202–5.

Valtonen, E. T., K. Pulkkinen, R. Poulin, and M. Julkunen. 2001. "The Structure of Parasite Component Communities in Brackish Water Fishes of the Northeastern Baltic Sea." *Parasitology* 122: 471–81.

Vizentin-Bugoni, J., P. K. Maruyama, V. J. Debastiani, L. d. S. Duarte, B. Dalsgaard, and M. Sazima. 2016. "Influences of Sampling Effort on Detected Patterns and Structuring Processes of a Neotropical Plant–Hummingbird Network." *Journal of Animal Ecology* 85: 262–72.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Llopis-Belenguer, Cristina, Juan Antonio Balbuena, Isabel Blasco-Costa, Anssi Karvonen, Volodimir Sarabeev, and Jukka Jokela. 2023. "Sensitivity of Bipartite Network Analyses to Incomplete Sampling and Taxonomic Uncertainty." *Ecology* 104(4): e3974. https://doi.org/10.1002/ecy.3974