

# Sampling networks of ecological interactions

Pedro Jordano<sup>\*a</sup>

<sup>a</sup>Integrative Ecology Group, Estación Biológica de Doñana, Consejo  
Superior de Investigaciones Científicas (EBD-CSIC), Avenida  
Americo Vesputio s/n, E-41092 Sevilla, Spain

In review, Functional Ecology, July 19, 2016

## Summary

1. Sampling ecological interactions presents similar challenges, problems, potential biases, and constraints as sampling individuals and species in biodiversity inventories. Sampling interactions is a fundamental step to build robustly estimated interaction networks, yet few analyses have attempted a formal approach to their sampling protocols. Robust estimates of the actual number of interactions (links) within diversified ecological networks require adequate sampling effort that needs to be explicitly gauged. Yet we still lack a sampling theory explicitly focusing on ecological interactions.

---

<sup>\*</sup>jordano@ebd.csic.es

2. While the complete inventory of interactions is likely impossible, a robust characterization of its main patterns and metrics is probably realistic. We must acknowledge that a sizeable fraction of the maximum number of interactions  $I_{max}$  among, say,  $A$  animal species and  $P$  plant species (i.e.,  $I_{max} = AP$ ) is impossible to record due to forbidden links, i.e., life-history restrictions. Thus, the number of observed interactions  $I$  in robustly sampled networks is typically  $I \ll I_{max}$ , resulting in sparse interaction matrices with low connectance.
3. Reasons for forbidden links are multiple but mainly stem from spatial and temporal uncoupling, size mismatches, and intrinsically low probabilities of interspecific encounter for most potential interactions of partner species. Adequately assessing the completeness of a network of ecological interactions thus needs knowledge of the natural history details embedded, so that forbidden links can be accounted for when addressing sampling effort.
4. Here I provide a review and outline a conceptual framework for interaction sampling by building an explicit analogue to individuals and species sampling, thus extending diversity-monitoring approaches to the characterization of complex networks of ecological interactions. Contrary to species inventories, a sizable fraction of non-observed pairwise interactions cannot be sampled, due to biological constraints that forbid their occurrence.
5. Recent implementations of inference methods for unobserved species or for individual-based data can be combined with the assessment of forbidden links. This can help in estimating their relative importance, simply by the difference between the asymptotic estimate of interaction richness *in a*

*robustly-sampled* assemblage and the maximum richness  $I_{max}$  of interactions.

This is crucial to assess the rapid and devastating effects of defaunation-driven loss of key ecological interactions and the services they provide and the analogous losses related to interaction gains due to invasive species and biotic homogenization.

## Keywords

complex networks, food webs, frugivory, mutualism, plant-animal interactions, pollination, seed dispersal

## Introduction

Biodiversity sampling is a labour-intensive activity, and sampling is often not sufficient to detect all or even most of the species present in an assemblage. Gotelli & Colwell (2011).

1 Biodiversity species assessment aims at sampling individuals in collections and  
2 determining the number of species represented. Given that, by definition, samples  
3 are incomplete, these collections do not enumerate the species actually present.  
4 The ecological literature dealing with robust estimators of species richness and di-  
5 versity in collections of individuals is immense, and a number of useful approaches  
6 have been used to obtain such estimates (Magurran, 1988; Gotelli & Colwell, 2001;  
7 Colwell *et al.*, 2004; Hortal *et al.*, 2006; Colwell, 2009; Gotelli & Colwell, 2011; Chao  
8 *et al.*, 2014). Recent effort has been also focused at defining essential biodiversity

9 variables (EBV) (Pereira *et al.*, 2013) that can be sampled and measured repeat-  
10 edly to complement biodiversity estimates. Yet sampling species or taxa-specific  
11 EBVs is just probing a single component of biodiversity; interactions among species  
12 are another fundamental component, one that supports the existence, but in some  
13 cases also the extinction, of species. For example, the extinction of interactions  
14 represents a dramatic loss of biodiversity because it entails the loss of fundamental  
15 ecological functions (Valiente-Banuet *et al.*, 2014). This missed component of bio-  
16 diversity loss, the extinction of ecological interactions, very often accompanies, or  
17 even precedes, species disappearance. Interactions among species are a key com-  
18 ponent of biodiversity and here I aim to show that most problems associated with  
19 sampling interactions in natural communities relate to, and are even worse than,  
20 problems associated with sampling species diversity. I consider pairwise interac-  
21 tions among species at the habitat level, in the context of alpha diversity and the  
22 estimation of local interaction richness from sampling data (Chao *et al.*, 2014). In  
23 the first part I provide a succinct overview of previous work addressing sampling  
24 issues for ecological interaction networks. In the second part, I discuss specific ra-  
25 tionales for sampling the biodiversity of ecological interactions. Finally, I provide  
26 a short overview of asymptotic diversity estimates (Gotelli & Colwell, 2001), and  
27 a discussion of its application to interaction sampling. Most of the examples come  
28 from the analysis of plant-animal interaction networks, yet are applicable to other  
29 types of interspecific interactions.

30 Interactions can be a much better indicator of the richness and diversity of  
31 ecosystem functions than a simple list of taxa and their abundances and/or re-  
32 lated biodiversity indicator variables (EBVs) (Memmott *et al.*, 2006; Valiente-  
33 Banuet *et al.*, 2014). Thus, sampling interactions should be a central issue when

identifying and diagnosing ecosystem services (e.g., pollination, seeding by frugivores, etc.). Fortunately, the whole battery of biodiversity-related tools used by ecologists to sample biodiversity (species, *sensu stricto*) can be extended and applied to the sampling of interactions (see Table 2 in Colwell *et al.*, 2004). Monitoring interactions is a type of biodiversity sampling and is subject to similar methodological shortcomings, especially under-sampling (Jordano, 1987; Jordano *et al.*, 2009; Coddington *et al.*, 2009; Vázquez *et al.*, 2009; Dorado *et al.*, 2011; Rivera-Hutinel *et al.*, 2012). For example, when we study mutualistic networks, our goal is to make an inventory of the distinct pairwise interactions that made up the network. We are interested in having a complete list of all the pairwise interactions among species (e.g., all the distinct, species-species interactions, or links, among the pollinators and flowering plants) that do actually exist in a given community. Sampling these interactions thus entails exactly the same problems, limitations, constraints, and potential biases as sampling individual organisms and species diversity. As Mao & Colwell (2005) put it, these are the workings of Preston’s demon, the moving “veil line” (Preston, 1948) between the detected and the undetected interactions as sample size increases.

Early efforts to recognize and solve sampling problems in analyses of interactions stem from research on food webs and to determine how undersampling biases food web metrics (Martinez, 1991; ?, 1993; Bersier *et al.*, 2002; Brose *et al.*, 2003; Banasek-Richter *et al.*, 2004; Wells & O’Hara, 2012). In addition, the myriad of classic natural history studies documenting animal diets, host-pathogen infection records, plant herbivory records, etc., represent efforts to document interactions occurring in nature. All of them share the problem of sampling incompleteness influencing the patterns and metrics reported. Yet, despite the early recogni-

tion that incomplete sampling may seriously bias the analysis of ecological networks (Jordano, 1987), only recent studies have explicitly acknowledged it and attempted to determine its influence (Ollerton & Cranmer, 2002; Nielsen & Bascompte, 2007; Vázquez *et al.*, 2009; Gibson *et al.*, 2011; Olesen *et al.*, 2011; Chacoff *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Bascompte & Jordano, 2014; Vizentin-Bugoni *et al.*, 2014, 2016; Frund *et al.*, 2015). The sampling approaches have been extended to predict patterns of coextinctions in interaction assemblages (e.g., hosts-parasites) (Colwell *et al.*, 2012). Most empirical studies provide no indication of sampling effort, implicitly assuming that the reported network patterns and metrics are robust. Yet recent evidences point out that number of partner species detected, number of actual links, and some aggregate statistics describing network patterns, are prone to sampling bias (Nielsen & Bascompte, 2007; Dorado *et al.*, 2011; Olesen *et al.*, 2011; Chacoff *et al.*, 2012; Rivera-Hutinel *et al.*, 2012; Olito & Fox, 2014; Frund *et al.*, 2015). Most of these evidences, however, come either from simulation studies (Frund *et al.*, 2015) or from relatively species-poor assemblages. Most certainly, sampling limitations pervade biodiversity inventories in tropical areas (Coddington *et al.*, 2009) and we might rightly expect that frequent interactions may be over-represented and rare interactions may be missed entirely in studies of mega-diverse assemblages (Bascompte & Jordano, 2014); but, to what extent?

## Sampling interactions: methods

When we sample interactions in the field we record the presence of two species that interact in some way. For example, Snow and Snow(1988) recorded an inter-

82 action whenever they saw a bird “touching” a fruit on a plant. We observe and  
 83 record feeding observations, visitation, occupancy, presence in pollen loads or in  
 84 fecal samples, etc., of *individual* animals or plants and accumulate pairwise inter-  
 85 actions, i.e., lists of species partners and the frequencies with which we observe  
 86 them. We assume that the matrix (species numbers) is predefined (i.e., all species  
 87 interacting are well documented). Therefore, estimating the sampling complete-  
 88 ness of pairwise interactions for a whole network, requires some gauging of how the  
 89 number (richness) of distinct pairwise interactions accumulates as sampling effort  
 90 is increased) and/or estimating the uncertainty around the missed links (Wells &  
 91 O’Hara, 2012).

