Species co-occurrence networks: Can they reveal trophic and non-trophic interactions in ecological communities?

Mara A. Freilich [iD],[1,2,3,11] Evie Wieters,[4] Bernardo R. Broitman,[5,6]
Pablo A. Marquet,[3,7,8,9,10] and Sergio A. Navarrete[4,9]

[1]*Department of Earth, Atmospheric and Planetary Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA*
[2]*Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 USA*
[3]*Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Alameda 340,
C.P. 6513677 Santiago, Chile*
[4]*Estación Costera de Investigaciones Marinas, Departamento de Ecología, Center for Marine Conservation,
Pontificia Universidad Católica de Chile, Santiago, Chile*
[5]*Centro de Estudios Avanzados en Zonas Áridas, Ossandon 877, Coquimbo, Chile*
[6]*Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Coquimbo, Chile*
[7]*Instituto de Ecología y Biodiversidad (IEB), Las Palmeras 3425, Santiago, Chile*
[8]*Instituto de Sistemas Complejos de Valparaíso (ISCV), Artillería 470, Cerro Artillería, Valparaíso, Chile*
[9]*Laboratorio Internacional en Cambio Global (LINCGlobal), Centro de Cambio Global (PUCGlobal), Pontificia Universidad Catolica de
Chile, Alameda 340, C.P. 6513677 Santiago, Chile*
[10]*The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501 USA*

Abstract. Co-occurrence methods are increasingly utilized in ecology to infer networks of species interactions where detailed knowledge based on empirical studies is difficult to obtain. Their use is particularly common, but not restricted to, microbial networks constructed from metagenomic analyses. In this study, we test the efficacy of this procedure by comparing an inferred network constructed using spatially intensive co-occurrence data from the rocky intertidal zone in central Chile to a well-resolved, empirically based, species interaction network from the same region. We evaluated the overlap in the information provided by each network and the extent to which there is a bias for co-occurrence data to better detect known trophic or non-trophic, positive or negative interactions. We found a poor correspondence between the co-occurrence network and the known species interactions with overall sensitivity (probability of true link detection) equal to 0.469, and specificity (true non-interaction) equal to 0.527. The ability to detect interactions varied with interaction type. Positive non-trophic interactions such as commensalism and facilitation were detected at the highest rates. These results demonstrate that co-occurrence networks do not represent classical ecological networks in which interactions are defined by direct observations or experimental manipulations. Co-occurrence networks provide information about the joint spatial effects of environmental conditions, recruitment, and, to some extent, biotic interactions, and among the latter, they tend to better detect niche-expanding positive non-trophic interactions. Detection of links (sensitivity or specificity) was not higher for well-known intertidal keystone species than for the rest of consumers in the community. Thus, as observed in previous empirical and theoretical studies, patterns of interactions in co-occurrence networks must be interpreted with caution, especially when extending interaction-based ecological theory to interpret network variability and stability. Co-occurrence networks may be particularly valuable for analysis of community dynamics that blends interactions and environment, rather than pairwise interactions alone.

Key words: anthropogenic impacts; co-occurrence; ecological networks; food webs; intertidal; keystone species; non-trophic interactions; species interactions.

Introduction

Most past ecological research on the network of species interactions within communities has focused on small subsets of interacting species, and usually on those interactions that can be easily inferred from observation of an organism's diet or physical contact between species, such as predation, pollination, and fruit dispersal (Dunne et al. 2002, Montoya and Solé 2002, Bascompte and Jordano 2014). Since such approaches usually require extensive efforts in

data collection, and because many types of interactions or ecological systems cannot be readily observed (e.g., microbial communities, endoparasites), it is increasingly common to infer or reconstruct interaction networks from pattern in species co-occurrence in either time or space (Stephens et al. 2009, Araújo et al. 2011, Faust and Raes 2012, Borthagaray et al. 2014).

Species co-occurrence is a simple and long sought after approach to infer species interactions within ecological systems (Morales-Castilla et al. 2015, Cazelles et al. 2016, Sander et al. 2017). Significant spatial co-occurrence is considered evidence of positive or mutualistic interactions, and co-exclusion considered evidence of negative (e.g., competitive) interactions (Faust et al. 2015, Fuhrman et al. 2015). However, co-occurrence patterns within a single

ecosystem can also be interpreted in terms of species response to environmental factors (Peres-Neto et al. 2001) or dispersal limitation (Ulrich 2004). The premise of the co-occurrence approach is that if species in a community are interacting with each other in a way that affects each others' abundance or presence over space, thereby influencing local community assembly patterns, then they will have non-random co-occurrence that could be revealed through an appropriate sampling design and statistical tests (Ulrich and Gotelli 2013, Borthagaray et al. 2014). For example, predators might be observed with their prey more frequently, and competitors might be observed together less frequently than expected from random assembly. This approach is closely related to the development of assembly rules (Diamond 1975) and null models in ecology (Gotelli and Graves 1996). While our understanding of co-occurrence patterns and the processes underlying community assembly is much more sophisticated and multifactorial than originally envisioned (Chase 2010, de Bello et al. 2012, Cazelles et al. 2016), the basic premise from the point of view of building ecological networks from such patterns is essentially the same (Peres-Neto et al. 2001).

The pattern of interactions among members of an ecological community has consequences for population dynamics and persistence of species, for network stability and for the maintenance of ecological function (Allesina and Pascual 2008, Faust and Raes 2012, Slessarev et al. 2016). It is therefore important to understand to what extent patterns of co-occurrence of species (or Operational Taxonomic Units [OTUs] in the case of microbes) reflect species interactions. In other words, to what extent are ecological networks built from co-occurrence patterns, for microscopic or metazoan organisms, commensurate with those built through direct observation, such as gut content analyses, direct observations of consumption, stable isotope analyses, or experimental manipulations (Dunne et al. 2008, Kéfi et al. 2015, 2016, Sander et al. 2017)?

There are important ecological and methodological reasons as to why ecological interactions may not translate into easily discernible patterns of co-occurrence. The method of constructing ecological networks from co-occurrence data has explicit spatial assumptions built into its design, which should be scrutinized further so that we better understand the limitations of the approach when it comes to interpreting network structure and dynamics. Predator and prey are expected to positively correlate over some spatial scale that allows predators to maximize prey encounters. Yet, over some smaller scale, effective predators should reduce or completely eliminate prey, sometimes restricting them to refuges beyond predator reach, generating strong negative associations. The actual spatial scale over which the direction of species correlation changes across an environmental or biotic gradient will also change depending on biological attributes of the species involved, such as body size and dispersal capacity (Borthagaray et al. 2014). This makes it quite challenging to detect such correlations across multiple-species assemblages using a fixed sampling size. In addition, species interaction networks, especially food web networks, are directed, which means the matrix of species interactions can be asymmetric (Cazelles et al. 2016). Covariance

structures estimated from spatial co-occurrence can only infer symmetric interactions. Further, species can coexist and exhibit a correlation in their abundances through either time or space because they are affected by a third species (as in apparent competition, e.g., Holt and Bonsall 2017), or through a common environmental factor, even if the species pair does not interact directly. Moreover, spatial variability in dispersal and subsequent settlement and recruitment can by itself generate spatial correlation patterns between species (de Bello et al. 2012, Shinen and Navarrete 2014). The multi-species interactive nature of real communities, where each species simultaneously interacts with many others and in different ways (Berlow et al. 2004, Kéfi et al. 2016), adds additional complications when trying to construct patterns of paired interaction from co-occurrence (Azaele et al. 2010). In this context, it is of great importance to understand the robustness of the network reconstructions based upon co-occurrence data.

