Inferring biotic interactions from proxies

Ignacio Morales-Castilla[1,2], Miguel Matías[1,3,4], Dominique Gravel[5,6] &
Miguel B. Araújo[1,3,7]

[1]InBio/CIBIO, Universidade de Évora, Largo dos Colegiais, 7000 Évora, Portugal

[2]Department of Biology, McGill University, 1205 ave Docteur Penfield,
Montreal, QC, H3A 1B1, Canada

[3]Departmento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales, CSIC,
Calle José Gutiérrez Abascal, 2, 28006, Madrid, Spain

[4]Imperial College London, Silwood Park Campus, Buckhurst Road, SL5 7PY Ascot, Berkshire,
United Kingdom

[5]Département de Biologie, Université du Québec à Rimouski, 300, Allée des Ursulines,
Rimouski, QC, G5L 3A1, Canada

[6]Quebec Center for Biodiversity Science, Montreal, QC, Canada

[7]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark,
University of Copenhagen, DK-2100 Copenhagen, Denmark

Abstract: Inferring biotic interactions from functional, phylogenetic and geographical proxies remains one great challenge in ecology. We propose a conceptual framework to infer links among species within regional species pools. Firstly, the framework starts with a characterisation of the backbone of biotic interaction networks. In this step, functional groups are identified to order links and remove forbidden interactions between species. Secondly, additional links are removed by examination of the geographical context in which species co-occur. Thirdly, hypotheses are proposed to establish interaction probabilities between species. We illustrate the framework using published food webs in terrestrial and marine systems. We conclude that preliminary descriptions of the web of life can be made by careful integration of data with theory.

Why inferring interactions?

Even if serious gaps in knowledge of biodiversity remain, much progress has been made in determining how many different types of organisms exist (the Linnaean shortfall[1]), what evolutionary relationships connect different lineages to a common ancestor (the Darwinian shortfall[2]), and where different species are distributed (the Wallacean shortfall[3]). Much less is known about the types of interactions that exist among species (the Eltonian shortfall[4]) and the importance of such interactions for the maintenance and evolution of life on earth. One reason why the Eltonian shortfall is so prevalent is that detection of direct and indirect interactions requires significant empirical efforts, even in simple systems. For example, identifying direct species interactions within an ecosystem with only seven species, would require accounting for forty-two potential links and up to 13,650 links if indirect interactions are considered [5]. Given that most systems have more than seven species, documenting all biotic interactions at any site (let alone across the world) would be a considerable undertaking. Furthermore, the widespread idea that biotic interactions affect ecological processes mainly at local scales of resolution and extent [6-8] has discouraged researchers working at biogeographical scales to address them: why deal with biotic interactions if they do not contribute to understanding the bigger ecological picture? Recent evidence, however, has shown that biotic interactions can affect species ranges [9-11] and co-distribution of species [12] at regional to continental scales, thus potentially mediating biological responses to environmental changes [13]. There is a demand for predictions about the dynamics and functioning of novel ecosystems emerging from differential responses of species to global changes [14, 15]. Therefore, and since determining the nature of all interactions among species is currently beyond reach, the development of a methodological framework for inferring interactions from proxies is a timely alternative.

But can biotic interactions be inferred? It has been argued that the study of pairwise interactions between species has failed to provide general principles about the dynamics and organization of communities [16], but the recent upsurge of interest in ecological networks [17] is gradually changing this perspective. We propose that, if appropriately analysed, existing data on functional traits, geographical distributions, and phylogenies, provides a starting point for making predictions about the likelihood of pairwise biotic interactions among species and the general properties of the interaction networks. The proposed framework is complementary to empirical approaches for network construction (whereby links are established only after observations are made [18]), and can be useful when empirical data on interactions are not accessible. Our approach leads to generating the backbone of an interactions network that can be used to identify plausible links among species in a given regional species pool.

What interactions to infer?

There are many different ways to describe a biotic interaction. Interactions may vary in their type (e.g., antagonistic or facilitative), their strength (e.g., weak or strong interactions), or their symmetry (e.g., symmetrical or asymmetrical). An important step for inferring biotic interactions is to determine what information is to be inferred. We propose building interaction networks bottom-up, i.e., predicting the links among species and then explore the collective properties of the networks (e.g., connectance, degree distribution, or distribution of motifs) emerging from these links [see also 13]. In the proposed framework, four key metrics are the focus of inference:

(i) *The probability of occurrence of a link* - The first step is to infer whether an interaction between any given pair of species is possible. Many species co-occur without interacting with each other [4] and while detection of links is the focus of much research in community ecology, predicting if a link is absent is of equal importance. Forbidden interactions, such as small fish eating big fish or grassland herbivores eating leaves in trees, can be easily identified thus helping to remove links from a full set of candidate links. Such process of pruning forbidden links from potential networks of interactions helps to limit the scope of inference to possible interactions alone (see Fig. 1). After removing the forbidden links from the tree, the next step is to calculate the probability with which possible links occur using a combination of data that can include traits, phylogenies and geographical features [19].

(ii) *The type of interaction* - Once a link between two species is deemed probable, the next step is to determine the expected type of interaction involved (e.g., antagonistic, facilitative, direct, and indirect). Such inference is not always straightforward. Out of the myriad of interactions in nature, not all are equally prevalent and not all are equally detectable. Variation in the prevalence and detectability of interactions partly explains why different interactions have received different attention. For example, in the past 20 years, competition attracted ca. 50% of all citations in the biotic interactions literature with commensalism and amensalism being almost residual (see Box 1). Species can also interact in more than one way making the classification of interaction particularly complex. For example, barnacles facilitate establishment of mussels in the rocky intertidal, while simultaneously competing with them for space. Whether (+/- vs. -/-) interactions are prevalent will depend on the net effect of the interactions, which will in turn determine the resulting pattern of attraction or repulsion [20]. Focusing on easily inferable direct interactions as a starting point (e.g., predator-prey interactions) can help because it will resolve the major interacting groups of the network.

(iii) *The strength of the interaction* - The direct effect that species have on one another's demography is termed interaction strength. Determining the strength of interactions between species is essential to understanding network dynamics, stability and robustness [21]. Approaches for measuring strength of interactions will vary with the specific types of interactions involved. Whereas interaction strength in a plant-pollinator network can be measured based on the frequency of visits between species pairs [22], in a food web it can be measured by the amount of energy ingested by the predator [23]. More generally, interaction strength can be measured as the total effect of a population on another (*per population*) or the effect of each individual of a population on another (*per capita*). Most attempts to estimate strength of interactions in the field will mix both. Here, we focus on strength of interactions at the population level.

(iv) *Asymmetry of interactions* – the degree to which the strength of the interaction between two species varies from one another is termed asymmetry. Interactions are generally asymmetrical [13, 22, 24], so that if one species A depends strongly on species B, species B depends weakly on species A. For instance, plants and pollinators vary in their degrees of mutual dependence with pollinators being typically more specialized than plants [22]. Another example is provided by predator-prey interactions, whereby only a fraction of the prey biomass consumed by the predator is converted into predator biomass. Failure to acknowledge these asymmetries will necessarily lead to overestimating the importance of particular links.

How to infer interactions?