92 Most types of ecological interactions can be illustrated with bipartite graphs,  
 93 with two or more distinct groups of interacting partners (Bascompte & Jordano,  
 94 2014); for illustration purposes I’ll focus more specifically on plant-animal inter-  
 95 actions. Sampling interactions requires filling the cells of an interaction matrix  
 96 with data. The matrix,  $\Delta = AP$  (the adjacency matrix for the graph representa-  
 97 tion of the network), is a 2D inventory of the interactions among, say,  $A$  animal  
 98 species (rows) and  $P$  plant species (columns) (Jordano, 1987; Bascompte & Jor-  
 99 dano, 2014). The matrix entries illustrate the values of the pairwise interactions  
 100 visualized in the  $\Delta$  matrix, and can be 0 or 1, for presence-absence of a given  
 101 pairwise interaction, or take a quantitative weight  $w_{ji}$  to represent the interaction  
 102 intensity or unidirectional effect of species  $j$  on species  $i$  (Bascompte & Jordano,  
 103 2014; Vazquez *et al.*, 2015). The outcomes of most ecological interactions are  
 104 dependent on frequency of encounters (e.g., visit rate of pollinators, number of  
 105 records of ant defenders, frequency of seeds in fecal samples). Thus, a frequently  
 106 used proxy for interaction intensities  $w_{ji}$  is just how frequent new interspecific

encounters are, whether or not appropriately weighted to estimate interaction effectiveness (Vazquez *et al.*, 2005).

We need to define two basic steps in the sampling of interactions: 1) which type of interactions we sample; and 2) which type of record we get to document the existence of an interaction. In step #1 we need to take into account whether we are sampling the whole community of interactor species (all the animals, all the plants) or just a subset of them, i.e., a sub matrix  $\Delta_{m,n}$  of  $m < A$  animal species and  $n < P$  plant species of the adjacency matrix  $\Delta_{AP}$  (i.e., the matrix representation of interactions among the partner species). Subsets can be: a) all the potential plants interacting with a subset of the animals (Fig. 1a); b) all the potential animal species interacting with a subset of the plant species (Fig. 1b); c) a subset of all the potential animal species interacting with a subset of all the plant species (Fig. 1c). While some discussion has considered how to establish the limits of what represents a network (Strogatz, 2001) (in analogy to discussion on food-web limits; Cohen, 1978), it must be noted that situations a-c in Fig. 1 do not represent complete interaction networks. Subnet sampling is generalized in studies of biological networks (e.g., protein interactions, gene regulation), yet it is important to recognize that most properties of subnetworks (even random subsamples) do not represent properties of whole networks (Stumpf *et al.*, 2005).

In step #2 above we face the problem of the type of record we take to sample interactions. This is important because it defines whether we approach the problem of filling up the interaction matrix in a “zoo-centric” way or in a “phyto-centric” way. Zoo-centric studies directly sample animal activity and document the plants ‘touched’ by the animal. For example, analysis of pollen samples recovered from the body of pollinators, analysis of fecal samples of frugivores, radio-tracking data, etc.



Phyto-centric studies take samples of focal individual plant species and document which animals ‘arrive’ or ‘touch’ the plants. Examples include focal watches of fruiting or flowering plants to record visitation by animals, raising insect herbivores from seed samples, identifying herbivory marks in samples of leaves, etc.

Most recent analyses of plant-animal interaction networks are phyto-centric; just 3.5% of available plant-pollinator ( $N=58$ ) or 36.6% plant-frugivore ( $N=22$ ) interaction datasets are zoo-centric (see Schleuning *et al.*, 2012). Moreover, most available datasets on host-parasite (parasitoid) or plant-herbivore interactions are “host-centric” or phyto-centric (e.g., Thébault & Fontaine, 2010; Morris *et al.*, 2013; Eklöf *et al.*, 2013). This may be related to a variety of causes, like preferred methodologies by researchers working with a particular group or system, logistic limitations, or inherent taxonomic focus of the research questions. A likely result of phyto-centric sampling would be adjacency matrices with large  $A : P$  ratios. In contrast, zoo-centric samplings might be prone to detect plants from outside the habitat, complicating the definition of network boundaries. In any case we don’t have a clear view of the potential biases that taxa-focused sampling may generate in observed network patterns, for example by generating consistently asymmetric interaction matrices (Dormann *et al.*, 2009).

Reasonably complete analyses of interaction networks can be obtained when combining both phyto-centric and zoo-centric sampling. For example, Bosch *et al.* (2009) showed that the addition of pollen load data on top of focal-plant sampling of pollinators unveiled a significant number of interactions, resulting in important network structural changes. Olesen *et al.* (2011) identified pollen loads on sampled insects and added the new links to an observation-based visitation matrix, with an extra 5% of links representing the estimated number of missing links

in the pollination network. The overlap between observational and pollen-load recorded links was only 33%, underscoring the value of combining methodological approaches. Zoo-centric sampling has recently been extended with the use of DNA-barcoding, for example with plant-herbivore (Jurado-Rivera *et al.*, 2009), host-parasitoid (Wirta *et al.*, 2014), and plant-frugivore interactions (González-Varo *et al.*, 2014). For mutualistic networks we would expect that zoo-centric sampling could help unveiling interactions of the animals with rare plant species or for relatively common plants species which are difficult to sample by direct observation. Future methodological work may provide significant advances showing how mixing different sampling strategies strengthens the completeness of network data. These mixed strategies may combine, for instance, timed watches at focal plants, spot censuses along walked transects, pollen load or seed contents analyses, monitoring with camera traps, and DNA barcoding records. However, there are no tested protocols and/or sampling designs for ecological interaction studies to suggest an optimum combination of approaches. Ideally, pilot studies would provide adequate information for each specific study setting.

## Sampling interactions: rationale

The number of distinct pairwise interactions that we can record in a landscape (an area of relatively homogeneous vegetation) is equivalent to the number of distinct classes in which we can classify the recorded encounters among *individuals* of two different species. Yet, individual-based interaction networks have been only recently studied (Dupont *et al.*, 2011; Wells & O'Hara, 2012). The most usual approach has been to pool individual-based interaction data into species-based

summaries, an approach that ignores the fact that only a fraction of individuals may actually interact given a per capita interaction effect (Wells & O’Hara, 2012). Wells & O’Hara (2012) illustrate the pros and cons of the approach. We walk in the forest and see a blackbird  $Tm$  picking an ivy  $Hh$  fruit and ingesting it: we have a record for  $Tm - Hh$  interaction. We keep advancing and record again a blackbird feeding on hawthorn  $Cm$  fruits so we record a  $Tm - Cm$  interaction; as we advance we encounter another ivy plant and record a blackcap swallowing a fruit so we now have a new  $Sa - Hh$  interaction, and so on. At the end we have a series of classes (e.g.,  $Sa - Hh$ ,  $Tm - Hh$ ,  $Tm - Cm$ , etc.), along with their observed frequencies.

We get a vector  $c = [c_1 \dots c_n]'$  where  $c_j$  is the number of classes represented  $j$  times in our sampling:  $c_1$  is the number of singletons (interactions recorded once),  $c_2$  is the number of twin pairs (interactions with just two records),  $c_3$  the number of triplets, etc. The problem thus turns to be estimating the number of distinct classes  $C$  from the vector of  $c_j$  values and the frequency of unobserved interactions (see “The real missing links” below).

More specifically, we usually obtain a type of reference sample (Chao *et al.*, 2014) for interactions: a series of repeated samples (e.g., observation days, 1h watches, etc.) with quantitative information, i.e., recording the number of instances of each interaction type on each day. This replicated abundance data, can be treated in three ways: 1) Abundance data within replicates: the counts of interactions, separately for each day; 2) Pooled abundance data: the counts of interactions, summed over all days (the most usual approach); and 3) Replicated incidence data: the number of days on which we recorded each interaction. Assuming a reasonable number of replicates, replicated incidence data is considered

to be the most robust statistically, as it takes account of heterogeneity among days (Colwell *et al.*, 2004, 2012; Chao *et al.*, 2014). Thus, both presence-absence and weighted information on interactions can be accommodated for this purpose.

## The species assemblage

When we consider an observed and recorded sample of interactions on a particular assemblage of  $A_{obs}$  and  $P_{obs}$  species (or a set of replicated samples) as a reference sample (Chao *et al.*, 2014) we may have three sources of undersampling error. These sources are ignored if we treat the reference sample as a true representation of the interactions in a well-defined assemblage: 1) some animal species are actually present but not observed (zero abundance or incidence in the interactions in the reference sample),  $A_0$ ; 2) some plant species are actually present but not observed (zero abundance or incidence in the interactions in the reference sample),  $P_0$ ; 3) some unobserved links (the zeroes in the adjacency matrix,  $UL$ ) may actually occur but not recorded. Thus a first problem is determining if  $A_{obs}$  and  $P_{obs}$  truly represent the actual species richness interacting in the assemblage. To this end we might use the replicated reference samples to estimate the true number of interacting animal  $A_{est}$  and plant  $P_{est}$  species as in traditional diversity estimation analysis (Chao *et al.*, 2014). If there are no uniques (species seen on only one day), then  $A_0$  and  $P_0$  will be zero (based on the Chao2 formula), and we have  $A_{obs}$  and  $P_{obs}$  as robust estimates of the actual species richness of the assemblage. If  $A_0$  and  $P_0$  are not zero they estimate the minimum number of undetected animal and plant species that can be expected with a sufficiently large number of replicates, taken from the same assemblage/locality by the same methods in the same time

228 period. We can use extrapolation methods (Colwell *et al.*, 2012) to estimate how  
 229 many additional replicate surveys it would take to reach a specified proportion  $g$   
 230 of  $A_{est}$  and  $P_{est}$ .