Despite these limitations, one can still argue that, if species interactions are important in determining the presence of species, then non-random patterns of co-occurrence must reflect the multiplicity of interactions, especially after controlling for environmental effects and indirect effects of third species (Peres-Neto et al. 2001, Azaele et al. 2010). Co-occurrence analyses can be particularly valuable because they have the potential to reveal which species respond in similar ways to ecosystem conditions. Indeed, many "true" observed links (e.g., a predator incidentally consuming a prey species) may be feeble and have little relevance in modulating prey abundance (but see Berlow 1999). In this sense, significant patterns of co-occurrence may reveal strong ecological interactions and filter out many weak effects.

Here we provide an empirical test of inferring species interactions from correlational studies using a well-resolved and comparatively specious empirically based ecological network from the wave exposed rocky shore of central Chile. We focus on examining the types of interactions for which there is the greatest correspondence between correlation-based and empirical approaches to species interactions. The only other study to conduct such an analysis focused on machine learning approaches and comparison to different ecosystems (Sander et al. 2017). We elucidate the aspects of the intertidal ecosystem that relate most directly to species co-occurrence.

We examine whether patterns of species co-occurrence, obtained through intensive field surveys conducted at multiple sites, may reflect the documented species interactions in the network. Importantly, not only consumptive interactions among the members (hereafter food web) of the local community have been considered, but also non-trophic interactions (hereafter NTI), such as facilitation, interference, or habitat provisioning have been described (Kéfi et al. 2015, 2016). Since many of the documented interactions in ecological communities are non-trophic, such as mutualistic relations and competition, we examine whether spatial co-occurrence patterns reflect best trophic, positive NTI or negative NTI type of interactions. Moreover, we assess whether known strong interactions are better resolved by spatial co-occurrence patterns, as well as assess the performance of the methods across strong environmental gradients.

Method

Network

The approach we follow here is to use a "known," independently constructed ecological network of species interactions for the wave-exposed rocky shore of central Chile (see Castilla and Durán 1985, and Castilla 1999 for an overview of this system) and examine whether such a network can be reconstructed from extensive species co-occurrence data. The interaction structure for this community is composed of the trophic interactions, negative NTIs, and positive NTIs recorded in Kéfi et al. (2015). Like all ecological networks, there is uncertainty as to the significance of some of the links connecting species in the network (see discussion in Kéfi et al. [2015]), but the network used here represents one of the most complete networks yet published in which both food web (trophic) interactions and non-trophic interactions are determined from long-term analysis of species' natural history, surveys and observations, and experimental studies. Further methodological details and analyses of the network structure can be found in Kéfi et al. 2015, 2016 and *online*.[12]

Field surveys in rocky shore communities

The spatial data is from surveys of species at multiple sites over multiple years along the rocky shores of central Chile. A total of 49 sites and 3,847 quadrats were considered in our analyses (Fig. 1). Out of these 49 sites, 46 sites were sampled during the time periods 1998–2000, 2003–2005, and 2010, 7–15 quadrats of 50 × 50 cm placed haphazardly along a 20–30 m long transect at each low and mid, and occasionally high, intertidal zones were sampled. This sampling effort was sufficient to capture the full species richness at each site (Broitman et al. 2011). Note that not every site was sampled every year and a different number of quadrats were sampled during each survey. Details of the surveys, methods and distribution of sites sampled can be found in Broitman et al. (2001, 2011), Navarrete et al. (2005), and Wieters et al. (2009). In 2013, three additional sites around the region of Copiapó (27° S) were surveyed. This time, 25 quadrats of 50 × 50 cm at three different tidal levels (high, mid, and low tide) along 50-m transects were sampled. For all surveys, mobile species were counted as individual organisms while sessile species were recorded as percent cover.

Inferring links from spatial structure

We used Pearson correlations on presence/absence data to test for spatial association between species. The sign and magnitude of the association is that of the correlation coefficient. Note that for presence/absence data, Pearson correlation is equivalent to Spearman and Kendall's tau correlations. We assigned a $\alpha = 0.1$ significance threshold, privileging a greater number of significant links over precision. The results are insensitive to alterations in this value, however it should be noted that 10% of the inferred associations are likely due to statistical effects.
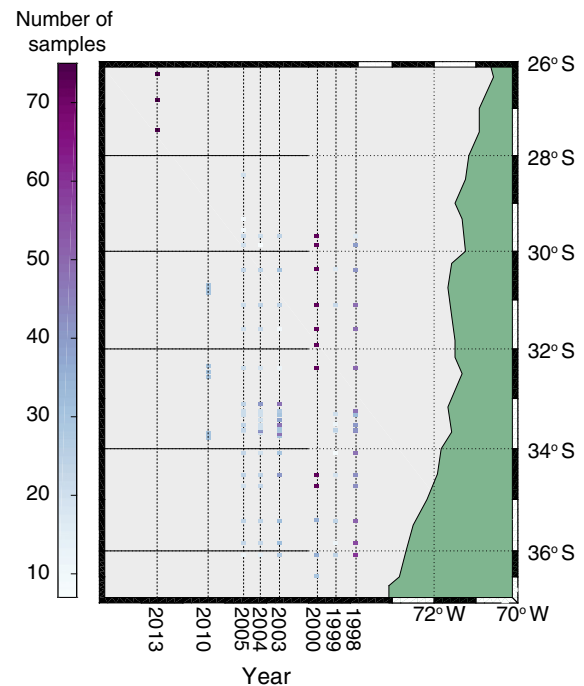
Fig. 1. Community surveys were done at 49 distinct coastal intertidal sites during the years 1998–2013. Not all sites were surveyed in all years and a different number of samples were taken during each survey. The figure indicates the latitude of sampling sites, which should be projected on the coast for the actual sampling location.

In *Results*, and for ease of presentation, we describe to what extent the species interactions recovered from co-occurrence data revealed "true" links as previously described in the intertidal ecological network, using the Pearson correlation approach and note the differences with the null model approach. Of course, as discussed above, identification of such "true" links are not free of assumptions (see Kéfi et al. [2015] for detailed discussion). We adopt the terminology of "interaction" for links in the "true" network and "association" for links in the co-occurrence network.