The proxies

When direct information about biotic interactions is unavailable, we must resort to indirect information or proxies to get insight about them. Three classes of proxies can help inferring interactions between species: traits, phylogenies, and geographical data (for review of examples see Table 1). Traits are usually defined as morphological, physiological, phenological, or behavioural characteristics of species that directly impact their fitness [25]. However, they are also expected to influence how species interact with one another. When information on traits is unavailable, or is incomplete, one option is to use phylogenies measuring evolutionary relationships among species as a surrogate for trait similarity. This approach is useful as long as there is phylogenetic signal in the traits of interest [26], and strong phylogenetic signals are more likely to exist when lineages are examined across broader geographical scales of extent or large the environmental gradients. Finally, in addition to information on traits, the geographical co-distribution of species, complemented with analysis of species-environment/species-habitat relationships can also be considered.

The three classes of proxies are not expected to provide equally valuable information (see Table 1). Traits ultimately determine if a given interaction is probable, or even possible, as they mediate species relationships with the environment. For example, different species might occupy a particular microhabitat differently and this can determine whether they are likely to compete for a particular resource. When trait information is lacking, phylogenies and geography can help filling the gaps; for example, by determining which species are likely to co-occur in the same parts of the world. We illustrate below how these proxies (e.g., body size, abiotic requirements, phylogenetic distance, spatial co-occurrence) can be utilized to infer species interactions.

Building the backbone of interaction networks

We propose a hierarchical approach to infer biotic interactions from proxies that is analogous to reconstructing large phylogenies based on collating smaller phylogenies on the top of a structuring backbone (i.e., ‘super trees’ or ‘meta trees’). Broad groups of interacting species would, ultimately, form a ‘super network’ or ‘meta web’ of interactions. The analogy with a super tree or meta tree is not new as it is implicit in the compilation of several networks of interactions [27], but it formalizes what is typically done for network construction.

The backbone of an interaction’s network starts with the identification of species more likely to share similar interactions. The concept is similar to modules [28], but we avoid this terminology as they are usually determined *a posteriori* and can also refer to simple assemblages of species such as linear food chains or apparent competition [29]. Instead, we define interacting groups based on *a priori* expectations of interactions. The concept is also analogous to that of guilds [30]. Guilds, however, are restricted to species sharing similar resources thus not encompassing non-consumptive interactions such as competition or niche construction [27]. A flexible definition of interacting groups based on traits, phylogenies and geographical distributions would enable combination of heterogeneous information. Such flexibility can be critical when inferring interactions with information that is not required across all species (e.g., foliage chemical defence compounds are not relevant to infer predator-prey interactions).

As an example of implementation of these ideas, we infer the backbone of food webs for two ecosystems: the Serengeti [31]; and the Cuban coral reef [32] (Fig. 1). Interacting groups of species are defined a priori in order to simplify the removal of forbidden links. The groups were defined based on the trophic hierarchy of the different species within the ecosystem (e.g., primary producers, grazers, small and large carnivores). This process of trophic classification of species led to identification of forbidden links and removal of ~30% of all potential direct links in the coral reef, and ~22% in the Serengeti (e.g., herbivores eating predators; Fig. 1). Refinement of the species groupings was achieved by considering the characteristics of the consumer species (e.g., distinguishing small vs. large carnivores in the Serengeti example, or separating invertebrate feeders, omnivorous and carnivore fish in the coral reef example; Fig. 1). Geography was then used to subdivide the producer groups, defining environments where only certain consumer species feed (e.g., grassland, woodland and shrubland are differenced in the Serengeti, and coral reefs and seagrasses are distinguished in the Cuban network). This step led to reducing up to 50% of the remaining links (Fig. 1). Notice that the proportion of links removed by the implementation of the framework was similar despite the distinct nature of the ecosystems and the reversed structure of the trophic networks (Fig.1).

There are few attempts in the literature that use trait-matching constraints to infer links. For example, Gravel et al. [25] used predator-prey body-size relationships to calibrate the niche model [26] and infer potential predator-prey interactions among pelagic Mediterranean fishes. Eklof et al. [33] showed that usually less than five dimensions were required to represent the structure of 200 networks of different types. When direct measurements of traits are unavailable, latent traits or phylogenetic information can also be used [27]. The ability to establish empirical relationships between traits (measured trait, latent traits, or phylogenetic relationships) and interactions will be key to predict whether different species coming into contact will interact or not. This would be the case when spread of exotic species leads them to colonizing new environments or when species shift their ranges in response to climate change.

Even though we illustrated our framework with examples of food webs (see Fig. 1), the framework is general and can be applied to other types of interactions. For example, after direct trophic interactions are mapped into the backbone of an interactions’ network, potential indirect interactions, such as exploitative and apparent competition [34, 35], or trophic cascades [36], can be inferred [27].

We note that theoretical [37] and experimental [38] studies have shown that pairwise interactions and network structure can be constrained by the environment, which leads to significant spatial variability in network structure [39]. Additionally, phylogenetic signal in networks can increase with increasing environmental stress as shown by parasitoids narrowing down their host-genotype niche breadth when temperature increases [40]. Further studies will need to address the relationship between the environment and network structure, as it is essential to predict feedbacks among species co-distributions, biotic interactions, and environmental change.

Assigning interaction probabilities

Identifying forbidden links in a network is relatively straightforward compared to the more subtle exercise of assigning probabilities of interactions between species. The null expectation is that, all other things being equal, the probability of interaction between two species is given by their prevalence (i.e., proportion of an area occupied by the species in a given geographical region) [20, 41]. Departures from this null expectation should arise whenever interaction strength between species is different from zero. In such cases, traits or the interaction between traits and the environment should modulate interaction probabilities (see Fig. 2). Zero probabilities of interaction should only be considered for cases of incompatible traits (e.g., herbivores do not feed on other animals).

Treating interactions as probabilistic is relatively novel. An approach that could represent our uncertainty in our model describing what governs the interaction (e.g., which traits to include), or alternatively, given that an interaction is feasible, it does not systematically occurs at a location where two species are found co-occurring. Some interactions might not realize themselves even though they are possible, either because some species are too rare or the environmental conditions (abiotic and biotic) are not suitable. It is as critically important to document them as the non-zero probabilities, and new empirical investigations of network structure should also aim at providing a confidence interval on them (a quantification of the probability of false negative).

Given that probabilities of interactions depend on the strength of the interactions between species, and their symmetry, calculating these metrics is important. The strength of interactions between pairs of species can be inferred, indirectly, from theoretical models or, directly, by field or laboratory experiments. Examples of the latter have involved examining the role of species in mutualistic networks [42] or changes in the structural complexity of the habitat [43, 44] in parasitoid-host networks. Standard predator or competitor removal experiments [45] have also provided estimates of *per population* interaction strength; *per capita* interactions would then be estimated providing that predator density was known. Mass-balance models can also be used to estimate interaction strength. For example, de Ruiter et al. [46] parameterized Lotka-Volterra equations based on prior identification of trophic links in soil food webs, measurements of population size and estimates of natural mortality rates. At equilibrium, the total energy consumed should balance mortality and predation and, therefore, it should be possible to derive interaction strength by deduction.