## 231 The interactions

232 We are then faced with assessing the sampling of interactions  $I$ . Table 1 summa-  
 233 rizes the main components and targets for estimation of interaction richness. In  
 234 contrast with traditional species diversity estimates, sampling networks has the  
 235 paradox that despite the potentially interacting species being present in the sam-  
 236 pled assemblage (i.e., included in the  $A_{obs}$  and  $P_{obs}$  species lists), some of their  
 237 pairwise interactions are impossible to record. The reason is forbidden links. In-  
 238 dependently of whether we sample full communities or subset communities we face  
 239 a problem: some of the interactions that we can visualize in the empty adjacency  
 240 matrix  $\Delta$  will simply not occur. With a total of  $A_{obs}P_{obs}$  “potential” interactions  
 241 (eventually augmented to  $A_{est}P_{est}$  in case we have undetected species), a fraction  
 242 of them are impossible to record, because they are forbidden (Jordano *et al.*, 2003;  
 243 Olesen *et al.*, 2011).

244 Our goal is to estimate the true number of non-null  $AP$  interactions, including  
 245 interactions that actually occur but have not been observed ( $I_0$ ) from the repli-  
 246 cated incidence frequencies of interaction types:  $I_{est} = I_{obs} + I_0$ . Note that  $I_0$   
 247 estimates the minimum number of undetected plant-animal interactions that can  
 248 be expected with a sufficiently large number of replicates, taken from the same  
 249 assemblage/locality by the same methods in the same time period. Therefore we  
 250 have two types of non-observed links:  $UL^*$  and  $UL$ , corresponding to the real

251 assemblage species richness and to the observed assemblage species richness, re-  
 252 spectively (Table 1).

253 Forbidden links are non-occurrences of pairwise interactions that can be ac-  
 254 counted for by biological constraints, such as spatio-temporal uncoupling (Jor-  
 255 dano, 1987), size or reward mismatching, foraging constraints (e.g., accessibility)  
 256 (Moré *et al.*, 2012), and physiological-biochemical constraints (Jordano, 1987). We  
 257 still have very little information about the frequency of forbidden links in natu-  
 258 ral communities (Jordano *et al.*, 2003; Stang *et al.*, 2009; Vázquez *et al.*, 2009;  
 259 Olesen *et al.*, 2011; Ibanez, 2012; Maruyama *et al.*, 2014; Vizentin-Bugoni *et al.*,  
 260 2014) (Table 1). Forbidden links are thus represented as structural zeroes in the  
 261 interaction matrix, i.e., matrix cells that cannot get a non-zero value. We might  
 262 expect different types of *FL* to occupy different parts of the  $\Delta$  matrix, with missing  
 263 cells due to phenological uncoupling, *FL<sub>P</sub>*, largely distributed in the lower-right  
 264 half  $\Delta$  matrix and actually missed links *ML* distributed in its central part (Ole-  
 265 sen *et al.*, 2010). Yet, most of these aspects remain understudied. Therefore, we  
 266 need to account for the frequency of these structural zeros in our matrix before  
 267 proceeding.

268 Our main problem then turns to estimate the number of true missed links,  
 269 i.e., those that can't be accounted for by biological constraints and that might  
 270 suggest undersampling. Thus, the sampling of interactions in nature, as the sam-  
 271 pling of species, is a cumulative process. In our analysis, we are not re-sampling  
 272 individuals, but interactions, so we made interaction-based accumulation curves.  
 273 If an interaction-based curve suggests a robust sampling, it does mean that no  
 274 new interactions are likely to be recorded, irrespectively of the species, as it is a  
 275 whole-network sampling approach (N. Gotelli, pers. com.). We add new, distinct,

interactions recorded as we increase sampling effort (Fig. 2). We can obtain an Interaction Accumulation Curve (*IAC*) analogous to a Species Curve (*SAC*) (see Supporting Information in the online data availability repository): the observed number of distinct pairwise interactions in a survey or collection as a function of the accumulated number of observations or samples (Colwell, 2009).

## Empirical data on Forbidden Links

Adjacency matrices are frequently sparse, i.e., they are densely populated with zeroes, with a fraction of them being structural (unobservable interactions) (Bascompte & Jordano, 2014). Thus, it would be a serious interpretation error to attribute the sparseness of adjacency matrices for bipartite networks to just the result of undersampling. The actual typology of link types in ecological interaction networks is thus more complex than just the two categories of observed and unobserved interactions (Table 1). Unobserved interactions are represented by zeroes and belong to two categories. Missing interactions may actually exist but require additional sampling or a variety of methods to be observed. Forbidden links, on the other hand, arise due to biological constraints limiting interactions and remain unobservable in nature, irrespectively of sampling effort (Table 1). Forbidden links *FL* may actually account for a relatively large fraction of unobserved interactions *UL* when sampling taxonomically-restricted subnetworks (e.g., plant-hummingbird pollination networks) (Table 1). Phenological uncoupling is also prevalent in most networks, and may add up to explain ca. 25-40% of the forbidden links, especially in highly seasonal habitats, and up to 20% when estimated relative to the total number of unobserved interactions (Table 2). In any case, we

might expect that a fraction of the missing links  $ML$  would be eventually explained by further biological reasons, depending on the knowledge of natural details of the particular systems. Our goal as naturalists would be to reduce the fraction of  $UL$  which remain as missing links; to this end we might search for additional biological constraints or increase sampling effort. For instance, habitat use patterns by hummingbirds in the Arima Valley network (Table 2; Snow & Snow, 1972) impose a marked pattern of microhabitat mismatches causing up to 44.5% of the forbidden links. A myriad of biological causes beyond those included as  $FL$  in Table 1 may contribute explanations for  $UL$ : limits of color perception, presence of secondary metabolites in fruit pulp and leaves, toxins and combinations of monosaccharides in nectar, etc. For example, aside from  $FL$ , some pairwise interactions may simply have an asymptotically-zero probability of interspecific encounter between the partner species, if they are very rare. However, it is surprising that just the limited set of forbidden link types considered in Table 1 explain between 24.6-77.2% of the unobserved links. Notably, the Arima Valley, Santa Virgínia, and Hato Ratón networks have  $> 60\%$  of the unobserved links explained, which might be related to the fact that they are subnetworks (Arima Valley, Santa Virgínia) or relatively small networks (Hato Ratón). All this means that empirical networks may have sizable fractions of structural zeroes. Ignoring this biological fact may contribute to wrongly inferring undersampling of interactions in real-world assemblages.

To sum up, two elements of inference are required in the analysis of unobserved interactions in ecological interaction networks: first, detailed natural history information on the participant species that allows the inference of biological constraints imposing forbidden links, so that structural zeroes can be identified in the adjacency matrix. Second, a critical analysis of sampling robustness and a robust



estimate of the actual fraction of missing links,  $M$ , resulting in a robust estimate of  $I$ . In the next sections we explore these elements of inference, using *IACs* as analogs to *SACs* to assess the robustness of interaction sampling.

## Asymptotic diversity estimates

Let's assume a sampling of the diversity in a specific locality, over relatively homogeneous landscape where we aim at determining the number of species present for a particular group of organisms. To do that we carry out transects or plot samplings across the landscape or use any other type of direct or indirect recording method, adequately replicated so we obtain a number of samples. Briefly,  $S_{obs}$  is the total number of species observed in a sample, or in a set of samples.  $S_{est}$  is the estimated number of species in the community represented by the sample, or by the set of samples, where *est* indicates an estimator. With abundance data, let  $S_k$  be the number of species each represented by exactly  $k$  individuals in a single sample. Thus,  $S_0$  is the number of undetected species (species present in the community but not included in the sample),  $S_1$  is the number of singleton species (represented by just one individual),  $S_2$  is the number of doubleton species (species with two individuals), etc. The total number of individuals in the sample would be:

$$n = \sum_{k=1}^{S_{obs}} S_k$$

A frequently used asymptotic, bias corrected, non-parametric estimator is  $S_{Chao1}$

344 (Hortal *et al.*, 2006; Chao, 2005; Colwell, 2013):

$$S_{Chao1} = S_{obs} + \frac{S_1(S_1 - 1)}{2(S_2 + 1)}$$

345 Another frequently used alternative is the Chao2 estimator,  $S_{Chao2}$  (Gotelli &  
346 Colwell, 2001), which has been reported to have a limited bias for small sample  
347 sizes (Colwell & Coddington, 1994; Chao, 2005). Instead of using counts it uses  
348 incidence frequencies ( $Q_k$ ) among samples (number of species present in just one  
349 sample, in two samples, etc.):

$$S_{Chao2} = S_{obs} + \frac{Q_1(Q_1 - 1)}{2(Q_2 + 1)}$$

350 A plot of the cumulative number of species recorded,  $S_n$ , as a function of some  
351 measure of sampling effort (say,  $n$  samples taken) yields the species accumulation  
352 curve (SAC) or collector's curve (Colwell & Coddington, 1994). Similarly, inter-  
353 action accumulation curves (IAC), analogous to SACs, can be used to assess the  
354 robustness of interactions sampling for plant-animal community datasets (Jordano,  
355 1987; Jordano *et al.*, 2009; Olesen *et al.*, 2011), as discussed in the next section.