We present heat maps of the full results in the Appendix S1: Figs. S2–S6 and present and discuss condensed versions of these results in the body of the manuscript.

Sensitivity, or the probability of detection of a true link, is computed as the ratio of the number of links that are correctly identified as ecological interactions (true positives) to the total number of links in the empirical network. Specificity, or the true negative rate, is the ratio of the number of links that are correctly identified non-interactions (true negatives) to the total number of non-interactions in the empirical network. Sensitivity and specificity are defined for any given subnetwork, including for single species, by counting only the links that connect to species within that subnetwork (links may originate or end outside of the subnetwork).

There is a plethora of approaches for reconstructing species interactions from presence/absence data, but most have the covariance matrix, the object of analysis in this study, as a common underpinning. An alternative conceptual approach, the "probabilistic approach," is to measure association based on significant deviations from the expected

probability of co-occurrence of two species based on the occurrence of single species (Araújo et al. 2011, Veech 2014). Significant deviation from the expectation can be determined using a null model approach (Gotelli and Graves 1996), or an exact P value if there are very few samples (Veech 2014). To complement our analysis, we used both the correlation and the probabilistic approach described here.

Strong interactions

The keystone species in this ecosystem that have been identified through experimental manipulations (Paine 1966, Castilla and Durán 1985, Oliva and Castilla 1986, Navarrete and Castilla 2003, Aguilera and Navarrete 2012) are carnivores: the seastar *Heliaster helianthus* and the muricid whelk *Concholepas concholepas*, and grazers: *Fissurella crassa* and *Fissurella limbata*. It should be noted that these grazers are highly omnivorous (Camus et al. 2008, 2013). There are also 14 species that are harvested by humans, including three of the keystone species (*C. concholepas*, *F. crassa*, and *F. limbata*).

RESULTS

Inferring links of the ecological network from non-random co-occurrence

The probabilistic approach performs substantially worse than the correlation-based approach due to a high false-positive rate (specificity of 0.283). It has slightly higher sensitivity (0.692) than the correlation-based approach, particularly for rare species, but much lower specificity. Consequently, we focus on analyzing the performance of the better-performing correlation method and present the results of the probabilistic approach in Appendix S1: Figs. S7, S8.

The overall sensitivity using the correlation-based approach with a significance of $\alpha = 0.1$ is 0.469, meaning just under one-half of the interactions in the empirical network are detected as significant associations. The specificity is 0.527, meaning that slightly over one-half of the detected

non-associations are not interactions. The lowest specificity (0.367) was in the sessile-sessile interactions, which are mostly negative NTIs. In comparison to other interactions, there are relatively few interactions where a sessile species affects a mobile species and both specificity (0.558) and sensitivity (0.531) are highest for these interactions.

As a species becomes more common across the region, sensitivity increases rapidly at first and then appears to level off around an occurrence of 1,000 quadrats out of the total of 3,847 quadrats (Fig. 2a). Similarly, specificity rapidly declines, i.e., more false or spurious interactions are detected with increasing occurrence of the species in the field surveys up to about 1,000 quadrats, where it levels off to between 0.2 and 0.3 (20–30% of non-interactions correctly classified as non-interactions; Fig. 2b). The specificity does not fall off as quickly as sensitivity increases; a linear best fit between the two metrics for each individual species compared to all interaction types has a slope of -0.861 (Fig. 2c). The detection is best for positive non-trophic interactions (Fig. 2c, purple line). We find no relationship between the total number of potential interactions and the average effect size (Appendix S1: Fig. S1).

Detection by interaction type

The ability to correctly detect a true link (i.e., sensitivity) varies across different interaction types with positive non-trophic interactions being most detectable by co-occurrence. Of the known positive non-trophic interactions, 77.4% were detected (Fig. 3). Negative non-trophic interactions and trophic interactions are less detected with 46.7% and 44.4% of the known interactions detected, respectively (Fig. 3).

At the spatial scale of the quadrats (0.25 m^2), the sign of significant correlation coefficients that correspond with trophic interactions is mainly positive, suggesting that the co-occurrence approach is best able to detect instances where species co-occur with their prey more often than they exclude their prey (Fig. 4c). This may be a general principle for this system, but we would need to collect additional evidence to support this hypothesis. The strongest positive
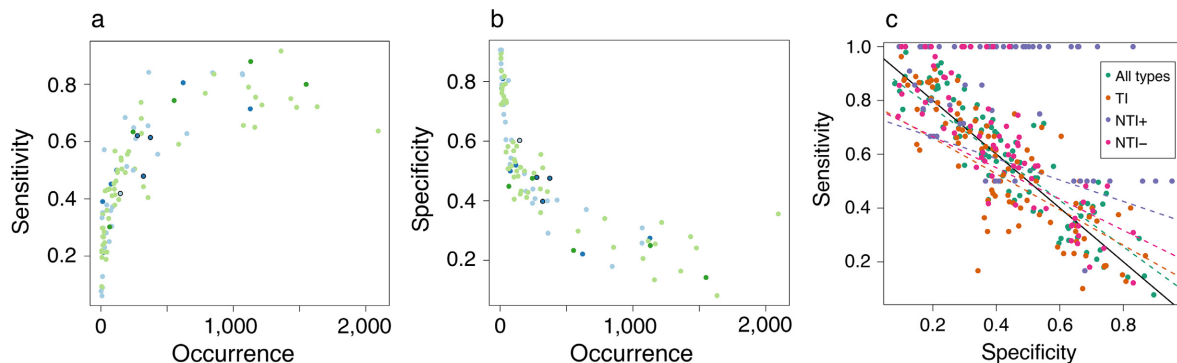