There are also several examples of indirect approaches for inference of interaction strength. For example, predator-prey body-size ratio is often used to estimate per capita effect of predators on their prey [47], or to derive this effect from metabolic scaling relationships [48]. Besides population size, interaction strength has also been associated with the frequency at which actual interactions occur. Neutral models of interactions were shown to predict relatively well the interaction strength between pairs of species within trophic guilds based on their local population sizes [49, 50]. In neutral models, the probability of species interacting together is proportional to the product of their relative abundance. Neutral forbidden links arise as two rare species are very unlikely to interact [41]. The network properties are consequently the direct result of frequency distributions of abundance of the different guilds, which in turn respond to changing environmental conditions [43, 44]. Increasing understanding of environmental-driven non-random changes in network structure will allow adjusting the probabilities of interaction, and generating more accurate predictions.

The asymmetry of interactions is the most understudied of the factors affecting interaction probabilities. Most modelling studies use a constant across all species to represent consumption inefficiency [51], but theoretical or empirical support for such an assumption is limited. Experimental studies are also usually conducted in a single direction; measuring the effect of predator removal on the prey biomass in the field [52] is often accomplished, but more challenging is to evaluate the effect of prey removal on the predator [but see 53]. The asymmetry of interactions has also been related to co-evolutionary dynamics [54] and it is likely that further indirect inferences of asymmetry will rely on phylogenetic relationships.

Based on the above, we list six simple hypotheses to guide inference about probabilities of interaction between pairs of species (see Fig. 2). To illustrate the implementation of these hypotheses, we apply one of them (i.e., increasing probability of predation with increasing predator-prey body size ratio) to infer probability of interaction for the subsets of consumer species within each of our two examples (Serengeti mammals and Cuban fish) (see Box 2).

Testing inferences about interactions

The usefulness of a theoretical model is partly dependent on it being successfully tested. However, inferences of biotic interactions by models are not easy because reliable absence data about interactions are generally unavailable. Similar problems exist in the species distributions modelling literature [4, 55], with the consequence that inferences of interactions must necessarily be interpreted as potential rather than realised. Indeed, observed interactions will typically constitute a small subset of all realised interactions and these are, themselves, a subset of all potential interactions within a given species pool. The consequence is that false positives (interactions predicted but not recorded) might not be an error at all; they might characterise existing but undetected interactions, or potential interactions that have yet not been realised. Whilst the meaning of false positives is often difficult to ascertain, false negatives (observed interactions not predicted by the model), provide clear indication that the inferences by the models are inaccurate. Our implementations of the proposed framework show how reductions of false positives are achieved by estimating probabilities of interaction based on a simple hypothesis (i.e., predator-prey body size ratios), while false negatives remain stable and at low levels (see Box2, Fig. II).

In most cases, inferred interactions will never be observed for several reasons, including low detectability, low abundances, dispersal barriers, and lack of overlap between ecological niches of the two species [19]. Detection of interactions will increase if sufficiently coarse spatial and temporal scales of observation are allowed. Likewise, non-interacting species might start interacting if novel circumstances prompt them to do so (e.g., stochastic long distance dispersal enabling contact of otherwise disjoint species, reshuffling of species geographic distributions due to shifts in the distributions of ecological niches).

In other cases, species might co-occur but interact only in certain parts of their ranges. It is the case of Australian tiger sharks, whose diet shifts geographically, selecting different (but equally available) prey species in different locations [56]. The Cuban Coral reef network example shows a substantial number of predicted links below the diagonal for which there is no empirical evidence (Box 2, Fig. Ii-j). The question arises as to whether these false positives could be realized in other locations. This question is relevant to the study of novel ecological communities expected under climate change [57].

The sheer numbers of potential interactions that exist in nature pose additional difficulties to testing inferences of interactions with experimental or observational approaches. Nevertheless, not all links affect properties of interactions networks equally and there is scope for simplification if only the most relevant species in the network, e.g., those with more or stronger connections, are examined [22].

Concluding remarks

We proposed a framework for inferring biotic interactions based on stepwise removal of forbidden links and calculation of probabilities of interaction for the remaining links. With such a process one is able to establish the backbone of an interactions network occurring in a given species pool. The pruning of the network is made using rules derived from functional traits, phylogenies and geographical proxies. For the sake of illustration we implemented the framework with food webs in two ecosystems (i.e., the Serengeti and a Cuban coral reef). We demonstrated how basic understanding of traits enables accurate predictions of the overall structure of the food webs (Fig. 1). While further testing of these ideas is warranted, several conceptual challenges remain. Among them, it will be critical to look more closely at how emerging properties such as mean trophic level, modularity, nestedness, change along the pruning sequence. This issue is reminiscent of the intense debates in the food web literature about the impact of sampling intensity on network properties [18]. Full understanding of the web of life seems distant, but sequentially building super networks or meta webs of biotic interactions networks will help unravelling key interactions and their potential effects on the distribution of life on earth.

Acknowledgements

We thank Natalia Melo for support with the design of the artwork. Research by IM-C, MGM, and MBA was supported through the Integrated Program of IC&DT (1/SAESCTN/ALENT-07-0224-FEDER-001755). IM-C, is currently funded by the Fonds de Recherches du Québec - Nature et Technologies (FQRNT) programme. MGM also acknowledges support by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (FORECOMM). DG received financial support from NSERC and Canada Research Chair program.

References

1 Brown, J.H. and Lomolino, M.V. (1998) *Biogeography*. Sinauer Press

2 Diniz-Filho, J.A.F.*, et al.* (2013) Darwinian shortfalls in biodiversity conservation. *Trends Ecol Evol* 28, 689-695

3 Whittaker, R.J.*, et al.* (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions* 11, 3-23

4 Peterson, A.T.*, et al.* (2011) *Ecological Niches and Geographical Distributions*. Princeton University Press

5 Dodds, W.K. and Nelson, J.A. (2006) Redefining the community: a species-based approach. *Oikos* 112, 464-472

6 Pearson, R.G. and Dawson, T.E. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12, 361-371

7 Benton, M.J. (2009) The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science (New York, N.Y.)* 323, 728-732

8 Loreau, M. (2010) *From populations to ecosystems: Theoretical foundations for a new ecological synthesis (MPB-46)*. Princeton University Press

9 Araújo, M.B. and Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16, 743-753

10 Heikkinen, R.K.*, et al.* (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecol Biogeogr* 16, 754-763

11 Bateman, B.L.*, et al.* (2012) Biotic interactions influence the projected distribution of a specialist mammal under climate change. *Diversity and Distributions* 18, 861-872

12 Gotelli, N.J.*, et al.* (2010) Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences* 107, 5030-5035

13 Araújo, M.B.*, et al.* (2011) Using species co-occurrence networks to assess the impacts of climate change. *Ecography* 34, 897-908

14 Montoya, J.M. and Raffaelli, D. (2010) Climate change, biotic interactions and ecosystem services. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 2013-2018

15 Blois, J.L.*, et al.* (2013) Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences* 110, 9374-9379

16 McGill, B.J.*, et al.* (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol* 21, 178-185

17 Bascompte, J. (2009) Disentangling the web of life. *Science* 325, 416

18 Martinez, N.D. (1991) Artifacts or Atributes? Effects of resolution on the Little Rock Lake food web. *Ecol Monogr* 61, 367-392