## 356 Assessing sampling effort when recording interac- 357 tions

358 The basic method we can propose to estimate sampling effort and explicitly show  
359 the analogues with rarefaction analysis in biodiversity research is to vectorize the  
360 interaction matrix  $AP$  so that we get a vector of all the potential pairwise interac-  
361 tions ( $I_{max}$ , Table 1) that can occur in the observed assemblage with  $A_{obs}$  animal

362 species and  $P_{obs}$  plant species. The new “species” we aim to sample are the pairwise  
 363 interactions (Table 3). So, if we have in our community *Turdus merula* ( $Tm$ ) and  
 364 *Rosa canina* ( $Rc$ ) and *Prunus mahaleb* ( $Pm$ ), our problem will be to sample 2 new  
 365 “species”:  $Tm - Rc$  and  $Tm - Pm$ . In general, if we have  $A = 1...i$ , animal species  
 366 and  $P = 1...j$  plant species (assuming a complete list of species in the assemblage),  
 367 we’ll have a vector of “new” species to sample:  $A_1P_1, A_1P_2, ...A_2P_1, A_2P_2, ...A_iP_j$ .  
 368 We can represent the successive samples where we can potentially get records of  
 369 these interactions in a matrix with the vectorized interaction matrix and columns  
 370 representing the successive samples we take (Table 3). This is simply a vectorized  
 371 version of the interaction matrix. This is analogous to a biodiversity sampling  
 372 matrix with species as rows and sampling units (e.g., quadrats) as columns (Jor-  
 373 dano *et al.*, 2009). The package *EstimateS* (Colwell, 2013) includes a complete set  
 374 of functions for estimating the mean IAC and its unconditional standard devia-  
 375 tion from random permutations of the data, or subsampling without replacement  
 376 (Gotelli & Colwell, 2001); it further reports asymptotic estimators for the expected  
 377 number of distinct pairwise interactions included in a given reference sample of  
 378 interaction records (see also the `specaccum` function in library `vegan` of the R  
 379 Package)(R Development Core Team, 2010; Jordano *et al.*, 2009; Olesen *et al.*,  
 380 2011). In particular, we may take advantage of replicated incidence data, as it  
 381 takes account of heterogeneity among samples (days, censuses, etc.; R.K Colwell,  
 382 pers. comm.) (see also Colwell *et al.*, 2004, 2012; Chao *et al.*, 2014).

383 In this way we effectively extend sampling theory developed for species diversity  
 384 to the sampling of ecological interactions. Yet future theoretical work will be  
 385 needed to formally assess the similarities and differences in the two approaches  
 386 and developing biologically meaningful null models of expected interaction richness

with added sampling effort.

Diversity-accumulation analysis (Magurran, 1988; Hortal *et al.*, 2006) comes up immediately with this type of dataset. This procedure plots the accumulation curve for the expected number of distinct pairwise interactions recorded with increasing sampling effort (Jordano *et al.*, 2009; Olesen *et al.*, 2011). Asymptotic estimates of interaction richness and its associated standard errors and confidence intervals can thus be obtained (Hortal *et al.*, 2006) (see Table 4 and Supplementary Online Material). The characteristic feature of interaction datasets is that, due to forbidden links, a number of pairwise interactions among the  $I_{max}$  number specified in the  $\Delta$  adjacency matrix cannot be recorded, irrespective of sampling effort.

We may expect undersampling specially in moderate to large sized networks with multiple modules (i.e., species subsets requiring different sampling strategies) (Jordano, 1987; Olesen *et al.*, 2011; Chacoff *et al.*, 2012); adequate sampling may be feasible when interaction subwebs are studied (Olesen *et al.*, 2011; Vizentin-Bugoni *et al.*, 2014), typically with more homogeneous subsets of species (e.g., bumblebee-pollinated flowers). In any case the sparseness of the  $\Delta$  matrix is by no means an indication of undersampling whenever the issue of structural zeroes in the interaction matrices is effectively incorporated in the estimates.

For example, mixture models incorporating detectabilities have been proposed to effectively account for rare species (Mao & Colwell, 2005). In an analogous line, mixture models could be extended to samples of pairwise interactions, also with specific detectability values. These detection rate/odds could be variable among groups of interactions, depending on their specific detectability. For example, detectability of flower-pollinator interactions involving bumblebees could have a

412 higher detectability than flower-pollinator pairwise interactions involving, say, ni-  
 413 tidulid beetles. These more homogeneous groupings of pairwise interactions within  
 414 a network define modules (Bascompte & Jordano, 2014), so we might expect that  
 415 interactions of a given module (e.g., plants and their hummingbird pollinators; Fig.  
 416 1a) may share similar detectability values, in an analogous way to species groups  
 417 receiving homogeneous detectability values in mixture models (Mao & Colwell,  
 418 2005). In its simplest form, this would result in a sample with multiple pairwise  
 419 interactions detected, in which the number of interaction events recorded for each  
 420 distinct interaction found in the sample is recorded (i.e., a column vector in Table  
 421 3, corresponding to, say, a sampling day). The number of interactions recorded for  
 422 the  $i_{th}$  pairwise interaction (i.e.,  $A_i P_j$  in Table 3),  $Y_i$  could be treated as a Poisson  
 423 random variable with a mean parameter  $\lambda_i$ , its detection rate. Mixture models  
 424 (Mao & Colwell, 2005) include estimates for abundance-based data (their analogs  
 425 in interaction sampling would be weighted data), where  $Y_i$  is a Poisson random  
 426 variable with detection rate  $\lambda_i$ . This is combined with the incidence-based model,  
 427 where  $Y_i$  is a binomial random variable (their analogous in interaction sampling  
 428 would be presence/absence records of interactions) with detection odds  $\lambda_i$ . Let  
 429  $T$  be the number of samples in an incidence-based data set. A Poisson/binomial  
 430 density can be written as (Mao & Colwell, 2005):

$$g(y; \lambda) = \begin{cases} \frac{\lambda^y}{y! e^\lambda} & [1] \\ \binom{T}{y} \frac{\lambda^y}{(1+\lambda)^T} & [2] \end{cases}$$

431 where [1] corresponds to a weighted network, and [2] to a qualitative network.  
 432 The detection rates  $\lambda_i$  depend on the relative abundances  $\phi_i$  of the interactions,

433 the probability of a pairwise interaction being detected when it is present, and the  
 434 sample size (the number of interactions recorded), which, in turn, is a function  
 435 of the sampling effort. Unfortunately, no specific sampling model has been devel-  
 436 oped along these lines for species interactions and their characteristic features. For  
 437 example, a complication factor might be that interaction abundances,  $\phi_i$ , in real  
 438 assemblages are a function of the abundances of interacting species that determine  
 439 interspecific encounter rates; yet they also depend on biological factors that ulti-  
 440 mately determine if the interaction occurs when the partner species are present.  
 441 For example,  $\lambda_i$  should be set to zero for all  $FL$ . In its simplest form,  $\phi_i$  could be  
 442 estimated from just the product of partner species abundances, an approach re-  
 443 cently used as a null model to assess the role of biological constraints in generating  
 444 forbidden links and explaining interaction patterns (Vizentin-Bugoni *et al.*, 2014).  
 445 Yet more complex models (e.g., Wells & O'hara 2012) should incorporate not  
 446 only interspecific encounter probabilities, but also interaction detectabilities, phe-  
 447 notypic matching and incidence of forbidden links. Mixture models are certainly  
 448 complex and for most situations of evaluating sampling effort better alternatives  
 449 include the simpler incidence-based rarefaction and extrapolation (Colwell *et al.*,  
 450 2012; Chao *et al.*, 2014).

## 451 The *real* missing links

452 Given that a fraction of unobserved interactions can be accounted for by forbidden  
 453 links, what about the remaining missing interactions? We have already discussed  
 454 that some of these could still be related to unaccounted constraints, and still others  
 455 would be certainly attributable to insufficient sampling. Would this always be the

case? A crucial ecological aspect limiting interactions within multispecific assemblages of distinct taxonomic relatedness (Fig. 2) is the probability of interspecific encounter, i.e., the probability that two individuals of the partner species actually encounter each other in nature.

Given log-normally distributed abundances of the two species groups, the expected probabilities of interspecific encounter ( $PIE$ ) would be simply the product of the two lognormal distributions. Thus, we might expect that for very low  $PIE$  values, pairwise interactions would be either extremely difficult to sample, or simply do not occur in nature. Consider the Nava de las Correhuelas interaction web (NCH, Table 2, 4), with  $A = 36$ ,  $P = 25$ ,  $I = 181$ , and almost half of the unobserved interactions not accounted for by forbidden links, thus  $M = 53.1\%$  (Jordano *et al.*, 2009). A sizable fraction of these possible but missing links would be simply not occurring in nature, most likely due to extremely low  $PIE$ , in fact asymptotically zero. Given the vectorized list of pairwise interactions for NCH, I computed the  $PIE$  values for each one by multiplying element-wise the two species abundance distributions. The  $PIE_{max} = 0.0597$ , being a neutral estimate, based on the assumption that interactions occur in proportion to the species-specific local abundances. With  $PIE_{median} < 1.4 \cdot 10^{-4}$  we may safely expect (note the quantile estimate  $Q_{75\%} = 3.27 \cdot 10^{-4}$ ) that a sizable fraction of these missing interactions may not occur according to this neutral expectation (Jordano, 1987; Olesen *et al.*, 2011) (neutral forbidden links, *sensu* Canard *et al.*, 2012).