FIG. 2. (a) Sensitivity, or percentage of links detected plotted against the total number of occurrences for a given species. (b) Specificity, or percentage of non-interactions detected as non-interactions plotted against the total number of occurrences for a given species. In these plots, each point is a different species. Blue dots are mobile species, green dots are sessile species. Darker dots are harvested species and species outlined in black are keystone species. (c) Specificity plotted against sensitivity. Each point is a different species. The colors indicate which network, trophic interactions (TI), positive non-trophic interactions (NTI+), negative non-trophic interactions (NTI−), or all interaction types is used as the "true" network. The dashed lines are best fit lines for each interaction type. The black line is a 1:1 line. Points above the 1:1 have better detection than random while point below the 1:1 line have worse detection than average.
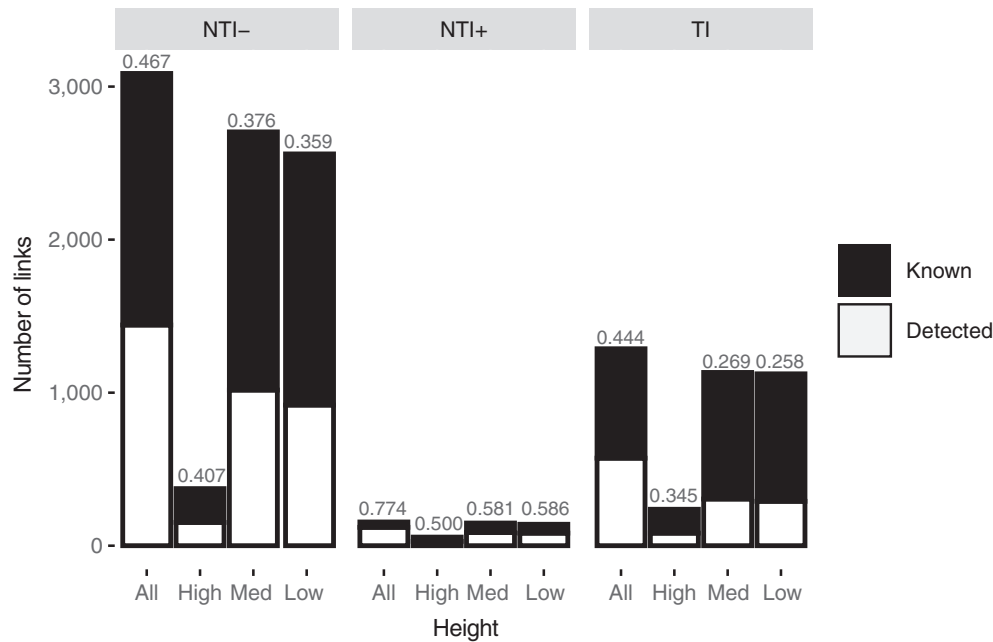
Fɪɢ. 3.   The bars show the number of links of each type including species found at all heights along the intertidal and species only present in the high, medium (med), and low heights in all samples. The white bars show the number of links detected using co-occurrence of species. Proportions above each bar are the proportion of links of each type detected using co-occurrence.

association observed was between the kelp *Lessonia* spp. (*L. spicata* and *L. beteroana*) and the grazing limpet *S. scurra*, which lives almost exclusively on the kelp. These species thus also share a positive non-trophic interaction. Almost all of the asymmetrical interactions in the interaction network are trophic interactions. Only 614 of the significant associations correspond to asymmetrical interactions; consequently 614 of 2,888 false positives are possibly accounted for by asymmetry in interactions.

Of the 120 positive non-trophic interactions detected as a significant association, 14 were detected as negative associations (Fig. 4d). Of these, four were species pairs that have both positive and negative non-trophic interactions (*Corallina officinalis–Perumytilus purpuratus*, *Phragmatopoma* spp.–*Ulva rigida*, *Porphyra* spp.–*Semimytilus algosus, Semimytilus algosus–Porphyra* spp.). Four of the positive NTIs that were detected as significant negative correlations are positive interactions initiated by *P. purpuratus*, which is both a strong competitor for space and a habitat engineering species.

Negative non-trophic interactions detected as significant associations in the co-occurrence analysis were overwhelmingly and incorrectly identified as positive associations (Fig. 4a). In the low intertidal zone, negative non-trophic interactions were more likely to be identified as negative associations in the surveys than at any other tidal height.

The intertidal stress gradient

The gradient from low to high intertidal zone represents a strong environmental stress gradient for intertidal organisms of marine origin (Connell 1961). We therefore performed separate analyses per tidal height, which is one way of controlling for or reducing the influence of environmental conditions.

Overall, a smaller percentage of the links were detected when separate analyses were conducted for the high, mid, and low shore heights, especially a lower percentage of positive non-trophic interactions and of trophic interactions (Fig. 3). The lower percentage of links detected is due, at least in part, to reduction in statistical power due to the smaller sample sizes. There are 1,610 quadrats for each of the low and mid tidal heights and 627 quadrats at the high tidal height. In order to control for the effects of reducing the sample size on the statistical power, we performed the co-occurrence analysis on 500 random subsamples with the same number of quadrats as were available for each shore height, 1,610 quadrats for comparison to the low and mid tidal zone and 627 quadrats for comparison to the high tidal zone. For all interaction types, a lower percentage of interactions were detected with a random subsample than with all samples, indicating that there is a reduction of statistical power with a reduced sample size. From these analyses, we found that species co-occurrence detected 34.6% ± 0.0187% (mean ± SD) of the negative non-trophic interactions with 627 randomly selected samples and 38.2% ± 0.0152% of the negative non-trophic interactions with 1,610 samples, 60.2% ± 0.0338% of the positive non-trophic interactions with 627 samples and 66.3% ± 0.0616% of the positive non-trophic interactions with 1,610 samples. Finally, species co-occurrence detected 28.0% ± 0.0141% of the trophic interactions with 627 samples and 33.8% ± 0.010% of the trophic interactions with 1,610 samples. These numbers should be compared with Fig. 3 to understand the interaction of tidal height and interaction type for detectability. Co-occurrence using just samples from the high tidal height detects a higher proportion of the known negative non-trophic and trophic interactions than a sample across the environmental gradient of the same size. Using the low and
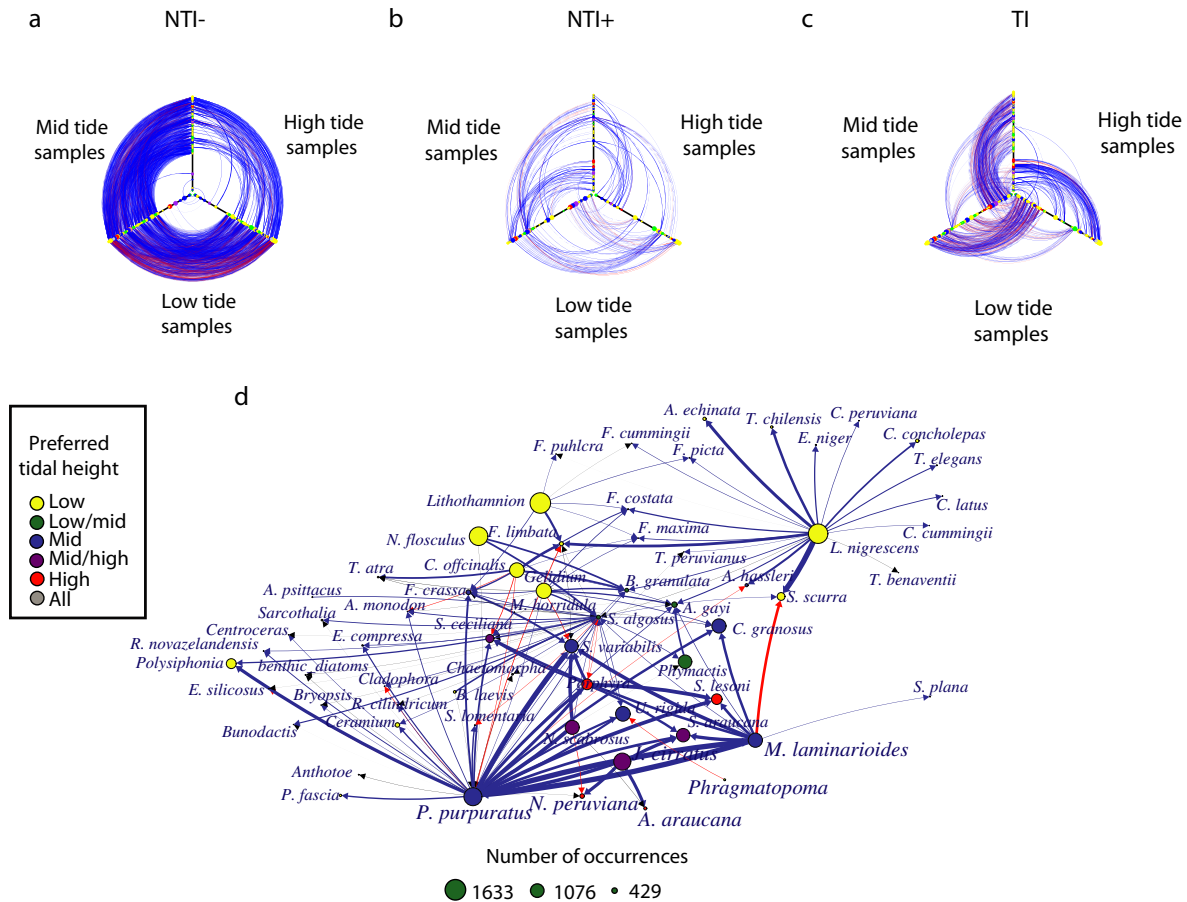
Fig. 4. Links between species inferred using three different data subsets (only high tide samples, only mid tide samples, and only low tide samples) for each of three interaction types (a) negative non-trophic interactions, (b) positive non-trophic interactions, and (c) trophic interactions. Species are arranged on the axes from inward to outward by high to low trophic position. The preferred tidal height of each species is indicated by the node color. Blue edges indicate an inferred positive association while red edges indicate an inferred negative association. (d) Network showing all positive non-trophic interactions. Blue edges indicate a positive association inferred by co-occurrence, red edges indicate a negative association inferred by co-occurrence, and black edges indicate no association. The node size represents the number of samples in which each species occurred while node color is the preferred tidal level for each species.