19 Poisot, T.*, et al.* (2014) Beyond species: why ecological interactions vary through space and time. *bioRxiv*

20 Araújo, M.B. and Rozenfeld, A. (2014) The geographic scaling of biotic interactions. *Ecography* 37, 406-415

21 Laska, M.S. and Wootton, J.T. (1998) Theoretical concepts and empirical approaches to measuring interaction strength. *Ecology* 79, 461-476

22 Bascompte, J.*, et al.* (2006) Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science* 312, 431-433

23 Wootton, J.T. (1997) Estimates and tests of per capita interaction strength: diet, abundance, and impact of intertidally foraging birds. *Ecol Monogr* 67, 45-64

24 Ruiter, P.C.d.*, et al.* (1993) Calculation of nitrogen mineralization in soil food webs. *Plant and Soil* 157, 263-273

25 Violle, C.*, et al.* (2007) Let the concept of trait be functional! *Oikos* 116, 882-892

26 Mouquet, N.*, et al.* (2012) Ecophylogenetics: advances and perspectives. *Biological Reviews* 87, 769-785

27 Kéfi, S.*, et al.* (2011) More than a meal… integrating non-feeding interactions into food webs. *Ecology Letters* 15, 291-300

28 Newman, M.E.J. (2010) *Networks: an introduction*. Oxford University Press

29 Gilman, S.E.*, et al.* (2010) A framework for community interactions under climate change. *Trends in Ecology & Evolution* 25, 325-331

30 Simberloff, D. and Dayan, T. (1991) The Guild Concept and the Structure of Ecological Communities. *Annual Review of Ecology and Systematics* 22, 115-143

31 Baskerville, E.B.*, et al.* (2011) Spatial guilds in the Serengeti food web revealed by a Bayesian group model. *PLoS Comp Biol* 7, e1002321

32 Roopnarine, P.D. and Hertog, R. (2012) Detailed Food Web Networks of Three Greater Antillean Coral Reef Systems: The Cayman Islands, Cuba, and Jamaica. *Dataset Papers in Ecology* 2013

33 Eklöf, A.*, et al.* (2013) The dimensionality of ecological networks. *Ecol Lett* 16, 577-583

34 Bonsall, M.B. and Hassel, M.P. (1997) Apparent competition structures ecological assemblages. *Nature* 388, 371-373

35 Holt, R.D. and Lawton, J.H. (1993) Apparent competition and enemy-free space of insect host-parasitid communities. *American Naturalist*???, 623

36 Borer, E.*, et al.* (2005) What determines the strength of a trophic cascade? *Ecology* 86, 528-537

37 Vasseur, D.A. and McCann, K.S. (2005) A mechanistic approach for modeling temperature-dependent consumer resource dynamics. *The American Naturalist* 166, 184-198

38 Woodward, G.*, et al.* (2010) Ecological networks in a changing climate. In *Advances in Ecological Research: Ecological Networks, Vol 42* (Woodward, G., ed), pp. 71-138, Elsevier Academic Press Inc

39 Poisot, T.*, et al.* (2012) The dissimilarity of species interaction networks. *Ecology Letters* 15, 1353-1361

40 Lavandero, B. and Tylianakis, J.M. (2013) Genotype matching in a parasitoid–host genotypic food web: an approach for measuring effects of environmental change. *Mol Ecol* 22, 229-238

41 Canard, E.*, et al.* (2012) Emergence of Structural Patterns in Neutral Trophic Networks. *PLoS ONE* 7, e38295

42 Aizen, M.A.*, et al.* (2012) Specialization and rarity predict nonrandom loss of interactions from mutualist networks. *Science* 335, 1486-1489

43 Tylianakis, J.M.*, et al.* (2007) Habitat modification alters the structure of tropical host–parasitoid food webs. *Nature* 445, 202-205

44 Laliberté, E. and Tylianakis, J.M. (2010) Deforestation homogenizes tropical parasitoid-host networks. *Ecology* 91, 1740-1747

45 Berlow, E.L.*, et al.* (2004) Interaction strengths in food webs: issues and opportunities. *Journal of Animal Ecology* 73, 585-598

46 Ruiter, P.C.d.*, et al.* (1995) Energetics, Patterns of Interaction Strengths, and Stability in Real Ecosystems. *Science* 269, 1257-1260

47 Emmerson, M.C. and Raffaelli, D. (2004) Predator–prey body size, interaction strength and the stability of a real food web. *Journal of Animal Ecology* 73, 399-409

48 Yodzis, P. and Innes, S. (1992) Body Size and Consumer-Resource Dynamics. *The American Naturalist* 139, 1151-1175

49 Vázquez, D.P.*, et al.* (2009) Evaluating multiple determinants of the structure of plant-animal mutualistic networks. *Ecology* 90, 2039-2046

50 Canard, E.*, et al.* (2014) Empirical Evaluation of Neutral Interactions in Host-Parasite Networks. *The American Naturalist* 183, 468-479

51 Brose, U.*, et al.* (2006) Allometric scaling enhances stability in complex food webs. *Ecology Letters* 9, 1228-1236

52 Wootton, J.T. (1992) Indirect effects, prey suceptibility, and habitat selection: impacts of birds on limpets and algae. *Ecology* 73, 91-98

53 Fordham, D.A.*, et al.* (2013) Adapted conservation measures are required to save the Iberian lynx in a changing climate. *Nature Clim. Change* 3, 899–903

54 Rezende, E.L.*, et al.* (2007) Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* 448, 925-928

55 Araújo, M.B. and Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology* 93, 1527-1539

56 Simpfendorfer, C.A., *et al.* (2001) Size, sex and geographic variation in the diet of the tiger shark, Galeocerdo cuvier, from Western Australian waters. *Environ Biol Fishes* 61, 37-46

57 Garcia, R.A., *et al.* (2014) Multiple Dimensions of Climate Change and Their Implications for Biodiversity. *Science* 344, 1247579

58 Boulangeat, I., *et al.* (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecol Lett* 15, 584-593

59 Gross, N., *et al.* (2009) Linking individual response to biotic interactions with community structure: a trait-based framework. *Funct Ecol* 23, 1167-1178

60 Pruitt, J.N. and Ferrari, M.C.O. (2011) Intraspecific trait variants determine the nature of interspecific interactions in a habitat-forming species. *Ecology* 92, 1902-1908

61 Schemske, D. (2002) Tropical diversity: patterns and processes. In *Ecological and Evolutionary Perspectives on the origin of Tropical Diversity* (Chazdon, R. and Whitmore, T., eds), pp. 163-173, University of Chicago Press

62 Rezende, E.L., *et al.* (2007) Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. *Oikos* 116, 1919-1929

63 Ovaskainen, O., *et al.* (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology* 91, 2514-2521

64 Ings, T.C., *et al.* (2009) Ecological networks--beyond food webs. *The Journal of animal ecology* 78, 253-269

65 Aliabadian, M., *et al.* (2007) Low predictive power of mid-domain effect to explain geographic species richness patterns in Palearctic songbirds. *Contributions to Zoology*, 197-204

66 Stouffer, D.B., *et al.* (2012) Evolutionary conservation of species' roles in food webs. *Science (New York, N.Y.)* 335, 1489-1492