When we consider the vectorized interaction matrix, enumerating all pairwise interactions for the  $AP$  combinations, the expected probabilities of finding a given interaction can be estimated with a Good-Turing approximation (Good, 1953). The technique, developed by Alan Turing and I.J. Good with applications to lin-

481 guistics and word analysis (Gale & Sampson, 1995) has been recently extended in  
 482 novel ways for ecological analyses (Chao *et al.*, 2015). It estimates the probability  
 483 of recording an interaction of a hitherto unseen pair of partners, given a set of past  
 484 records of interactions between other species pairs. Let a sample of  $N$  interactions  
 485 so that  $n_r$  distinct pairwise interactions have exactly  $r$  records. All Good-Turing  
 486 estimators obtain the underlying frequencies of events as:

$$P(X) = \frac{(N_X + 1)}{T} \left(1 - \frac{E(1)}{T}\right) \quad (1)$$

487 where  $X$  is the pairwise interaction,  $N_X$  is the number of times interaction  $X$   
 488 is recorded,  $T$  is the sample size (number of distinct interactions recorded) and  
 489  $E(1)$  is an estimate of how many different interactions were recorded exactly once.  
 490 Strictly speaking Equation (1) gives the probability that the next interaction type  
 491 recorded will be  $X$ , after sampling a given assemblage of interacting species. In  
 492 other words, we scale down the maximum-likelihood estimator  $\frac{n}{T}$  by a factor of  
 493  $\frac{1-E(1)}{T}$ . This reduces all the probabilities for interactions we have recorded, and  
 494 makes room for interactions we haven't seen. If we sum over the interactions we  
 495 have seen, then the sum of  $P(X)$  is  $1 - \frac{1-E(1)}{T}$ . Because probabilities sum to one,  
 496 we have the left-over probability of  $P_{new} = \frac{E(1)}{T}$  of seeing something new, where  
 497 new means that we sample a new pairwise interaction.

## 498 Discussion

499 Recent work has inferred that most data available for interaction networks are  
 500 incomplete due to undersampling, resulting in a variety of biased parameters and



501 network patterns (Chacoff *et al.*, 2012). It is important to note, however, that in  
 502 practice, most surveyed networks to date have been subnets of much larger net-  
 503 works. This is also true for protein interaction, gene regulation, and metabolic  
 504 networks, where only a subset of the molecular entities in a cell have been sam-  
 505 pled (Stumpf *et al.*, 2005). Despite recent attempts to document whole ecosystem  
 506 meta-networks (Pocock *et al.*, 2012), it is likely that most ecological interaction  
 507 networks will illustrate just major ecosystem compartments. Due to their high  
 508 generalization, high temporal and spatial turnover, and high complexity of associ-  
 509 ation patterns, adequate sampling of ecological interaction networks is challenging  
 510 and requires extremely large sampling effort. Undersampling of ecological net-  
 511 works may originate from the analysis of assemblage subsets (e.g., taxonomically  
 512 or functionally defined), and/or from logistically-limited sampling effort. It is ex-  
 513 tremely hard to robustly sample the set of biotic interactions even for relatively  
 514 simple, species-poor assemblages; thus, we need to assess how robust is the char-  
 515 acterization of the adjacency matrix  $\Delta$ . Concluding that an ecological network  
 516 dataset is undersampled just by its sparseness would be unrealistic. The reason  
 517 stems from a biological fact: a sizeable fraction of the maximum, potential links  
 518 that can be recorded among two distinct sets of species is simply unobservable,  
 519 irrespective of sampling effort (Jordano, 1987). In addition, sampling effort needs  
 520 to be explicitly gauged because of its potential influence on parameter estimates  
 521 for the network.

522 Missing links are a characteristic feature of all plant-animal interaction net-  
 523 works, and likely pervade other ecological interactions. Important natural history  
 524 details explain a fraction of them, resulting in unrealizable interactions (i.e., for-  
 525 bidden interactions) that define structural zeroes in the interaction matrices and

contribute to their extreme sparseness. Sampling interactions is a way to monitor biodiversity beyond the simple enumeration of component species and to develop efficient and robust inventories of functional interactions. Yet no sampling theory for interactions is available. Focusing just on the realized interactions or treating missing interactions as the expected unique result of sampling bias would miss important components to understand how all sorts of interactions coevolve within complex webs of interdependence among species.

Contrary to species inventories, a sizable fraction of non-observed pairwise interactions cannot be sampled, due to biological constraints that forbid their occurrence. Moreover, recent implementations of inference methods for unobserved species (Chao *et al.*, 2015) or for individual-based data (Wells & O’Hara, 2012) can be combined with the forbidden link approach. They do not account either for the existence of these ecological constraints, but can help in estimating their relative importance, simply by the difference between the asymptotic estimate of interaction richness *in a robustly-sampled* assemblage and the maximum richness  $I_{max}$  of interactions.

Ecological interactions provide the wireframe supporting the lives of species, and they also embed crucial ecosystem functions which are fundamental for supporting the Earth system. We still have a limited knowledge of the biodiversity of ecological interactions, and they are being lost (extinct) at a very fast pace, frequently preceding species extinctions (Valiente-Banuet *et al.*, 2014). We urgently need robust techniques to assess the completeness of ecological interactions networks because this knowledge will allow the identification of the minimal components of their ecological complexity that need to be restored to rebuild functional ecosystems after perturbations.

## Acknowledgements

I am indebted to Jens M. Olesen, Alfredo Valido, Jordi Bascompte, Thomas Lewinshon, John N. Thompson, Nick Gotelli, Carsten Dormann, and Paulo R. Guimaraes Jr. for useful and thoughtful discussion at different stages of this manuscript. Jeferson Vizentin-Bugoni kindly helped with the Sta Virginia data. Jens M. Olesen kindly made available the Grundvad dataset; together with Robert K. Colwell, Néstor Pérez-Méndez, JuanPe González-Varo, and Paco Rodríguez provided most useful comments to a final version of the ms. Robert Colwell shared a number of crucial suggestions that clarified my vision of sampling ecological interactions, and the final manuscript was greatly improved with comments from three anonymous reviewers. The study was supported by a Junta de Andalucía Excellence Grant (RNM-5731), as well as a Severo Ochoa Excellence Award from the Ministerio de Economía y Competitividad (SEV-2012-0262). The Agencia de Medio Ambiente, Junta de Andalucía, provided generous facilities that made possible my long-term field work in different natural parks.

## Data accessibility

This review does not use new raw data, but includes some re-analyses of previously published material. All the original data supporting the paper, R code, supplementary figures, and summaries of analytical protocols is available at the author's GitHub repository ([https://github.com/pedroj/MS\\_Network-Sampling](https://github.com/pedroj/MS_Network-Sampling)), with DOI: 10.5281/zenodo.29437.

## References

- Banasek-Richter, C., Cattin, M. & Bersier, L. (2004) Sampling effects and the robustness of quantitative and qualitative food-web descriptors. *Journal of Theoretical Biology* **226**, 23–32.
- Bascompte, J. & Jordano, P. (2014) *Mutualistic networks*. Monographs in Population Biology, No. 53, Princeton University Press, Princeton, NJ.
- Bersier, L., Banasek-Richter, C. & Cattin, M. (2002) Quantitative descriptors of food-web matrices. *Ecology* **83**, 2394–2407.
- Bosch, J., Martín González, A.M., Rodrigo, A. & Navarro, D. (2009) Plant-pollinator networks: adding the pollinator’s perspective. *Ecology Letters* **12**, 409–419.
- Brose, U., Martinez, N. & Williams, R. (2003) Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* **84**, 2364–2377.
- Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012) Emergence of structural patterns in neutral trophic networks. *PLoS ONE* **7**, e38295.
- Chacoff, N.P., Vazquez, D.P., Lomascolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012) Evaluating sampling completeness in a desert plant-pollinator network. *Journal of Animal Ecology* **81**, 190–200.
- Chao, A. (2005) Species richness estimation. *Encyclopedia of Statistical Sciences*, pp. 7909–7916, Oxford University Press, New York, USA.

- 594 Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Elli-  
595 son, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework  
596 for sampling and estimation in species diversity studies. *Ecological Monographs*  
597 **84**, 45–67.
- 598 Chao, A., Hsieh, T.C., Chazdon, R.L., Colwell, R.K. & Gotelli, N.J. (2015) Un-  
599 veiling the species-rank abundance distribution by generalizing the Good-Turing  
600 sample coverage theory. *Ecology* **96**, 1189–1201.
- 601 Coddington, J.A., Agnarsson, I., Miller, J.A., Kuntner, M. & Hormiga, G. (2009)  
602 Undersampling bias: the null hypothesis for singleton species in tropical arthro-  
603 pod surveys. *Journal of Animal Ecology* **78**, 573–584.
- 604 Cohen, J.E. (1978) *Food webs and niche space*. Princeton University Press, Prince-  
605 ton, New Jersey, US.
- 606 Colwell, R. & Coddington, J. (1994) Estimating terrestrial biodiversity through ex-  
607 trapolation. *Philosophical Transactions Of The Royal Society Of London Series*  
608 *B-Biological Sciences* **345**, 101–118.
- 609 Colwell, R.K. (2009) Biodiversity: concepts, patterns, and measurement. *The*  
610 *Princeton Guide to Ecology* (ed. S.A. Levin), pp. 257–263, Princeton Univer-  
611 sity Press, Princeton.
- 612 Colwell, R.K. (2013) EstimateS: Biodiversity Estimation. - pp. 1–33.
- 613 Colwell, R.K., Dunn, R.R. & Harris, N.C. (2012) Coextinction and persistence of  
614 dependent species in a changing world. *Annual Review of Ecology Evolution and*  
615 *Systematics* **43**, 183–203.