mid tidal elevation samples, co-occurrence detects fewer links than in a sample of the same size across the environmental gradient for all interaction types.

Indirect interactions

The empirical network is dense, with 4,458 links representing both trophic and non-trophic interactions, resulting in a connectance of 0.47. Consequently, indirect interactions are extremely abundant in this system. At path length 4, the network is complete (every species is connected to every other species by a path of length 4). An indirect interaction of path length 2 is an interaction between two species that is mediated by a third species (e.g., species A and species B have an interaction, species C and species B have an interaction, so the interaction between species A and species C is a path length two indirect interaction). There are 8,833 interactions of path length two; of these, 1,537 correspond to significant co-occurrences that did not correspond to links in the empirical networks of known direct interactions ("false positives"), making them likely candidates for pairwise co-occurrence

driven by interactions with a third species. One example of an indirect interaction is competition for prey. In this network, 1,324 path length 2 indirect interactions are between predators that share prey. Of these 1,324 links, 496 links were discovered as significant association between species.

Strong interactions: keystone species, anthropogenic effects, and effect size

We did not find that the keystone species are distinguishable in either the specificity or sensitivity from other species in the community. Harvested species tend to have a slightly larger effect size (stronger correlation) for a given node degree (Fig. 5), whereas no noticeable distinction between keystone and other species was observed in terms of the effect size (Fig. 5). The median effect size of correlations increases with increasing detected node degree and sessile species show a more gradual increasing trend than mobile species. The species with the largest median effect sizes were small generalist herbivores, the barnacles *N. scabrosus* and *J. cirratus* and the algae *M. laminarioides* and *Porphyra* spp.
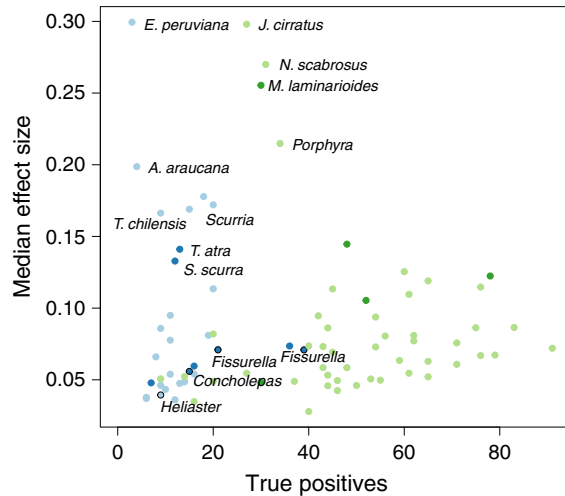
Fig. 5. True positives, plotted against the median effect size (correlation intensity) of all detected links. A true positive is a known species interaction that is also detected as a significant association. This analysis only includes the most common species (occurrence greater than 62 quadrats). Sixty-two occurrences was chosen, because it is the inflection point in Fig. 2a and b, in order to avoid the loss of power at low occurrence. Blue dots are mobile species, green dots are sessile species. Darker dots are harvested species and species outlined in black are keystone species.

DISCUSSION

The construction of ecological networks from patterns of species co-occurrences is rapidly expanding in the microbial ecological and biomedical sciences (Fuhrman et al. 2015), and spreading to metazoan communities (Araújo et al. 2011, Borthagaray et al. 2014). There is no doubt that such co-occurrence networks reveal aspects of a local community that have important bearing on network dynamics, stability, and resilience (Faust and Raes 2012), and, in the case of microbial environmental genomic studies, there are as of now few other sensible approaches to get a glimpse into the complex matrix of interactions among the members of these highly diverse communities. Our results do not dispute the importance of such networks. However, in many ways, authors have interpreted co-occurrence as revealing "classical" species interaction networks (e.g., competition, predation, facilitation). Our results complement and expand recent empirical findings (Sander et al. 2017), which suggest this to be highly unlikely by using a spatially extensive data set that reflects the type of presence–absence data that is commonly available for co-occurrence studies.

Inferring links

Given the multiplicity of ecological and environmental processes that affect species assemblages, the prevalence of weak links in the empirical network, and the potentially large number of species in this community, the result that many species interactions are still detectable, with 25–70% of species interactions detected, depending on the interaction type and data subset used, is indicative of a strong role of species interactions in this community.

There are distinct reasons, statistical and artefactual, for why there may be false positives as opposed to false negatives.