67 Berlow, E.L., *et al.* (2009) Simple prediction of interaction strengths in complex food webs. *Proceedings of the National Academy of Sciences of the United States of America* 106, 187-191

68 Rudolf, V.H.W. (2012) Seasonal shifts in predator body size diversity and trophic interactions in size-structured predator-prey systems. *The Journal of animal ecology* 81, 524-532

69 Broitman, B.R., *et al.* (2009) Predator-prey interactions under climate change: the importance of habitat vs body temperature. *Oikos* 118, 219-224

70 Kowalewski, M. (2002) The fossil record of predation: an overview of analytical methods. *Paleontological Society Papers* 8, 3-42

71 Gilbert, G.S., *et al.* (2012) Evolutionary tools for phytosanitary risk analysis: phylogenetic signal as a predictor of host range of plant pests and pathogens. *Evolutionary applications* 5, 869-878

72 Futuyma, D.J., ed (1998) *Evolutionary Biology*. Sinauer Associates

73 Keeling, M.J., *et al.* (2001) Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science (New York, N.Y.)* 294, 813-817

74 Kunstler, G., *et al.* (2012) Competitive interactions between forest trees are driven by species' trait hierarchy, not phylogenetic or functional similarity: implications for forest community assembly. *Ecol Lett* 15, 831-840

75 Valiente-Banuet, A. and Verdú, M. (2008) Temporal shifts from facilitation to competition occur between closely related taxa. *J Ecol* 96, 489-494

76 Cavender-Bares, J., *et al.* (2009) The merging of community ecology and phylogenetic biology. *Ecol Lett* 12, 693-715

77 Graham, C.H., *et al.* (2009) Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl, 19673-19678

78 Latimer, a.M., *et al.* (2009) Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. *Ecol Lett* 12, 144-154
79 Sebastián-González, E., *et al.* (2010) Testing the heterospecific attraction hypothesis with time-series data on species co-occurrence. *Proceedings. Biological sciences / The Royal Society* 277, 2983-2990
80 Sáyago, R., *et al.* (2013) Evaluating factors that predict the structure of a commensalistic epiphyte-phorophyte network. *Proceedings of the Royal Society B* 280, 20122821-20122821

Box 1. Are all biotic interactions equally prevalent in nature? Can they be predicted from co-occurrence?

Four broad types of biotic interactions can be distinguished based on whether the net effect of the interaction on each interacting species is detrimental for both of them (i.e., competition [-,-]), beneficial for both of them (i.e., mutualism [+,+]), positive for one species and negative for the other (e.g., predation or parasitism [+,-]), or positive for one species and indifferent for the other (i.e., commensalism [+,0]). Less common is when one interacting species is indifferent while the other is negatively affected (i.e. amensalism [0,-]. Quantification of the prevalence of the different types of interactions in nature is still lacking, but have they received similar treatment in the literature? A review of the literature of the past two decades reveals that antagonistic interactions (e.g., competition and predation) have been the focus of more than three quarters of all published studies on biotic interactions (Fig. I). These studies often refer to biotic interactions generically, even if only one interaction type—typically competition—is addressed [58].

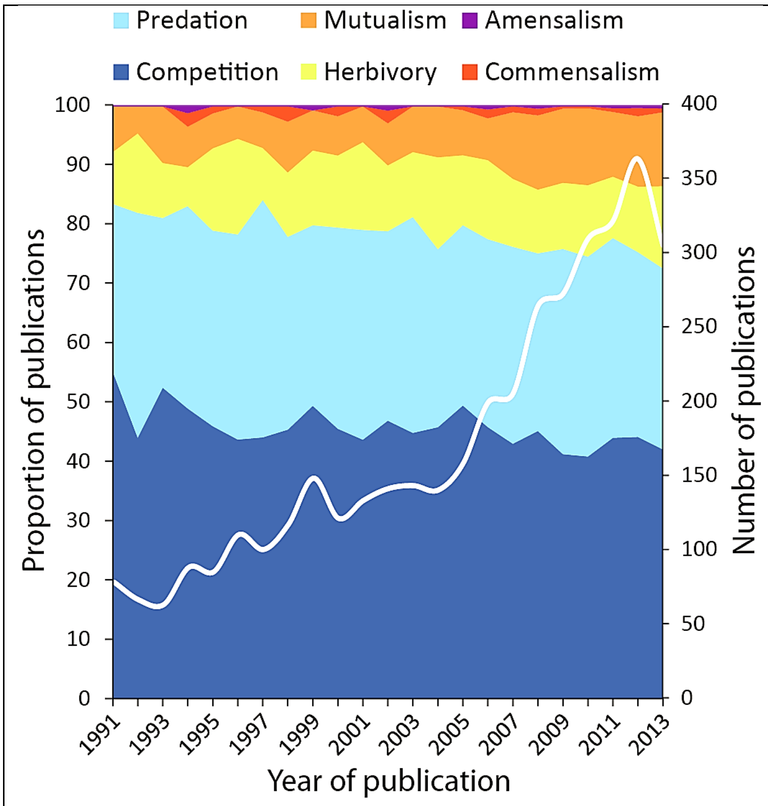

Fig. I. Biotic interaction types in the literature since 1991. Search performed with the Web of Science including the terms ‘ecology’ AND ‘interaction’ AND ‘interaction type’.

Why are antagonistic interactions so popular in the ecological literature? (1) Is there evidence that certain types of biotic interactions are more important than others in controlling for the dynamics of species, communities and ecosystems? (2) Are antagonistic interactions easier to detect in the field than facilitative interactions? (3) Are antagonistic interactions more prevalent? Answering these questions requires a comprehensive comparative analysis across all types of biotic interactions. Such analysis is not feasible given the lack of systematic descriptions of biotic interactions across taxa and regions. But simulations may offer an alternative. For example, a recent study provided the first comprehensive simulation of the expected co-occurrence between two species arising from all possible combinations of direct biotic interaction types [20]. The study shows that similar co-occurrences can be achieved by different interactions, leading to the conclusion that co-occurrences alone are not sufficient to provide insight into the biotic interactions generating them [e.g., 13] (Fig. II).