- 616 Colwell, R.K., Mao, C.X. & Chang, J. (2004) Interpolating, extrapolating, and  
617 comparing incidence-based species accumulation curves. *Ecology* **85**, 2717–2727.
- 618 Dorado, J., Vazquez, D.P., Stevani, E.L. & Chacoff, N.P. (2011) Rareness and  
619 specialization in plant-pollinator networks. *Ecology* **92**, 19–25.
- 620 Dormann, C.F., Frund, J., Bluthgen, N. & Gruber, B. (2009) Indices, graphs and  
621 null models: Analyzing bipartite ecological networks. *Open Ecology Journal* **2**,  
622 7–24.
- 623 Dupont, Y.L., Trøjelsgaard, K., Hagen, M., Henriksen, M.V., Olesen, J.M., Ped-  
624 ersen, N.M.E. & Kissling, W.D. (2014) Spatial structure of an individual-based  
625 plant-pollinator network. *Oikos* **123**, 1301–1310.
- 626 Dupont, Y.L., Trøjelsgaard, K. & Olesen, J.M. (2011) Scaling down from species  
627 to individuals: a flower-visitation network between individual honeybees and  
628 thistle plants. *Oikos* **120**, 170–177.
- 629 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P.,  
630 Dalsgaard, B., de Sassi, C., Galetti, M., Guimaraes, P.R., Lomáscolo, S.B.,  
631 Martín González, A.M., Pizo, M.A., Rader, R., Rodrigo, A., Tylianakis, J.M.,  
632 Vazquez, D.P. & Allesina, S. (2013) The dimensionality of ecological networks.  
633 *Ecology Letters* **16**, 577–583.
- 634 Frund, J., McCann, K.S. & Williams, N.M. (2015) Sampling bias is a challenge  
635 for quantifying specialization and network structure: lessons from a quantitative  
636 niche model. *Oikos* pp. n/a–n/a.

- 637 Gale, W.A. & Sampson, G. (1995) Good-Turing frequency estimation without  
638 tears. *Journal of Quantitative Linguistics* **2**, 217–237.
- 639 Gibson, R.H., Knott, B., Eberlein, T. & Memmott, J. (2011) Sampling method  
640 influences the structure of plant–pollinator networks. *Oikos* **120**, 822–831.
- 641 González-Varo, J.P., Arroyo, J.M. & Jordano, P. (2014) Who dispersed the seeds?  
642 The use of DNA barcoding in frugivory and seed dispersal studies. *Methods in*  
643 *Ecology and Evolution* **5**, 806–814.
- 644 Good, I.J. (1953) The population frequencies of species and the estimation of  
645 population parameters. *Biometrika* **40**, 237–264.
- 646 Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: procedures and pitfalls  
647 in the measurement and comparison of species richness. *Ecology Letters* **4**, 379–  
648 391.
- 649 Gotelli, N.J. & Colwell, R.K. (2011) Estimating species richness. *Biological Di-*  
650 *versity Frontiers in Measurement and Assessment* (eds. A.E. Magurran & B.J.  
651 McGill), pp. 39–54, Oxford University Press, Oxford, UK.
- 652 Hortal, J., Borges, P. & Gaspar, C. (2006) Evaluating the performance of species  
653 richness estimators: sensitivity to sample grain size. *Journal of Animal Ecology*  
654 **75**, 274–287.
- 655 Ibanez, S. (2012) Optimizing size thresholds in a plant–pollinator interaction web:  
656 towards a mechanistic understanding of ecological networks. *Oecologia* **170**, 233–  
657 242.

- 658 Jordano, P. (1987) Patterns of mutualistic interactions in pollination and seed dis-  
 659 persal: connectance, dependence asymmetries, and coevolution. *The American*  
 660 *Naturalist* **129**, 657–677.
- 661 Jordano, P., Bascompte, J. & Olesen, J. (2003) Invariant properties in coevolu-  
 662 tionary networks of plant-animal interactions. *Ecology Letters* **6**, 69–81.
- 663 Jordano, P., Vázquez, D. & Bascompte, J. (2009) Redes complejas de interac-  
 664 ciones planta—animal. *Ecología y evolución de interacciones planta-animal* (eds.  
 665 R. Medel, R. Dirzo & R. Zamora), pp. 17–41, Editorial Universitaria, Santiago,  
 666 Chile.
- 667 Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. & Gomez-Zurita,  
 668 J. (2009) DNA barcoding insect-host plant associations. *Proceedings Of The*  
 669 *Royal Society B-Biological Sciences* **276**, 639–648.
- 670 Magurran, A. (1988) *Ecological diversity and its measurement*. Princeton Univer-  
 671 sity Press, Princeton, US.
- 672 Mao, C. & Colwell, R.K. (2005) Estimation of species richness: mixture models,  
 673 the role of rare species, and inferential challenges. *Ecology* **86**, 1143–1153.
- 674 Martinez, N. (1991) Artifacts or attributes? Effects of resolution on food-web  
 675 patterns in Little Rock Lake food web. *Ecological Monographs* **61**, 367–392.
- 676 Martinez, N.D. (1993) Effects of resolution on food web structure. *Oikos* **66**, 403–  
 677 412.
- 678 Maruyama, P.K., Vizentin-Bugoni, J., Oliveira, G.M., Oliveira, P.E. & Dalsgaard,



- 679 B. (2014) Morphological and spatio-temporal mismatches shape a neotropical  
680 savanna plant-hummingbird network. *Biotropica* **46**, 740–747.
- 681 Memmott, J., Alonso, D., Berlow, E., Dobson, A., Dunne, J.A., Solé, R.V. &  
682 Weitz, J.S. (2006) Biodiversity loss and ecological network structure. *Ecological*  
683 *Networks Linking Structure to Dynamics in Food Webs* (eds. M. Pascual & J.A.  
684 Dunne), pp. 325–347, Oxford University Press, Oxford, UK.
- 685 Moré, M., Amorim, F.W., Benitez-Vieyra, S., Medina, A.M., Sazima, M. &  
686 Cocucci, A.A. (2012) Armament Imbalances: Match and Mismatch in Plant-  
687 Pollinator Traits of Highly Specialized Long-Spurred Orchids. *PLoS ONE* **7**,  
688 e41878.
- 689 Morris, R.J., Gripenberg, S., Lewis, O.T. & Roslin, T. (2013) Antagonistic inter-  
690 action networks are structured independently of latitude and host guild. *Ecology*  
691 *Letters* **17**, 340–349.
- 692 Nielsen, A. & Bascompte, J. (2007) Ecological networks, nestedness and sampling  
693 effort. *Journal of Ecology* **95**, 1134–1141–1141.
- 694 Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H. & Jordano, P. (2011)  
695 Missing and forbidden links in mutualistic networks. *Proceedings Of The Royal*  
696 *Society B-Biological Sciences* **278**, 725–732.
- 697 Olesen, J.M., Dupont, Y.L., O’gorman, E., Ings, T.C., Layer, K., Melin, C.J.,  
698 Trjelsgaard, K., Pichler, D.E., Rasmussen, C. & Woodward, G. (2010) From  
699 Broadstone to Zackenberg. *Advances in Ecological Research* **42**, 1–69.
- 700 Olito, C. & Fox, J.W. (2014) Species traits and abundances predict metrics of

- 701 plant-pollinator network structure, but not pairwise interactions. *Oikos* **124**,  
702 428–436.
- 703 Ollerton, J. & Cranmer, L. (2002) Latitudinal trends in plant-pollinator interac-  
704 tions: are tropical plants more specialised? *Oikos* **98**, 340–350.
- 705 Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes,  
706 R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops,  
707 N., Dulloo, E., Faith, D., Freyhof, J., Gregory, R.D., Heip, C., Hoft, R., Hurtt,  
708 G., Jetz, W., Karp, D.S., Mcgeoch, M., Obura, D., Onoda, Y., Pettorelli, N.,  
709 Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S., Turak, E., Walpole, M.  
710 & Wegmann, M. (2013) Essential biodiversity variables. *Science* **339**, 277–278.
- 711 Pocock, M.J.O., Evans, D.M. & Memmott, J. (2012) The Robustness and Restora-  
712 tion of a Network of Ecological Networks. *Science* **335**, 973–977.
- 713 Preston, F. (1948) The commonness, and rarity, of species. *Ecology* **29**, 254–283.
- 714 R Development Core Team (2010) *R: A language and environment for statis-*  
715 *tical computing*. R Foundation for Statistical Computing. Vienna, Austria.  
716 <http://www.R-project.org>, Vienna, Austria.
- 717 Rivera-Hutinel, A., Bustamante, R.O., Marín, V.H. & Medel, R. (2012) Effects  
718 of sampling completeness on the structure of plant-pollinator networks. *Ecology*  
719 **93**, 1593–1603.
- 720 Schleuning, M., Frund, J., Klein, A.M., Abrahamczyk, S., Alarcón, R., Albrecht,  
721 M., Andersson, G.K.S., Bazarian, S., Böhning-Gaese, K., Bommarco, R., Dals-  
722 gaard, B., Dehling, D.M., Gotlieb, A., Hagen, M., Hickler, T., Holzschuh, A.,