The correlation test can only infer symmetric interactions but true interactions can be asymmetrical, leading to false positives. In addition, some of the false positives could represent correlations based on shared environmental preference, common settlement patterns, or indirect interactions. Indirect interactions could also result in false negatives, if multiple interactions cancel out (Cazelles et al. 2016). There may be other reasons for the lack of sensitivity, such as a priority effect operating on sessile or highly territorial species, for instance. If a priority effect is important, then the order of arrival of larva to the shore, a stochastic process that may also depend on species-specific life history traits, may matter more for coexistence than does the outcome of competitive interactions among juveniles or adults once on the shore (Berkley et al. 2010, Aiken and Navarrete 2014, Oróstica et al. 2014).

Increasing occurrence of individual species decreases the specificity while increasing sensitivity. This diminishing return has both statistical and biological explanations. As occurrence increases, more information is available about new species combinations, but as occurrence asymptotes to match the number of samples, there is less information available again. The saturating relationships in Fig. 2a and b is likely a unimodal relationship with zero sensitivity as occurrence approaches 1 and as occurrence approaches the number of samples. We cannot verify this hypothesis since no species were present in all samples. This suggests that there is a limit to detectability of interactions when using occurrence alone, a limitation that in theory could be resolved with quantitative abundance data. Previous studies have attributed similar results based on occurrences to significant ecological processes, hypothesizing that less common species are more affected by biotic interactions than by environmental preferences (Azaele et al. 2010). This may be the reason why including both biotic and abiotic predictor variables usually improves the modeling of species interactions and spatial distribution (González-Salazar et al. 2013, Stephens et al. 2017). In our system, we have no information that could help us test this hypothesis. Further work along this research direction should be encouraged.

We found that the Pearson's correlation test performed better than the probabilistic approach. The robustness of Pearson's correlation has been noted in other studies. Notably, Sander et al. (2017) found that two machine-learning approaches were not superior to using Pearson's correlation.

Niche-based and interaction-based processes

Positive non-trophic interactions are more detectable than either negative non-trophic interactions or trophic interactions. The fact that over 70% of positive non-trophic interactions were correctly detected indicates the usefulness of occurrence data to infer species interactions. A similar result was also obtained by Sander et al. (2017) using occurrence data and both Pearson's correlation and Dynamic Bayesian Networks to infer non-trophic interactions (positive and negative pooled), which suggest that the increased sensitivity of occurrence data for this type of data may be a general phenomenon. Positive non-trophic interactions generally ameliorate environmental stress or provide habitat for other species, in effect expanding the possible niche for the species involved (Jones et al. 1994, Wright et al. 2002, Hastings et al.

2007, Stachowicz 2012). Trophic and non-trophic interactions can only operate within the niche defined by environmental constraints. This may be a key attribute of positive non-trophic interactions that makes them more detectable when sampling across space and environmental variability.

Power to detect species interactions decreases when controlling for tidal height except in the high tidal zone. Many of the detected co-occurrences may be mainly driven by shared environmental preferences rather than species interactions. Controlling for tidal height might have been expected to increase the detectability of interactions because the interaction network is modular with the modules related to the height at which species are found (Kéfi et al. 2015, 2016). However, many strong interactions occur between tidal levels and lead to tidal segregation of species as shown in Chile and in other rocky shore communities, i.e., predators delimiting lower end of mussels (Paine 1966, Castilla and Durán 1985, Menge et al. 1994, Menge et al. 2004), competitive monopolization or domination of the mid-tidal zone restricting other species to higher or lower elevations (Connell 1961, Navarrete and Castilla 1990, Berlow and Navarrete 1997, Branch and Steffani 2004). This tidal segregation between strong interactors would go undetected when examining within tidal levels. On the other hand, the overlap between environmental preference and potential species interactions may artificially inflate detection of species interactions when samples from all tidal heights are included.

Habitat preferences and environmental processes may be especially important for shaping species distributions if biotic interactions are weak (Shinen and Navarrete 2014). There are considerably more negative non-trophic interactions and trophic interactions in the empirical network than there are positive non-trophic interactions, so there may also be proportionally more weak negative non-trophic and trophic interactions (Berlow et al. 2004, Lopez et al. 2017). Weak interactions may have either less of an impact or a more variable impact on the spatial structure of the community (Berlow 1999). While weak interactions can be quite important for population dynamics and community stability, they may be less detectable using co-occurrence alone, especially if there are many and diverse weak interactions. One implication of this work is that studies based on pairwise interactions may not be generalizable to understanding the whole community composition.

Indirect interactions

The pattern of species co-occurrence is affected by indirect interactions (those mediated by a third species) in addition to direct interactions (Cazelles et al. 2016). Exploitation competition for prey might affect predator co-occurrence, while the effects of multiple consumers on a single prey species might blur the relationship between a single consumer species and its prey. In terms of non-trophic interactions, a species' competitors are also likely competing with each other due to the density of the negative non-trophic interaction network.

Indirect interactions mediated by the habitat provisioning species can be very strong. For example, of the species that have significant associations with both *Lessonia* spp. and *P. purpuratus*, 55 of 64 species have negative association with one and a positive association with the other. *Lessonia* and *P. purpuratus* compete for space on the low shore and have a negative non-trophic interaction. The reversed sign of the association might be an indirect interaction mediated by the competition between *Lessonia* and *P. purpuratus*.

Lack of significant co-occurrence could be generated by indirect interactions if multiple interactions between species cancel the effects of pairwise interactions. If this were the case, one would expect that species with many interactions (high node degree) would have fewer links detected than species with fewer interactions (low node degree). In this analysis, we find no relationship between total node degree and the percent of links detected. There is a weak inverse relationship between outgoing node degree and the percent of links detected, suggesting that specialist predators are more likely to co-occur with their prey, since trophic interactions are the main asymmetrical interactions in this network. This is most likely related to mobility patterns of these predators.

Strong interactions: keystone species, anthropogenic influence, and effect size

Harvested species have a larger effect size of realized co-occurrences than would be expected based on the number of links detected. This might be related to perturbations in the system aiding detection because species interactions are most evident as the species occurrences return to equilibrium after a disturbance. It may also reflect the fact that humans usually remove the larger bodied species within a given assemblage, which may have stronger effects than other species in the assemblage.

The median of the absolute value of the effect size increases as the number of detected interactions increases (Fig. 5) and there is no relationship between the total number of potential interactions and the average effect size (Appendix S1: Fig. S1). These results largely contradict the assertion by Cazelles et al. (2016) that "the strength of an interaction decreases with the total number of interactions a species experiences." The trend of increasing effect size as the detected node degree increases could be statistical; when there are many links, strong links are more likely to be detected. In sectors of the interaction matrix with few interactions (e.g., sessile species affecting mobile species), there is high specificity and high sensitivity, indicating that when there are fewer interactions at the community level, interactions may be more detectable. More generally, these results indicate that co-occurrence may mostly identify interactions above a certain threshold and miss truly weak links.