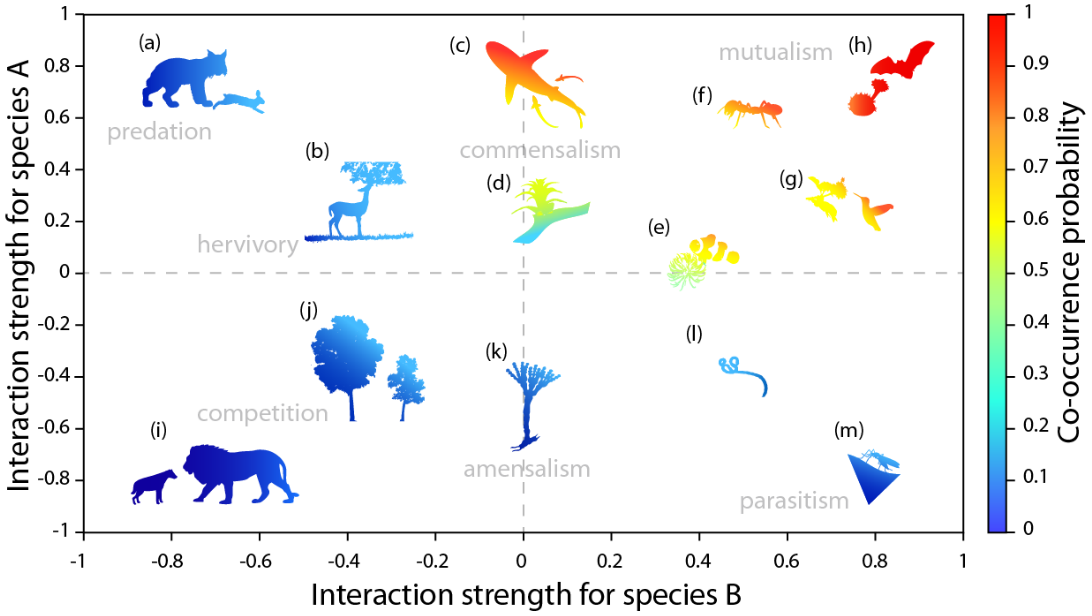

Fig. II. Co-occurrence probability between two different species across biotic interaction space. Biotic interaction space is a representation of all possible types of interactions across two axes, each indicating the direction and the strength of the interaction for each species. Examples are for (a) predation of *Oryctolagus cuniculus* by *Lynx pardinus*, (b) hervibory by *Odocoileus virginianus*, commensalism by (c) *Remora brachyptera* and *Carcharhinus melanopterus*, (d) epiphytic bromeliad (fam. Bromeliaceae) and (e) *Amphiprion percula* and *Entacmea quadricolor*; examples of mutualism for (f) shelter-defense interaction between *Pseudomyrmex ferruginea* and *Cecropia peltata*, of (g) pollination of *Helicornia caribaea* by *Eulampis jugularis* and of (h) pollination of *Stenocereus thurberi* by *Leptonycteris curasoae*; competition between (i) *Panthera leo* and *Crocuta crocuta* and between (j) *Swietenia mahagoni* individuals; amensalism produced by (k) *Penicillium expansum*, and parasitism of (l) virus of genus *Ebolavirus* and (m) *Anopheles gambiae* mosquito, which is itself host for *Plasmodium falciparum*.

Box 2. Assigning probabilities and testing inferred interactions

Estimating probabilities of interactions between pairs of species is challenging, but using proxies coupled with simple sets of rules can help (Fig. 2). As an example, we use simple allometric rules to infer probabilities of interactions on a subset of mammal species from the Serengeti food web and fish species from the Cuban coral reefs food web. Following the procedure described in Figure 1, we removed forbidden links from the full matrix of potential interactions (Fig. Ia-b,f-g). The next goal of our framework is to assign probabilities to the interaction matrix. To compute probabilities we followed a method that modifies previous similar approaches [59]. It fits a model describing the relationship between predator-prey body sizes and uses its parameters (describing the optimum and the range of predator's niches) to compute the probability of interaction given the size of the predator and prey species, and assuming a gaussian function. We parameterized the model with the observed log-sizes of predator and prey species based on the realized interactions. Note that this approach allows estimating a probability function resembling the niche of each predator species. We show here an average of the probability density function across different values of predator-prey body size ratio for simplicity (Fig. Ic,h). We finally assigned a probability to each remaining link in order to obtain the inferred matrix of interactions (Fig. Id,i). Although we restrict this example to allometric relationships, additional hypotheses (i.e. phylogenetic distance, niche distance, see Fig. 2) could be incorporated as subsequent steps.

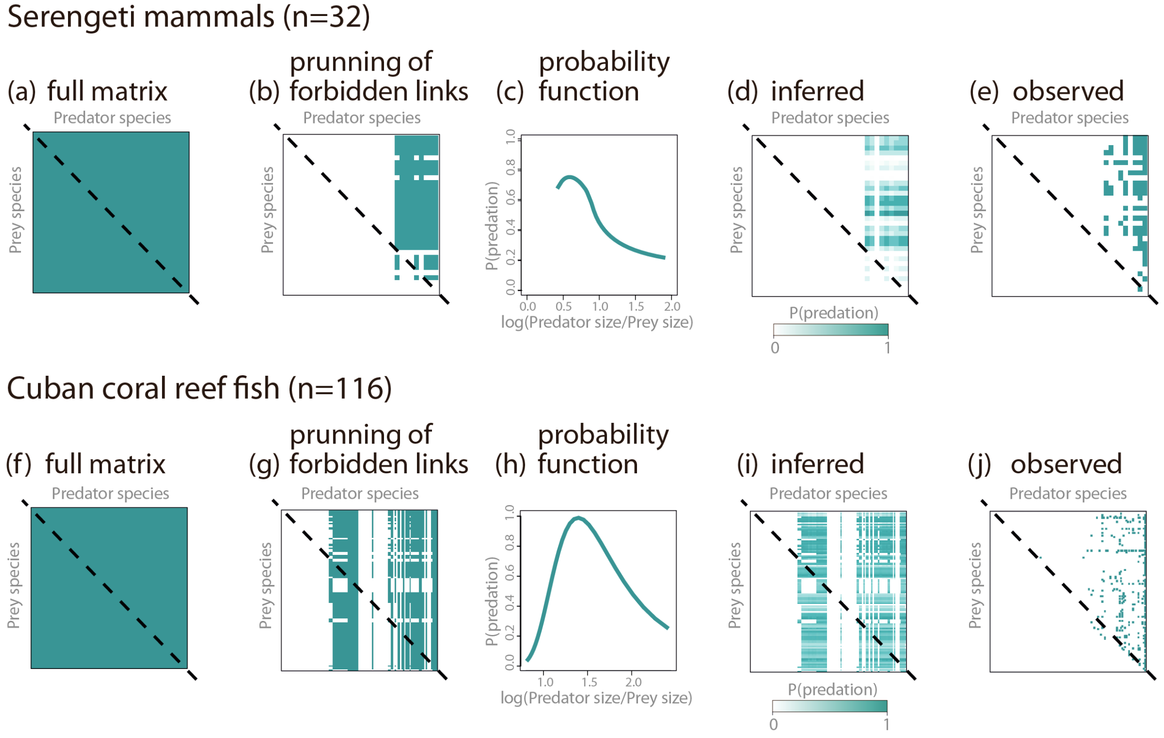

Fig. I. Predicting probabilities interactions for a subset of 32 mammal species of the Serengeti (a-e) and for 116 fish species present in the Cuban coral reefs dataset (f-j). Forbidden links are initially removed by grouping species based on their trophic level, their traits (size or diet) and their geography (b,g) (as for Figure 1). Probability density functions for predator-prey body size ratios (c,h) are then applied to estimate the probability of interaction of each link in the inferred matrices (d,i). Finally, inferred matrices of interaction are compared against the observed interactions (e,j).

Once interactions are inferred, a last stage involves testing the accuracy of the inferences against the observed interactions (Fig. Ie,j). In this example, we show how applying simple rules to prune forbidden links and to infer probabilities of interaction contributed to decrease false positive error rate, whilst leaving false negatives do not change much (Fig. II). It is to note, however, that whereas the fourth step (i.e. assignation of probabilities) does not enhance the inferences for the Serengeti mammals (Fig. IIa), it is able to improve the accuracy of the inferences in the case of the Cuban coral reefs (see Fig. IIb). In the former example, predator-prey body size relationships are not informative and thus, different hypotheses should guide the assignation of probabilities (see Fig. 2).