- 723 Kaiser-Bunbury, C.N., Kreft, H., Morris, R.J., Sandel, B., Sutherland, W.J.,  
 724 Svenning, J.C., Tschardtke, T., Watts, S., Weiner, C.N., Werner, M., Williams,  
 725 N.M., Winqvist, C., Dormann, C.F. & Blüthgen, N. (2012) Specialization of  
 726 mutualistic interaction networks decreases toward tropical latitudes. *Current*  
 727 *Biology* **22**, 1925–1931.
- 728 Snow, B. & Snow, D. (1972) Feeding niches of hummingbirds in a Trinidad valley.  
 729 *Journal of Animal Ecology* **41**, 471–485.
- 730 Snow, B. & Snow, D. (1988) *Birds and berries*. Poyser, Calton, UK.
- 731 Stang, M., Klinkhamer, P., Waser, N.M., Stang, I. & van der Meijden, E. (2009)  
 732 Size-specific interaction patterns and size matching in a plant-pollinator inter-  
 733 action web. *Annals Of Botany* **103**, 1459–1469.
- 734 Strogatz, S. (2001) Exploring complex networks. *Nature* **410**, 268–276.
- 735 Stumpf, M.P.H., Wiuf, C. & May, R.M. (2005) Subnets of scale-free networks  
 736 are not scale-free: Sampling properties of networks. *Proceedings of the National*  
 737 *Academy of Sciences USA* **102**, 4221–4224.
- 738 Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the  
 739 architecture of mutualistic and trophic networks. *Science* **329**, 853–856.
- 740 Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A.,  
 741 Galetti, M., García, M.B., García, D., Gomez, J.M., Jordano, P., Medel, R.,  
 742 Navarro, L., Obeso, J.R., Oviedo, R., Ramírez, N., Rey, P.J., Traveset, A.,  
 743 Verdú, M. & Zamora, R. (2014) Beyond species loss: the extinction of ecological  
 744 interactions in a changing world. *Functional Ecology* **29**, 299–307.

- 745 Vazquez, D., Morris, W. & Jordano, P. (2005) Interaction frequency as a surrogate  
746 for the total effect of animal mutualists on plants. *Ecology Letters* **8**, 1088–1094.
- 747 Vázquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009) Evaluating multiple determi-  
748 nants of the structure of plant-animal mutualistic networks. *Ecology* **90**, 2039–  
749 2046.
- 750 Vazquez, D.P., Ramos-Jiliberto, R., Urbani, P. & Valdovinos, F.S. (2015) A con-  
751 ceptual framework for studying the strength of plant-animal mutualistic inter-  
752 actions. *Ecology Letters* **18**, 385–400.
- 753 Vizentin-Bugoni, J., Maruyama, P.K., Debastiani, V.J., Duarte, L.d.S., Dalsgaard,  
754 B. & Sazima, M. (2016) Influences of sampling effort on detected patterns and  
755 structuring processes of a neotropical plant–hummingbird network. *Journal of*  
756 *Animal Ecology* **85**, 262–272.
- 757 Vizentin-Bugoni, J., Maruyama, P.K. & Sazima, M. (2014) Processes entangling  
758 interactions in communities: forbidden links are more important than abundance  
759 in a hummingbird-plant network. *Proceedings Of The Royal Society B-Biological*  
760 *Sciences* **281**, 20132397–20132397.
- 761 Wells, K. & O’Hara, R.B. (2012) Species interactions: estimating per-individual  
762 interaction strength and covariates before simplifying data into per-species eco-  
763 logical networks. *Methods in Ecology and Evolution* **4**, 1–8.
- 764 Wirta, H.K., Hebert, P.D.N., Kaartinen, R., Prosser, S.W., Várkonyi, G. & Roslin,  
765 T. (2014) Complementary molecular information changes our perception of food  
766 web structure. *Proceedings of the National Academy of Sciences USA* **111**, 1885–  
767 1890.

## Figure captions

**Figure 1.** Sampling ecological interaction networks (e.g., plant-animal interactions) usually focus on different types of subsampling the full network, yielding submatrices  $\Delta[m, n]$  of the full interaction matrix  $\Delta$  with  $A$  and  $P$  animal and plant species. a) all the potential plants interacting with a subset of the animals (e.g., studying just the hummingbird-pollinated flower species in a community); b) all the potential animal species interacting with a subset of the plant species (e.g., studying the frugivore species feeding on figs *Ficus* in a community); and c) sampling a subset of all the potential animal species interacting with a subset of all the plant species (e.g., studying the plant-frugivore interactions of the rainforest understory).

**Figure 2.** Sampling species interactions in natural communities. Suppose an assemblage with  $A = 3$  animal species (red, species 1–3 with three, two, and 1 individuals, respectively) and  $P = 3$  plant species (green, species a-c with three individuals each) (colored balls), sampled with increasing effort in steps 1 to 6 (panels). In Step 1 we record animal species 1 and plant species 1 and 2 with a total of three interactions (black lines) represented as two distinct interactions:  $1 - a$  and  $1 - b$ . As we advance our sampling (panels 1 to 6, illustrating e.g., additional sampling days) we record new distinct interactions. Note that we actually sample and record interactions among individuals, yet we pool the data across species to get a species by species interaction matrix. Few network analyses have been carried out on individual data (Dupont *et al.*, 2014). Above and below each panel are the cumulative number of distinct species and interactions sampled, so

792 that panel 6 illustrates the final network.

793

794 **Figures**

Figure 1:

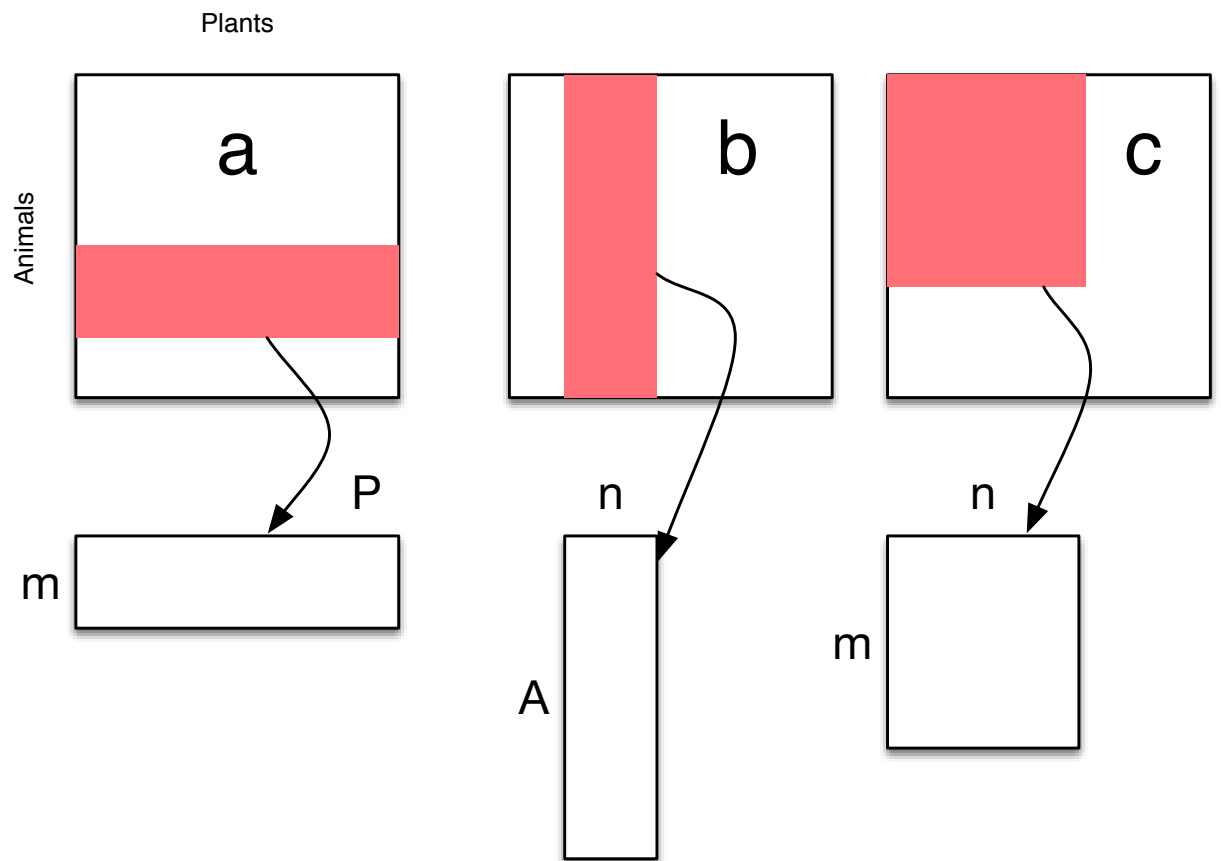
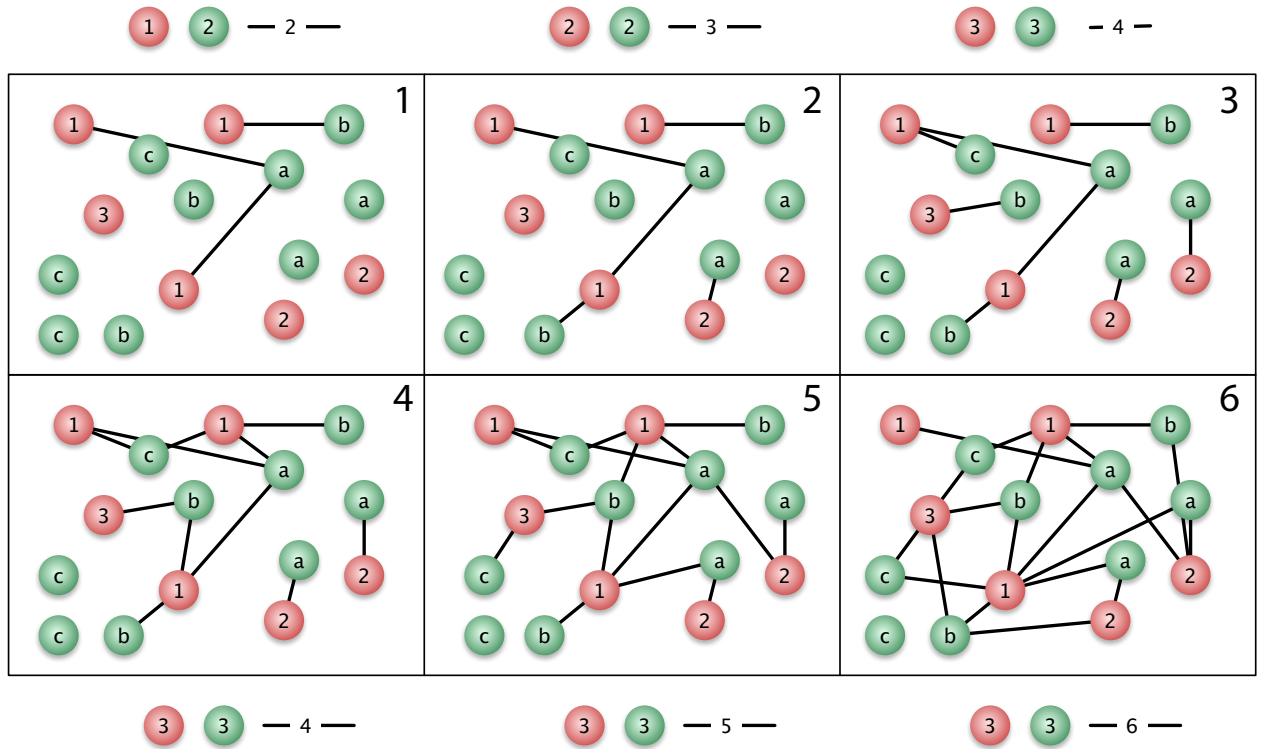