We find that small grazers have the largest median effect size of correlations. This is in line with the conclusions of Borthagaray et al. 2014 that smaller species form tightly linked subgroups (here manifested as large effect size) and could be an effect of the relatively small quadrats, which may capture the co-occurrence of smaller-sized species better than larger species. There is also slightly higher edge density among the species in the steeply increasing section of Fig. 5 (0.25) as compared to edge density between those species and the species on the slowly increasing section of the trend (0.11). We additionally demonstrate that habitat engineering sessile species can also form tightly linked subgroups, but that not all sessile species form tightly

linked groups. Most sessile species have relatively weak co-occurrence with other species.

We do not find a distinct signal of keystone species. We would have expected keystone species to have a large effect size (Menge et al. 1994, Power et al. 1996), however, it is possible that keystone species instead have large effect size only when abundance is taken into account, or they may have strong effects only on a small subset of the species with which they interact. Moreover, it is possible that the effect of keystone species may be more noticeable on indirect than direct interactions. It is important to keep in mind, however, that the ability to detect a predator-prey interaction using spatial co-occurrence depends not only on the strength of the interaction, but also on how homogeneous or variable are across space the other factors that simultaneously influence species distribution, such as recruitment and environmental tolerances.

Conclusions

Co-occurrence networks do not reproduce interaction networks, but they do provide interesting and interpretable information about community assembly. In cases in which spatial or environmental effects operate synergistically with species interactions to determine the presence and absence of species (or other interacting biological component such as a protein or an OTU), a co-occurrence network can be a valuable object of analysis, albeit at times difficult to interpret due to the influence of indirect interactions and stochastic processes.

Both ecological and statistical effects can limit the interpretability of co-occurrence networks. As the occurrence of a given species increases, the probability of detecting a greater number of statistically significant associations increases as well. However, these associations may not necessarily correspond to interactions; both the number of true positives and false positives increases as the occurrence of a species increases. In terms of the important ecological effects, both environmental effects, including here recruitment, and species interactions determine the species range in the intertidal zone. Consequently, it can be difficult to disentangle which associations are related to interactions and which are related only to shared environmental preferences or correlated settlement. We suggest that environmental and settlement preferences may possibly outweigh biotic interactions in determining whole community co-occurrences. This is not to say that environment is more important in structuring communities than biotic interactions but that environmental variability may leave a more discernable signal in spatial co-occurrence patterns. Having said this, habitat engineering species and non-trophic positive interactions may leave a more detectable signal than other interaction types because they expand or create niche spaces for the species with which they interact.

In summary, ecological patterns observed in co-occurrence networks must be interpreted with caution, especially when extending interaction-based ecological theory to interpret network variability and stability. Co-occurrence networks may be particularly valuable for analysis of community dynamics as an epiphenomenon combining interactions and environment, rather than simply as the result of pairwise interactions.

Literature Cited

Aguilera, M. A., and S. A. Navarrete. 2012. Interspecific competition for shelters in territorial and gregarious intertidal grazers: consequences for individual behaviour. PLoS ONE 7:e46205.

Aiken, C. M., and S. A. Navarrete. 2014. Coexistence of competitors in marine metacommunities: environmental variability, edge effects, and the dispersal niche. Ecology 95:2289–2302.

Allesina, S., and M. Pascual. 2008. Network structure, predator-prey modules, and stability in large food webs. Theoretical Ecology 1:55–64.

Araújo, M. B., A. Rozenfeld, C. Rahbek, and P. A. Marquet. 2011. Using species co-occurrence networks to assess the impacts of climate change. Ecography 34:897–908.

Azaele, S., R. Muneepeerakul, A. Rinaldo, and I. Rodriguez-Iturbe. 2010. Inferring plant ecosystem organization from species occurrences. Journal of Theoretical Biology 262:323–329.

Bascompte, J., and P. Jordano. 2014. Mutualistic networks. Princeton University Press, Princeton, New Jersey, USA.

Berkley, H. A., B. E. Kendall, S. Mitarai, and D. A. Siegel. 2010. Turbulent dispersal promotes species coexistence. Ecology Letters 13:360–371.

Berlow, E. L. 1999. Strong effects of weak interactions in ecological communities. Nature 398:330–334.

Berlow, E. L., and S. A. Navarrete. 1997. Spatial and temporal variation in rocky intertidal community organization: lessons from repeating field experiments. Journal of Experimental Marine Biology and Ecology 214:195–229.

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

Borthagaray, A. I., M. Arim, and P. A. Marquet. 2014. Inferring species roles in metacommunity structure from species co-occurrence networks. Proceedings of the Royal Society B 281:20141425.

Branch, G. M., and C. N. Steffani. 2004. Can we predict the effects of alien species? A case-history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). Journal of Experimental Marine Biology and Ecology 300:189–215.

Broitman, B. R., S. A. Navarrete, F. Smith, and S. D. Gaines. 2001. Geographic variation of southeastern Pacific intertidal communities. Marine Ecology Progress Series 224:21–34.

Broitman, B. R., F. Veliz, T. Manzur, E. A. Wieters, G. R. Finke, P. A. Fornes, N. Valdivia, and S. A. Navarrete. 2011. Geographic variation in diversity of wave exposed rocky intertidal communities along central Chile. Revista Chilena de Historia Natural 84:143–154.

Camus, P. A., K. Daroch, and L. F. Opazo. 2008. Potential for omnivory and apparent intraguild predation in rocky intertidal herbivore assemblages from northern Chile. Marine Ecology Progress Series 361:35–45.

Camus, P. A., P. A. Arancibia, and I. Ávila-Thieme. 2013. A trophic characterization of intertidal consumers on Chilean rocky shores. Revista de Biología Marina Y Oceanografía 48:431–450.

Castilla, J. C. 1999. Coastal marine communities: trends and perspectives from human-exclusion experiments. Trends in Ecology and Evolution 14:280–283.

Castilla, J. C., and L. R. Durán. 1985. Human exclusion from the rocky intertidal zone of central Chile: the effects on *Concholepas concholepas* (Gastropoda). Oikos 39:1–399.

Cazelles, K., M. B. Araújo, N. Mouquet, and D. Gravel. 2016. A theory for species co-occurrence in interaction networks. Theoretical Ecology 9:39–48.

Chase, J. M. 2010. Stochastic community assembly causes higher biodiversity in more productive environments. Science 328:1388–1391.

Connell, J. H. 1961. The influence of interspecific competition and other factors on the distribution of the barnacle *Chthamalus stellatus*. Ecology 42:710–723.

de Bello, F., et al. 2012. Functional species pool framework to test for biotic effects on community assembly. Ecology 93:2263–2273.

Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 *in* M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Belknap Press, Cambridge, Massachusetts, USA.