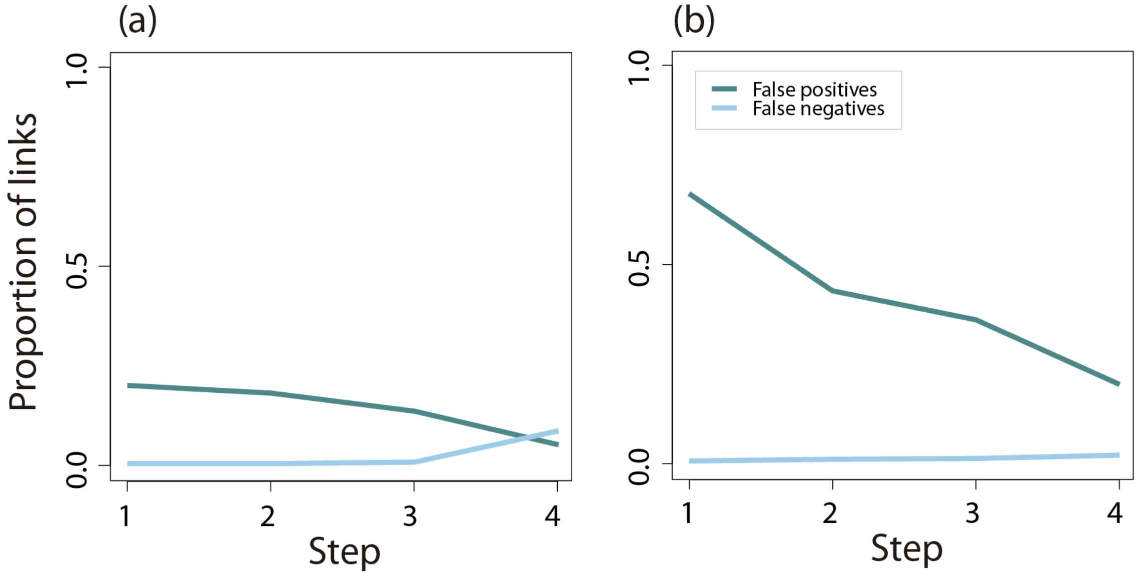

Fig. II. False positive and false negative error rate achieved in different steps of the implementation of our framework.

Several network properties (e.g. connectance, modularity, etc.) can also be computed for each backbone grouping and for the inferred interaction matrices. For instance, connectance decreases from the first backbone obtained grouping species by their trophic level to the inferred matrix of interactions by 92.3% in the Serengeti, and by 87.2% in the Cuban coral reefs. In both examples the properties of the inferred networks approximate better the properties of the observed ones.

Table 1 – Examples of proxies (traits, phylogenies and geography) used in inferences of biotic interactions.

Interaction type	**Traits**	**Phylogeny**	**Geography**
Mutualism (+/+)	Functional composition of subalpine grasslands is related to facilitation when resources are scarce [59] Behavioural trait composition can turn the interactions of spider *Anelosimus studiosus* into mutualism or commensalism [60] Flower size can determine the number of pollinator species [61]	Phylogenetic history plays a relevant role in structuring plant-animal mutualistic networks [62] and could be used to predict co-extinctions [54]	Distribution of woodpeckers improves species distribution models of boreal owls [10] Positive co-occurrence can be used to infer facilitative interactions [63] The probability of geographic co-occurrence is higher for (+/+) interactions [20]
Predation /Parasitism (+/-)	Body size is a determinant of consumer–resource interactions in aquatic food webs [64-66], and can also determine interaction strength [67] Shifts in body size affect type and strength of predator-prey interactions [68] Body temperature can have major implications of our understanding of how thermal stress modulates predator-prey relationships under field conditions [69]	Phylogenetic data has commonly been used to infer predation in paleontological data [70] The phylogenetic signal in host range can be used to predict which plant species are likely to be susceptible to a particular pest or pathogen [71] Phylogeny as a proxy for unmeasured trait information explains much of food web structure [66, 72]	Spatial distribution can affect disease interactions for the Foot-and-mouth livestock disease [73] In very specialized predator-prey interactions, prey geographic distribution might enhance predictions for the distribution of the predator [48]
Competition (-/-)	Functional composition of subalpine grasslands is related to facilitation when there is no resource limitation [59] The prevalence of competition versus environmental filtering can be inferred based on functional traits [74]	Plant interactions turn from facilitation into competition with increasing phylogenetic relatedness [75] Competition is assumed in community phylogenetics when communities are over-dispersed [76, 77] The prevalence of competition versus environmental filtering is not affected by phylogenetic relationships for tree species [74]	Pairwise negative residual spatial co-variation indicates potential competitive interactions [78] Negative co-occurrence patterns can be used to infer competitive interactions [63, 79]
Commensalism (+/0)	Behavioural trait composition can turn the interactions of spider *Anelosimus studiosus* into mutualism or commensalism [60]	Phylogenetic relationships do not seem to affect network structure in an epiphyte – phorophyte network [80]	Host plants distribution can improve distribution model for butterfly *Parnassius mnemosyne* [9] Species abundance, species spatial overlap and host size largely predicted pairwise interactions and several network metrics [80]
Amensalism (-/0)	Behavioural trait composition can be related to amensalism of spider *Anelosimus studiosus* with interacting species [60]		

Figure captions

Figure 1. Inferring biotic interactions from proxies and testing them with two examples of food webs from the Serengeti grassland ecosystem (a-e) and from Cuban coral reefs (f-j). The examples differ in the numbers of species and geographic realm but are comparable in scales of extent as data was pooled across large regions (>25,000Km²). Whereas the first example comprises 161 species inhabiting the Serengeti National Park, (north of Tanzania), the second comprises 265 species found in all Cuban coral reefs within the 100m depth platform surrounding the island. By applying three nested successive constrains the number of potential links was reduced by more than two thirds. The first step involved assigning species into three major trophic groups (b,g): carnivores (red circles), herbivores (dark green circles) and primary producers (light green circles). The second step involved classification of consumer species based on their diets (c,h). The third step involved refinement of groups by accounting for the geographic location of species or by establishing where producer species feed in environmental space (d,i). The trophic structure of the two examples is reversed, with more diversity of primary producers in the Serengeti. Yet, the ability to remove forbidden links and match empirical food webs is similar (i.e. 66.49% for the Serengeti and 67.74% for the Cuban coral reef).