Figure 2:



Jordano - Figure 1



## Table captions

**Table 1.** A taxonomy of link types for ecological interactions (Olesen *et al.* 2011).

$A$ , number of animal species;  $P$ , number of plant species;  $I$ , number of observed links;  $C = 100I/(AP)$ , connectance;  $FL$ , number of forbidden links; and  $ML$ , number of missing links. As natural scientists, our ultimate goal is to eliminate  $ML$  from the equation  $FL = AP - I - ML$ , which probably is not feasible given logistic sampling limitations. When we, during our study, estimate  $ML$  to be negligible, we cease observing and estimate  $I$  and  $FL$ .

**Table 2.** Frequencies of different type of forbidden links in natural plant-animal interaction assemblages.  $AP$ , maximum potential links,  $I_{max}$ ;  $I$ , number of observed links;  $UL$ , number of unobserved links;  $FL$ , number of forbidden links;  $FL_P$ , phenology;  $FL_S$ , size restrictions;  $FL_A$ , accessibility;  $FL_O$ , other types of restrictions;  $ML$ , unknown causes (missing links). Relative frequencies (in parentheses) calculated over  $I_{max} = AP$  for  $I$ ,  $ML$ , and  $FL$ ; for all forbidden links types, calculated over  $FL$ . References, from left to right: Olesen *et al.* 2008; Olesen & Myrthue unpubl.; Snow & Snow 1972 and Jordano *et al.* 2006; Vizentin-Bugoni *et al.* 2014; Jordano *et al.* 2009; Olesen *et al.* 2011.

**Table 3.** A vectorized interaction matrix.

**Table 4.** Sampling statistics for three plant-animal interaction networks (Olesen *et al.* 2011). Symbols as in Table 1;  $N$ , number of records;  $Chao1$  and  $ACE$  are asymptotic estimators for the number of distinct pairwise interactions  $I$  (Hortal

819 *et al.* 2006), and their standard errors;  $C$ , sample coverage for rare interactions  
820 (Chao & Jost 2012). Scaled asymptotic estimators and their confidence intervals  
821 ( $CI$ ) were calculated by weighting *Chao1* and *ACE* with the observed frequencies  
822 of forbidden links.

823

## 824 Tables

Table 1:

Link type	Formulation	Definition
Potential links	$I_{max} = A_{obs}P_{obs}$	Size of observed network matrix, i.e. maximum number of potentially observable interactions; $A_{obs}$ and $P_{obs}$ , numbers of interacting animal and plant species, respectively. These might be below the real numbers of animal and plant species, $A_{est}$ and $P_{est}$ .
Observed links	$I_{obs}$	Total number of observed links in the network given a sufficient sampling effort. Number of ones in the adjacency matrix.
True links	$I_{est}$	Total number of links in the network given a sufficient sampling effort; expected for the augmented $A_{est}P_{est}$ matrix.
Unobserved links	$UL = I_{max} - I_{obs}$	Number of zeroes in the adjacency matrix.
True unobserved links	$UL* = I_{max} - I_{obs}$	Number of zeroes in the augmented adjacency matrix that, eventually, includes unobserved species.
Forbidden links	$FL$	Number of links, which remain unobserved because of linkage constraints, irrespectively of sufficient sampling effort.
Observed Missing links	$ML = A_{obs}P_{obs} - I_{obs} - FL$	Number of links, which may exist in nature but need more sampling effort and/or additional sampling methods to be observed.
True Missing links	$ML* = A_{est}P_{est} - I_{est} - FL$	Number of links, which may exist in nature but need more sampling effort and/or additional sampling methods to be observed. Augments $ML$ for the $A_{est}P_{est}$ matrix.

Table 2:

Link type	Pollination			Seed dispersal		
	Zackenberg	Grundvad	Arima Valley	Sta. Virginia	Hato Ratón	Nava Correhuelas
$I_{max}$	1891	646	522	423	272	825
$I$	268 (0.1417)	212 (0.3282)	185 (0.3544)	86 (0.1042)	151 (0.4719)	181 (0.2194)
$UL$	1507 (0.7969)	434 (0.6718)	337 (0.6456)	337 (0.4085)	169 (0.5281)	644 (0.7806)
$FL$	530 (0.3517)	107 (0.2465)	218 (0.6469)	260 (0.7715)	118 (0.6982)	302 (0.4689)
$FL_P$	530 (1.0000)	94 (0.2166)	0 (0.0000)	120 (0.1624)	67 (0.3964)	195 (0.3028)
$FL_S$	$\dots(\dots)$	8 (0.0184)	30 (0.0890)	140 (0.1894)	31 (0.1834)	46 (0.0714)
$FL_A$	$\dots(\dots)$	5 (0.0115)	150 (0.445) <sup>a</sup>	$\dots(\dots)$	20 (0.1183)	61 (0.0947)
$FL_O$	$\dots(\dots)$	$\dots(\dots)$	38 (0.1128) <sup>b</sup>	$\dots(\dots)$	$\dots(\dots)$	363 (0.5637)
$ML$	977 (0.6483)	327 (0.7535)	119 (0.3531)	77 (0.1042)	51 (0.3018)	342 (0.5311)

<sup>a</sup>, Lack of accessibility due to habitat uncoupling, i.e., canopy-foraging species vs. understory species.

<sup>b</sup>, Colour restrictions, and reward per flower too small relative to the size of the bird.

Dots indicate no data available for the  $FL$  type.

Table 3:

Interaction	Sample 1	Sample 2	Sample 3	...	Sample $i$
A1 - P1	12	2	0	...	6
A1 - P2	0	0	0	...	1
...	...	...	...	...	...
A5 - P3	5	0	1	...	18
A5 - P4	1	0	1	...	3
...	...	...	...	...	...
A <sub>i</sub> - P <sub>i</sub>	1	0	1	...	2

Table 4:

	Hato Ratón	Nava Correhuelas	Zackenberg
$A$	17	33	65
$P$	16	25	31
$I_{max}$	272	825	1891
$N$	3340	8378	1245
$I$	151	181	268
$C$	0.917	0.886	0.707
$Chao1$	$263.1 \pm 70.9$	$231.4 \pm 14.2$	$509.6 \pm 54.7$
$ACE$	$240.3 \pm 8.9$	$241.3 \pm 7.9$	$566.1 \pm 14.8$
% <i>unobserved</i> <sup>a</sup>	8.33	15.38	47.80

<sup>a</sup>, estimated with library Jade (R Core Development Team 2010, Chao *et al.* 2015)