Dunne, J. A., R. J. Williams, and N. D. Martinez. 2002. Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecology Letters 5:558–567.

Dunne, J. A., R. J. Williams, N. D. Martinez, R. A. Wood, and D. H. Erwin. 2008. Compilation and network analyses of Cambrian food webs. PLoS Biology 5:e102.

Faust, K., L. Lahti, D. Gonze, W. M. de Vos, and J. Raes. 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology 25:56–66.

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

Fuhrman, J. A., J. A. Cram, and D. M. Needham. 2015. Marine microbial community dynamics and their ecological interpretation. Nature Reviews Microbiology 13:133–146.

González-Salazar, C., C. R. Stephens, and P. A. Marquet. 2013. Comparing the relative contributions of biotic and abiotic factors as mediators of species' distributions. Ecological Modelling 248:57–70.

Gotelli, N. J., and G. R. Graves. 1996. Null models in ecology. Smithsonian Institution Press, Washington, D.C., USA.

Hastings, A., J. E. Byers, J. A. Crooks, K. Cuddington, C. G. Jones, J. G. Lambrinos, T. S. Talley, and W. G. Wilson. 2007. Ecosystem engineering in space and time. Ecology Letters 10:153–164.

Holt, R. D., and M. B. Bonsall. 2017. Apparent competition. Annual Review of Ecology, Evolution, and Systematics 48:447–471.

Jones, C. G., J. H. Lawton, and M. Shachak. 1994. Organisms as ecosystem engineers. Oikos 69:373–386.

Kéfi, S., E. L. Berlow, E. A. Wieters, L. N. Joppa, S. A. Wood, U. Brose, and S. A. Navarrete. 2015. Network structure beyond food webs: mapping non-trophic and trophic interactions on Chilean rocky shores. Ecology 96:291–303.

Kéfi, S., V. Miele, E. A. Wieters, S. A. Navarrete, and E. L. Berlow. 2016. How structured is the entangled bank? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience. PLoS Biology 14:e1002527.

Lopez, D. N., P. A. Camus, N. Valdivia, and S. A. Estay. 2017. High temporal variability in the occurrence of consumer-resource interactions in ecology networks. Oikos 126:1699–1707.

Menge, B. A., E. L. Berlow, C. A. Blanchette, S. A. Navarrete, and S. B. Yamada. 1994. The Keystone species concept: variation in interaction strength in a rocky intertidal habitat. Ecological Monographs 64:249–286.

Menge, B. A., C. Blanchette, P. Raimondi, T. Freidenburg, S. Gaines, J. Lubchenco, D. Lohse, G. Hudson, M. Foley, and J. Pamplin. 2004. Species interaction strength: testing model predictions along an upwelling gradient. Ecological Monographs 74:663–684.

Montoya, J. M., and R. V. Solé. 2002. Small world patterns in food webs. Journal of Theoretical Biology 214:405–412.

Morales-Castilla, I., M. G. Matias, D. Gravel, and M. B. Araujo. 2015. Inferring biotic interactions from proxies. Trends in Ecology & Evolution 30:347–356.

Navarrete, S. A., and J. C. Castilla. 1990. Barnacle walls as mediators of intertidal mussel recruitment: effects of patch size on the utilization of space. Marine Ecology Progress Series. Oldendorf 68:113–119.

Navarrete, S. A., and J. C. Castilla. 2003. Experimental determination of predation intensity in an intertidal predator guild: dominant versus subordinate prey. Oikos 100:251–262.

Navarrete, S. A., E. A. Wieters, B. R. Broitman, and J. C. Castilla. 2005. Scales of benthic–pelagic coupling and the intensity of species interactions: from recruitment limitation to top-down control. Proceedings of the National Academy of Sciences USA 102:18046–18051.

Oliva, D., and J. C. Castilla. 1986. The effect of human exclusion on the population structure of key-hole limpets *Fissurella crassa* and *F. limbata* on the coast of central Chile. Marine Ecology 7:201–217.

Oróstica, M. H., M. A. Aguilera, G. A. Donoso, J. A. Vásquez, and B. R. Broitman. 2014. Effect of grazing on distribution and recovery of harvested stands of *Lessonia berteroana* kelp in northern Chile. Marine Ecology Progress Series 511:71–82.

Paine, R. T. 1966. Food web complexity and species diversity. American Naturalist 100:65–75.

Peres-Neto, P. R., J. D. Olden, and D. A. Jackson. 2001. Environmentally constrained null models: site suitability as occupancy criterion. Oikos 93:110–120.

Power, M. E., D. Tilman, J. A. Estes, B. A. Menge, W. J. Bond, S. Mills, G. Daily, J. C. Castilla, J. Lubchenco, and R. T. Paine. 1996. Challenges in the quest for keystones. BioScience 46:609–620.

Sander, E. L., J. T. Wootton, and S. Allesina. 2017. Ecological network inference from long-term presence-absence data. Scientific Reports 7:7154.

Shinen, J. L., and S. A. Navarrete. 2014. Lottery coexistence on rocky shores: Weak niche differentiation or equal competitors engaged in neutral dynamics? American Naturalist 183:342–362.

Slessarev, E. W., Y. Lin, N. L. Bingham, J. E. Johnson, Y. Dai, J. P. Schimel, and O. A. Chadwick. 2016. Water balance creates a threshold in soil pH at the global scale. Nature 540:567–569.

Stachowicz, J. 2012. Niche expansion by positive interactions: realizing the fundamentals. A comment on Rodriguez-Cabal et al. Ideas in Ecology and Evolution 5:42–43.

Stephens, C. R., J. G. Heau, C. González, C. N. Ibarra-Cerdeña, V. Sánchez-Cordero, and C. González-Salazar. 2009. Using biotic interaction networks for prediction in biodiversity and emerging diseases. PLoS ONE 4:e5725.

Stephens, C. R., V. Sánchez-Cordero, and C. González Salazar. 2017. Bayesian inference of ecological interactions from spatial data. Entropy 19:547.

Ulrich, W. 2004. Species co-occurrences and neutral models: reassessing JM Diamond's assembly rules. Oikos 107:603–609.

Ulrich, W., and N. J. Gotelli. 2013. Pattern detection in null model analysis. Oikos 122:2–18.

Veech, J. A. 2014. The pairwise approach to analysing species co-occurrence. Journal of Biogeography 41:1029–1035.

Wieters, E. A., B. R. Broitman, and G. M. Branch. 2009. Benthic community structure and spatiotemporal thermal regimes in two upwelling ecosystems: comparisons between South Africa and Chile. Limnology and Oceanography 54:1060–1072.

Wright, J. P., C. G. Jones, and A. S. Flecker. 2002. An ecosystem engineer, the beaver, increases species richness at the landscape scale. Oecologia 132:96–101.

SUPPORTING INFORMATION