Figure 2. Overview of the hierarchical framework to infer biotic interactions from proxies. (a) Initial saturated interaction matrix containing all species in the regional species pool and all possible directed links amongst them. (b) Constrained matrix after removal of forbidden links based on trait information. (c) Constrained matrix after removal of forbidden links based on geographical considerations. (d) Filtering of matrix based on estimated probabilities of interaction (see below) derived from traits, phylogenies and/or species geographical distributions. (e) Comparison of the inferred

interaction matrix against the actual (observed) interaction matrix to evaluate the accuracy of the inference. Probabilities of interaction between groups are derived from the following predictions. *P1:* The strength of competitive interactions between two species is inversely related to their proximity in trait or phylogenetic space. Functionally similar, or closely related, species are more likely to compete with each other. *P2:* Closely related species are more likely to share interactions with other species than distantly related species. *P3:* The probability that a species predates on another in certain food webs (i.e. marine food webs) is likely to increase with an increase of their body-size ratio. *P4:* The more similar the ecological niches of two species, the greater the probability that they will co-occur and hence the higher their interaction probability. *P5:* Related to predictions 4, interactions will also affect co-distribution. The current understanding of co-occurrence is that negative interactions will generate repulsion; positive interactions aggregation and exploitative interactions (+/-) will generate asymmetric co-occurrence (where one species is most often associated to the other and the other one less often). *P*6: Interaction strength will be proportional to the product of relative abundance of the two species (neutral expectation).

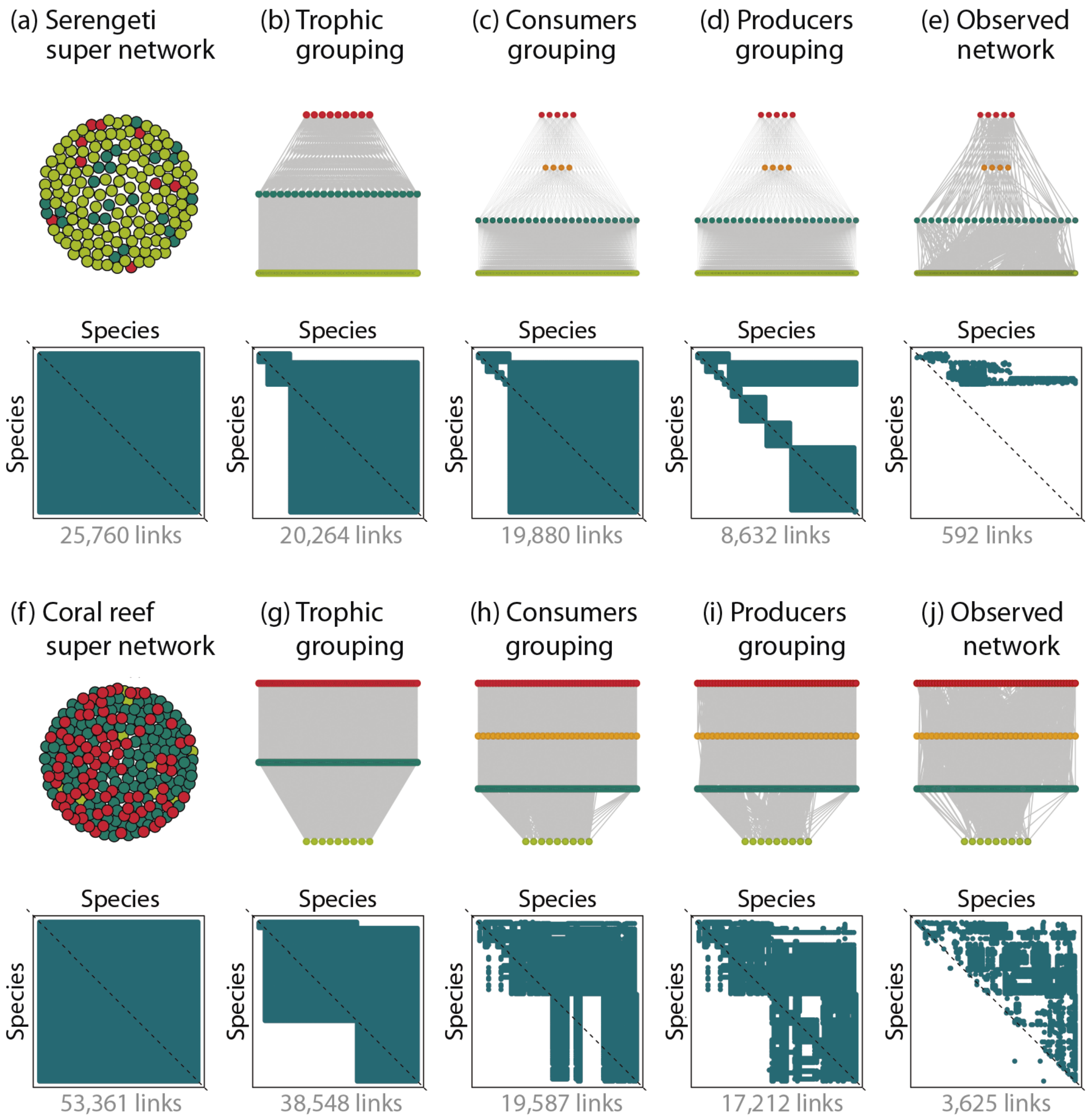
(a) Serengeti super network
(b) Trophic grouping
(c) Consumers grouping
(d) Producers grouping
(e) Observed network
Species
Species
25,760 links
20,264 links
19,880 links
8,632 links
592 links
(f) Coral reef super network
(g) Trophic grouping
(h) Consumers grouping
(i) Producers grouping
(j) Observed network
53,361 links
38,548 links
19,587 links
17,212 links
3,625 links

Figure 1

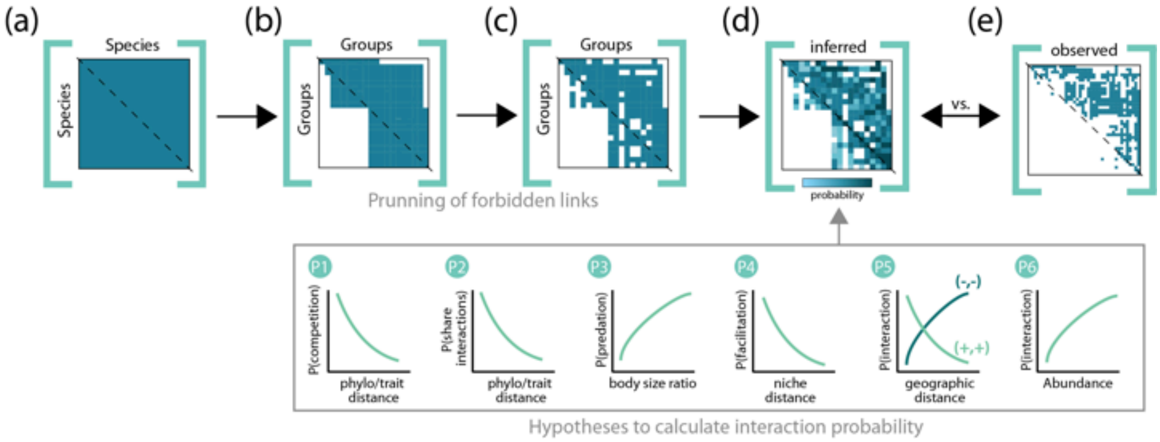
(a)
Species
Species
(b)
Groups
Groups
(c)
Groups
Groups
Prunning of forbidden links
(d)
inferred
probability
vs.
(e)
observed
P1
P(competition)
phylo/trait distance
P2
P(share interactions)
phylo/trait distance
P3
P(predation)
body size ratio
P4
P(facilitation)
niche distance
P5
P(interaction)
(-,-)
(+,+)
geographic distance
P6
P(interaction)
Abundance
Hypotheses to calculate interaction probability

Figure 2